



Guida per gli sviluppatori

Amazon Machine Learning



Version Latest

Copyright © 2022 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon Machine Learning: Guida per gli sviluppatori

Copyright © 2022 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e il trade dress di Amazon non possono essere utilizzati in relazione ad alcun prodotto o servizio che non sia di Amazon, in alcun modo che possa causare confusione tra i clienti, né in alcun modo che possa denigrare o screditare Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

.....	ix
Cos'è Amazon Machine Learning?	1
Amazon Machine Learning	1
Origini dati	1
Modelli ML	3
Valutazioni	4
Previsioni in batch	5
Previsioni in tempo reale	6
Accesso ad Amazon Machine Learning	6
Regioni ed endpoint	7
Prezzi di Amazon ML	7
Stima dei costi delle previsioni in batch	8
Stima dei costi delle previsioni in tempo reale	10
Concetti di Machine Learning	11
Risoluzione dei problemi aziendali con Amazon Machine Learning	11
Quando usare il Machine Learning	12
Sviluppo di un'applicazione di machine learning	13
Formulazione del problema	13
Raccolta di dati etichettati	14
Analisi dei dati	14
Elaborazione delle caratteristiche	15
Divisione dei dati in dati di addestramento e di valutazione	17
Addestramento del modello	17
Valutazione dell'accuratezza del modello	21
Miglioramento dell'accuratezza del modello	26
Utilizzo del modello per effettuare previsioni	27
Riqualificazione dei modelli sui nuovi dati	28
Il processo Amazon Machine Learning	28
Impostazione di Amazon Machine Learning	31
Registrati ad AWS	31
Tutorial: Utilizzare Amazon ML per prevedere le risposte a un'offerta di marketing	32
Prerequisito	32
Fasi	32
Fase 1: Preparazione dei dati	33

Fase 2: Creare un'origine dati di addestramento	35
Fase 3: Creare un modello ML	41
Fase 4: Esaminare le prestazioni predittive del modello ML e impostare un punteggio soglia	42
Fase 5: Utilizzare il modello ML per generare previsioni	45
Fase 6: Eliminare	53
Creazione e utilizzo delle origini dati	55
Comprendere il formato dei dati per Amazon ML	55
Attributes	56
Requisiti relativi al formato dei file di input	56
Utilizzo di più file come input di dati per Amazon ML	57
Caratteri di fine riga nel formato CSV	57
Creazione di uno schema di dati per Amazon ML	58
Esempio di schema	59
Utilizzo del campo targetAttributeName	61
Utilizzo del campo rowID	61
Utilizzo del campo AttributeType	62
Fornire uno schema ad Amazon ML	64
Divisione dei dati	65
Pre-divisione dei dati	65
Divisione sequenziale dei dati	66
Divisione casuale dei dati	67
Informazioni sui dati	69
Statistiche descrittive	69
Accesso alle informazioni sui dati nella console di Amazon ML	70
Uso di Amazon S3 con Amazon ML	80
Caricamento di dati in Amazon S3	81
Autorizzazioni	81
Creazione di un'origine dati Amazon ML dai dati in Amazon Redshift	82
Parametri obbligatori per la procedura guidata Crea origine dati	82
Creazione di un'origine dati con Amazon Redshift Data (console)	87
Risoluzione dei problemi di Amazon Redshift	91
Utilizzo dei dati da un database Amazon RDS per creare un'origine dati Amazon ML	96
Identificatore di istanza di database RDS	98
Nome database MySQL	98
Credenziali utente di database	98
Informazioni di protezione AWS Data Pipeline	98

Informazioni sulla sicurezza di Amazon RDS	99
Query SQL MySQL	99
Percorso di output S3	100
Addestramento dei modelli ML	101
Tipi di modelli ML	101
Modello di classificazione binario	101
Modello di classificazione multiclasse	102
Modello di regressione	102
Processo di addestramento	102
Parametri di addestramento	103
Massima dimensione del modello	103
Numero massimo di passate sui dati	104
Tipo di mescolamento dei dati di addestramento	105
Tipo e quantità di regolarizzazione	106
parametri di addestramento: Tipi e valori predefiniti	106
Creazione di un modello ML	108
Prerequisiti	109
Creazione di un modello ML con opzioni predefinite	109
Creazione di un modello ML con opzioni personalizzate	110
Trasformazioni dei dati per il Machine Learning	112
L'importanza della trasformazione delle caratteristiche	112
Trasformazioni delle caratteristiche con le composizioni dati	113
Riferimenti relativi al formato delle composizioni	113
Gruppi	114
Assegnazioni	114
Output	115
Esempio completo di composizione	117
Composizioni suggerite	118
Riferimento per le trasformazioni di dati	119
Trasformazione N-gramma	120
Trasformazione OSB (Orthogonal Sparse Bigram)	121
Trasformazione in minuscolo	122
Trasformazione con rimozione della punteggiatura	122
Trasformazione binning quantile	123
Trasformazione di normalizzazione	124
Trasformazione del prodotto cartesiano	124

Riordino dei dati	126
Parametri DataRearrangement	127
Valutazione dei modelli ML	131
Informazioni del modello ML	132
Informazioni sul modello binario	132
Interpretazione delle previsioni	132
Informazioni del modello multiclasse	136
Interpretazione delle previsioni	136
Informazioni sul modello di regressione	139
Interpretazione delle previsioni	139
Prevenzione dell'overfitting	141
Convalida incrociata	142
Regolazione dei modelli	144
Avvisi relativi alla valutazione	145
Generazione e interpretazione delle previsioni	147
Creazione di una previsione in batch	147
Creazione di una previsione in batch (console)	148
Creazione di una previsione in batch (API)	148
Revisione dei parametri delle previsioni in batch	149
Revisione dei parametri delle previsioni in batch (console)	150
Revisione dei parametri e dei dettagli delle previsioni in batch (API)	150
Lettura dei file di output delle previsioni in batch	150
Individuazione del file manifest delle previsioni in batch.	150
Lettura del file manifest	151
Recupero dei file di output delle previsioni in batch	152
Interpretazione dei contenuti dei file della previsione in batch per un modello ML di classificazione binaria	152
Interpretazione dei contenuti dei file della previsione in batch per un modello ML di classificazione multiclasse	153
Interpretazione dei contenuti dei file della previsione in batch per un modello ML di regressione	154
Richiesta di previsioni in tempo reale	155
Prova delle previsioni in tempo reale	156
Creazione di un endpoint in tempo reale	158
Individuazione dell'endpoint delle previsioni in tempo reale (console)	159
Individuazione dell'endpoint delle previsioni in tempo reale (API)	160

Creazione di una richiesta di previsioni in tempo reale	160
Eliminazione di un endpoint in tempo reale	163
Gestione di oggetti Amazon ML	164
Elenco degli oggetti	164
Elencazione degli oggetti (Console)	165
Elencazione degli oggetti (API)	166
Recupero delle descrizioni degli oggetti	167
Descrizioni dettagliate nella console	167
Descrizioni dettagliate dall'API	167
Aggiornamento di oggetti	168
Eliminazione di oggetti	168
Eliminazione di oggetti (console)	169
Eliminazione di oggetti (API)	169
Monitoraggio di Amazon ML con i parametri di Amazon CloudWatch	171
Registrazione delle chiamate API di Amazon ML con AWS CloudTrail	172
Informazioni su Amazon ML in CloudTrail	172
Esempio: Voci del file di log di Amazon ML	174
Tagging degli oggetti	178
Nozioni di base sui tag	178
Limitazioni applicate ai tag	179
Tagging degli oggetti Amazon ML (console)	180
Tagging degli oggetti Amazon ML (API)	182
Riferimento Amazon Machine Learning	183
Concessione ad Amazon ML delle autorizzazioni per leggere i dati da Amazon S3	183
Concessione ad Amazon ML delle autorizzazioni per generare previsioni in Amazon S3	185
Controllo dell'accesso alle risorse Amazon ML con IAM	187
Sintassi delle policy	188
Specificazione delle azioni politiche IAM per Amazon ML	189
Specificazione degli ARN per le risorse Amazon ML nelle policy IAM	190
Esempi di policy per Amazon MLs	190
Prevenzione del confused deputy tra servizi	194
Gestione delle dipendenze nelle operazioni asincrone	195
Verifica dello stato della richiesta	196
Limiti di sistema	197
Nomi e ID per tutti gli oggetti	198
Durate degli oggetti	199

Risorse	200
Cronologia dei documenti	201

Non aggiorniamo più il servizio Amazon Machine Learning né accettiamo nuovi utenti. Questa documentazione è disponibile per gli utenti esistenti, ma non la aggiorneremo più. Per ulteriori informazioni, consulta la pagina [Cos'è Amazon Machine Learning](#).

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.

Cos'è Amazon Machine Learning?

Non stiamo più aggiornando il servizio Amazon Machine Learning (Amazon ML) o accettiamo nuovi utenti. Questa documentazione è disponibile per gli utenti esistenti, ma non la stiamo più aggiornando.

AWSora fornisce un efficace servizio basato su cloud (Amazon SageMaker) in modo che sviluppatori di qualunque livello possano utilizzare la tecnologia di machine learning. SageMaker è un servizio completamente gestito di machine learning che consente di creare potenti modelli di machine learning. Grazie a SageMaker, i data scientist e gli sviluppatori possono sviluppare e formare modelli di machine learning, che possono poi essere distribuiti direttamente in ambienti ospitati pronti per la produzione.

Per ulteriori informazioni, consulta la [Documentazione SageMaker](#).

Argomenti

- [Amazon Machine Learning](#)
- [Accesso ad Amazon Machine Learning](#)
- [Regioni ed endpoint](#)
- [Prezzi di Amazon ML](#)

Amazon Machine Learning

Questa sezione riepiloga i seguenti concetti fondamentali e descrive in modo più dettagliato come vengono utilizzati all'interno di Amazon ML:

- [Origini dati](#) contengono i metadati associati agli input dei dati relativi ad Amazon ML
- [Modelli ML](#) (modelli ML) generano previsioni utilizzando i modelli estratti dai dati di input
- [Valutazioni](#) valuta la qualità dei modelli ML
- [Previsioni in batch](#) generano in modo asincrono previsioni per più osservazioni di dati di input
- [Previsioni in tempo reale](#) generano in modo sincrono previsioni per singole osservazioni di dati

Origini dati

Un'origine dati è un oggetto che contiene metadati relativi ai dati di input. Amazon ML legge i dati di input, calcola le statistiche descrittive sui rispettivi attributi e memorizza le statistiche, insieme a uno

schema e ad altre informazioni, come parte dell'oggetto origine dati. Successivamente Amazon ML utilizza l'origine dati per addestrare e valutare un modello ML e generare previsioni in batch.

Important

Un'origine dati non archivia una copia dei dati in entrata. Archivia invece un riferimento alla posizione di Amazon S3 in cui si trovano i dati di input. Se si sceglie di spostare o modificare il file di Amazon S3, Amazon ML non può accedervi o utilizzarlo per creare un modello ML, generare valutazioni o generare previsioni.

La tabella seguente definisce i termini correlati alle origini dati.

Termine	Definizione
Attributo	<p>Una proprietà denominata univoca all'interno di un'osservazione. Nel caso di dati sotto forma di tabella, come ad esempio fogli di calcolo o file di valori separati da virgola (.csv), le intestazioni delle colonne rappresentano gli attributi e le righe contengono i valori per ogni attributo.</p> <p>Sinonimi: variabile, nome variabile, campo, colonna</p>
Nome origine dati	(facoltativo) consente di definire un nome in formato leggibile per un'origin e dati. Questi nomi consentono di individuare e gestire le origini dati nella console di Amazon ML.
Dati di input	Nome collettivo per tutte le osservazioni a cui un'origine dati fa riferimento.
Ubicazione	Posizione dei dati di input. Al momento, Amazon ML può utilizzare i dati archiviati all'interno dei bucket di Amazon S3, dei database Amazon Redshift, oppure dei database MySQL in Amazon Relational Database Service (RDS).
Osservazione	<p>Una singola unità dei dati di input. Ad esempio, se si crea un modello ML per rilevare transazioni fraudolente, i dati di input saranno costituiti da molte osservazioni, ciascuna delle quali rappresenta una singola transazione.</p> <p>Sinonimi: record, esempio, istanza, riga</p>

Termine	Definizione
ID riga	<p>(facoltativo) Un flag che, se specificato, identifica un attributo dei dati di input da includere nell'output di previsione. Questo attributo semplifica l'associazione tra una previsione e la corrispondente osservazione.</p> <p>Sinonimi: identificatore riga</p>
Schema	<p>Le informazioni necessarie per interpretare i dati di input, inclusi i nomi degli attributi e i relativi tipi di dati, nonché i nomi degli attributi speciali.</p>
Statistiche	<p>Riepilogo delle statistiche per ogni attributo dei dati di input. Queste statistiche hanno due funzioni:</p> <p>La console di Amazon ML li visualizza in formato grafico per consentire di comprendere i dati subito e identificare irregolarità o errori.</p> <p>Amazon ML li utilizza durante il processo di addestramento per migliorare la qualità del modello ML risultante.</p>
Stato	<p>Indica lo stato attuale dell'origine dati, ad esempio In Progress (In corso), Completed (Completato) o Failed (Non riuscito).</p>
Attributo di destinazione	<p>Nel contesto dell'addestramento di un modello ML, l'attributo di destinazione identifica il nome dell'attributo nei dati di input che contengono le risposte "corrette". Amazon ML lo utilizza per individuare i pattern dei dati di input e generare un modello ML. Nel contesto della valutazione e generazione delle previsioni, l'attributo di destinazione è l'attributo il cui valore verrà previsto da un modello ML addestrato.</p> <p>Sinonimi: target</p>

Modelli ML

Un modello ML è un modello matematico in grado di generare previsioni individuando pattern nei dati. Amazon ML supporta tre tipi di modelli ML: di classificazione binaria, di classificazione multiclasse e di regressione.

La tabella seguente definisce i termini correlati ai modelli ML.

Termine	Definizione
Regressione	L'obiettivo dell'addestramento di un modello ML di regressione è prevedere un valore numerico.
Multiclasse	L'obiettivo dell'addestramento di un modello ML multiclasse è prevedere i valori che appartengono a una serie limitata e predefinita di valori consentiti.
Binario	L'obiettivo dell'addestramento di un modello ML binario è prevedere i valori che possono avere solo uno di due stati, ad esempio true (vero) o false (falso).
Dimensione del modello	I modelli ML acquisiscono e memorizzano pattern. Più sono i pattern memorizzati da un modello ML, maggiori sono le sue dimensioni. La dimensione del modello ML è descritta in Mbyte.
Numero di passate	Quando si addestra un modello ML, è possibile utilizzare i dati provenienti da un'origine dati. È talvolta vantaggioso utilizzare ogni record di dati più di una volta nel processo di addestramento. Il numero di volte che si permette ad Amazon ML di usare lo stesso record di dati viene chiamato numero di passate.
Regolarizzazione	La regolarizzazione è una tecnica di machine learning che è possibile usare per ottenere modelli di qualità superiore. Amazon ML offre un'impostazione predefinita che funziona ottimamente per la maggior parte dei casi.

Valutazioni

Una valutazione misura la qualità del modello ML e stabilisce se è performante.

La tabella seguente definisce i termini correlati alle valutazioni.

Termine	Definizione
Informazioni del modello	Amazon ML fornisce un parametro e un numero di informazioni che è possibile utilizzare per valutare le prestazioni predittive del modello.

Termine	Definizione
AUC	L'AUC (Area Under the ROC Curve) misura la capacità di un modello ML binario di prevedere un punteggio più elevato per gli esempi positivi rispetto agli esempi negativi.
Punteggio macro-medio F1	Il punteggio macro-medio F1 viene utilizzato per valutare le prestazioni predittive dei modelli ML multiclasse.
RMSE	L'RMSE (Root Mean Square Error) è un parametro utilizzato per valutare le prestazioni predittive dei modelli ML di regressione.
Interruzione	I modelli ML funzionano generando punteggi di previsioni numeriche. Applicando un valore di interruzione, il sistema converte tali punteggi in etichette 0 e 1.
Accuratezza	L'accuratezza misura la percentuale di previsioni corrette.
Precisione	La precisione mostra la percentuale di istanze positive effettive (anziché falsi positivi) tra le istanze recuperate (quelle previste come positive). In altre parole, quanti elementi selezionati sono positivi?
Recall	Il recall mostra la percentuale di positivi effettivi nel numero totale di istanze pertinenti (positivi effettivi). In altre parole, quanti elementi positivi sono selezionati?

Previsioni in batch

Le previsioni in batch prevedono un insieme di osservazioni che possono essere eseguite contemporaneamente. È l'ideale per le analisi predittive che non hanno un requisito di tempo reale.

La tabella seguente definisce i termini correlati alle previsioni in batch.

Termine	Definizione
Percorso di output	I risultati di una previsione in batch sono memorizzati nel percorso di output di un bucket S3.

Termine	Definizione
File manifest	Il file manifest si riferisce a ciascun file di dati di input con i relativi risultati della previsione in batch. È memorizzato nel percorso di output del bucket S3.

Previsioni in tempo reale

Le previsioni in tempo reale sono per le applicazioni con un requisito di bassa latenza, ad esempio applicazioni Web interattive, mobili o desktop. È possibile eseguire query relative a previsioni su qualsiasi modello ML utilizzando l'API di previsione in tempo reale a bassa latenza.

La tabella seguente definisce i termini correlati alle previsioni in tempo reale.

Termine	Definizione
API di previsione in tempo reale	L'API di previsione in tempo reale accetta una sola osservazione di input nel payload della richiesta e restituisce la previsione nella risposta.
Endpoint di previsione in tempo reale	Per utilizzare un modello ML con l'API di previsione in tempo reale, è necessario creare un endpoint di previsione in tempo reale. Una volta creato, l'endpoint contiene l'URL utilizzabile per richiedere previsioni in tempo reale.

Accesso ad Amazon Machine Learning

È possibile accedere ad Amazon ML utilizzando una delle seguenti opzioni:

Console Amazon ML

È possibile accedere alla console Amazon ML accedendo alla Console di gestione AWS e aprendo la console Amazon ML in <https://console.aws.amazon.com/machinelearning/>.

AWS CLI

Per ulteriori informazioni su come installare e configurare AWS CLI, consultare la pagina relativa alla configurazione dell'interfaccia a riga di comando di AWS nella [AWS Command Line Interface Guida per l'utente di](#).

API Amazon ML

Per ulteriori informazioni sull'API Amazon ML, consulta [Riferimento dell'API Amazon ML](#).

SDK AWS

Per ulteriori informazioni sui SDK AWS, consultare la pagina relativa agli [strumenti per Amazon Web Services](#).

Regioni ed endpoint

Amazon Machine Learning (Amazon ML) supporta gli endpoint di previsione in tempo reale nelle due regioni seguenti:

Nome regione	Regione	Endpoint	Protocollo
US East (N. Virginia)	us-east-1	machinelearning.us-east-1.amazonaws.com	HTTPS
Europe (Ireland)	eu-west-1	machinelearning.eu-west-1.amazonaws.com	HTTPS

È possibile effettuare l'hosting di set di dati, addestrare e valutare modelli e attivare previsioni in qualsiasi regione.

È consigliabile mantenere tutte le risorse nella stessa regione. Se i dati di input si trova in una regione differente rispetto alle risorse Amazon ML, vengono addebitati costi per il trasferimento transregionale dei dati. È possibile richiamare un endpoint di previsione in tempo reale da qualsiasi regione, ma richiamare un endpoint da una regione che non ha l'endpoint che si sta chiamando può avere ripercussioni sulle latenze delle previsioni in tempo reale.

Prezzi di Amazon ML

I prezzi dei servizi AWS sono calcolati solo in base all'uso effettivo. Non sono previste tariffe minime né impegni anticipati.

Amazon Machine Learning (Amazon ML) addebita una tariffa oraria per il tempo di elaborazione impiegato per calcolare i dati statistici e addestrare e valutare modelli; successivamente si paga in base al numero di previsioni generate per l'applicazione. Per le previsioni in tempo reale si paga anche una tariffa oraria per la capacità riservata, in base alle dimensioni del modello.

Amazon ML stima i costi per le previsioni solo nella [Console Amazon ML](#).

Per ulteriori informazioni sui prezzi di Amazon ML, consulta [Prezzi Amazon Machine Learning](#).

Argomenti

- [Stima dei costi delle previsioni in batch](#)
- [Stima dei costi delle previsioni in tempo reale](#)

Stima dei costi delle previsioni in batch

Quando si richiedono previsioni in Batch da un modello Amazon ML utilizzando la procedura guidata Crea previsione in batch, Amazon ML stima il costo di tali previsioni. Il metodo per calcolare la stima varia in base al tipo di dati disponibili.

Stima dei costi delle previsioni in batch quando sono disponibili dati statistici

La stima più accurata dei costi si ottiene quando Amazon ML ha già calcolato le statistiche di riepilogo sull'origine dati utilizzata per richiedere le previsioni. Queste statistiche sono sempre calcolate per origini dati create utilizzando la console di Amazon ML. Gli utenti di API devono impostare il flag `ComputeStatistics` su `True` quando creano origini dati tramite codice utilizzando le API [CreateDataSourceFromS3](#), [CreateDataSourceFromRedshift](#) o [CreateDataSourceFromRDS](#). Affinché le statistiche siano disponibili, l'origine dati deve essere nello stato `READY`.

Una delle statistiche che Amazon ML calcola è il numero di record di dati. Quando è disponibile il numero di record di dati, la procedura guidata Amazon ML Crea previsione in Batch stima il numero di previsioni moltiplicando il numero di record di dati per [commissione per le previsioni dei lotti](#).

Il costo effettivo può variare rispetto a tale stima per i motivi seguenti:

- Alcuni dei record di dati potrebbe fallire l'elaborazione. Le previsioni da record di dati non riusciti non vengono fatturate.
- La stima non considera i crediti preesistenti o altre modifiche applicate da AWS.

The screenshot shows the 'Batch prediction results' page in the Amazon Machine Learning console. The page is divided into four steps: 1. ML model for batch prediction, 2. Data for batch prediction, 3. Batch prediction results (current step), and 4. Review. A blue box highlights the estimated cost: 'The estimated cost for generating your predictions is \$4.20. This estimate is based on the 41188 data records included in your prediction request.' Below this, it states 'The Amazon ML fee for batch predictions is \$0.10/1000 predictions rounded to nearest penny. [Learn more](#)'. The user is prompted to 'Type the path to the S3 location in which the prediction results will be saved.' The 'S3 destination' field contains 's3:// Bucket-name/Folder-name/'. The 'Batch prediction name (Optional)' field contains 'Batch prediction: ML model: Banking.csv'. At the bottom, there are three buttons: 'Cancel', 'Previous', and 'Review' (highlighted in blue). The footer includes 'Feedback', 'English', '© 2008 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved.', 'Privacy Policy', and 'Terms of Use'.

Stima dei costi delle previsioni in batch quando sono disponibili solo le dimensioni

Quando si richiede una previsione in batch e i dati statistici per l'origine dati richiesta non sono disponibili, Amazon ML stima il costo in base a quanto segue:

- La dimensione totale dei dati che è calcolata e persiste durante la convalida dell'origine dati
- La dimensione media dei record di dati che Amazon ML stima leggendo e analizzando i primi 100 MB dei file di dati

Per stimare l'importo della previsione in batch, Amazon ML divide la dimensione totale dei dati per la dimensione media dei record di dati. Questo metodo di previsione del costo è meno preciso del metodo utilizzato quando è disponibile il numero di record di dati perché i primi record del file di dati potrebbero non rappresentare accuratamente la dimensione media dei record.

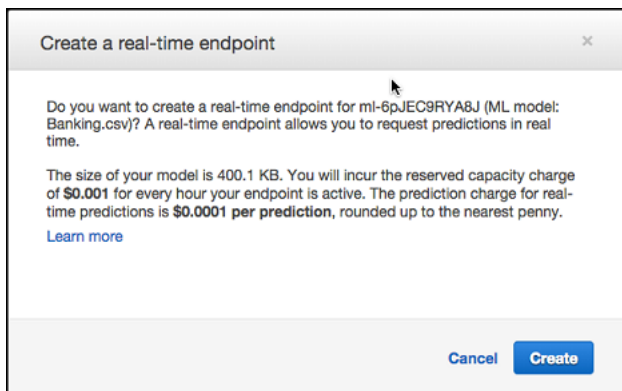
Stima dei costi delle previsioni in batch quando non sono disponibili né dati statistici né dimensioni dei dati

Quando non sono disponibili né dati statistici né dimensioni dei dati, Amazon ML non riesce a stimare il costo delle previsioni in batch. Ciò avviene frequentemente quando l'origine dati che si sta utilizzando per richiedere le previsioni in batch non è ancora stata convalidata da Amazon ML. Questo può accadere quando è stata creata un'origine dati che si basa su una query Amazon Redshift (Amazon Redshift) o Amazon Relational Database Service (Amazon RDS) e il trasferimento dei dati non è ancora stato completato, oppure quando la creazione dell'origine dati è in attesa dietro

ad altre operazioni nell'account. In questo caso, Amazon ML di informa l'utente in merito alle tariffe per la previsione in batch. Si può scegliere di procedere con la richiesta di previsione in batch senza una stima, oppure di annullare la procedura guidata e di ritornare quando l'origine dati utilizzata per le previsioni si trova nello stato INPROGRESS o READY.

Stima dei costi delle previsioni in tempo reale

Quando viene creato un endpoint di previsione in tempo reale utilizzando la console di Amazon ML, viene visualizzata la stima della tariffa per la capacità riservata applicata per riservare l'endpoint per l'elaborazione della previsione. Questo addebito varia in base alle dimensioni del modello, come spiegato nella [pagina dei prezzi dei servizi](#). L'utente, inoltre, viene informato della tariffa standard Amazon ML per le previsioni in tempo reale.



Concetti di Machine Learning

Il Machine learning (ML) consente di utilizzare dati storici per prendere migliori decisioni di business. Gli algoritmi ML rilevano pattern nei dati e creano modelli matematici utilizzando quanto rilevato. Questi modelli possono quindi essere utilizzati per generare previsioni basate su dati futuri. Ad esempio, una possibile applicazione di un modello di machine learning potrebbe essere la previsione della probabilità che un cliente acquisti un determinato prodotto in base al suo comportamento precedente.

Argomenti

- [Risoluzione dei problemi aziendali con Amazon Machine Learning](#)
- [Quando usare il Machine Learning](#)
- [Sviluppo di un'applicazione di machine learning](#)
- [Il processo Amazon Machine Learning](#)

Risoluzione dei problemi aziendali con Amazon Machine Learning

È possibile usare Amazon Machine Learning per applicare il machine learning a problemi riguardo ai quali si dispone di esempi di risposte reali. Ad esempio, se si desidera utilizzare Amazon Machine Learning per prevedere se un'e-mail è spam, è necessario raccogliere esempi di e-mail che vengono correttamente etichettate come spam o non spam. È quindi possibile utilizzare il machine learning per generalizzare questi esempi di e-mail, per prevedere la probabilità che una nuova e-mail sia spam o meno. Questo approccio di apprendimento da dati che sono stati etichettati con la risposta reale è noto come machine learning controllato.

È possibile utilizzare gli approcci ML controllati per queste attività specifiche di machine learning: classificazione binaria (prevedere uno di due possibili risultati), classificazione multiclasse (prevedere uno di più di due risultati) e regressione (prevedere un valore numerico).

Esempi di problemi di classificazione binaria:

- Il cliente acquisterà questo prodotto oppure no?
- Questa e-mail è spam o non spam?
- Questo prodotto è un libro o un animale da fattoria?
- Questa recensione è scritta da un cliente o da un robot?

Esempi di problemi di classificazione multilclasse:

- Questo prodotto è un libro, un film o un abito?
- Questo film è una commedia romantica, un documentario o un thriller?
- Quale categoria di prodotti è più interessante per questo cliente?

Esempi di problemi di classificazione di regressione:

- Che temperatura ci sarà domani a Seattle?
- Per questo prodotto, quante unità si venderanno?
- Quanti giorni trascorreranno prima che questo cliente interrompa l'uso dell'applicazione?
- A che prezzo sarà venduta questa casa?

Quando usare il Machine Learning

È importante ricordare che ML non è una soluzione per ogni tipo di problema. Vi sono alcuni casi in cui è possibile sviluppare soluzioni affidabili senza usare le tecniche ML. Ad esempio, non è necessario servirsi di ML se è possibile determinare un valore di destinazione utilizzando regole semplici, calcoli o fasi predeterminati che possono essere programmati senza che occorra un apprendimento basati sui dati.

Utilizzare il machine learning per le seguenti situazioni:

- Non è possibile codificare le regole: Molte attività umane (ad esempio riconoscere se un'e-mail è spam o non spam) non possono essere svolte adeguatamente utilizzando una semplice (deterministica) soluzione basata su regole. Un numero elevato di fattori possono influenzare la risposta. Quando le regole dipendono da troppi fattori e molte di queste regole si sovrappongono o devono essere ottimizzate in modo molto accurato, le persone incontrano difficoltà a effettuare una codifica accurata di tali regole. È possibile utilizzare ML in modo efficiente per risolvere il problema.
- Non è possibile scalare: È possibile riconoscere manualmente alcune centinaia di e-mail e decidere se si tratta di spam o meno. Tuttavia, questa operazione diventa noiosa quando si parla di milioni di e-mail. Le soluzioni ML sono efficaci quando si tratta di gestire problemi su vasta scala.

Sviluppo di un'applicazione di machine learning

Lo sviluppo di applicazioni ML è un processo iterativo che richiede una sequenza di fasi. Per sviluppare un'applicazione ML, seguire queste fasi generali:

1. Individuare i principali problemi di ML in termini di ciò che si osserva e di quale risposta si desidera che il modello preveda.
2. Raccogliere, ripulire e preparare i dati per renderli adatti all'utilizzo da parte degli algoritmi di addestramento del modello ML. Visualizzare e analizzare i dati per l'esecuzione di controlli di integrità, per convalidare la qualità dei dati e per comprendere i dati.
3. Spesso i dati grezzi (variabili di input) e la risposta (target) non sono rappresentati in un modo utilizzabile per addestrare un modello altamente predittivo. Pertanto, in genere è necessario tentare di costruire rappresentazioni o funzionalità di input più predittive dalle variabili grezze.
4. Fornire le caratteristiche risultanti all'algoritmo di apprendimento per creare i modelli e valutare la qualità dei modelli sui dati che non sono stati utilizzati per lo sviluppo del modello.
5. Utilizzare il modello per generare previsioni della risposta target per nuove istanze di dati.

Formulazione del problema

Il primo passo nel machine learning è quello di decidere ciò che si desidera prevedere, ovvero l'etichetta o la risposta target. Si può immaginare uno scenario in cui si desidera fabbricare prodotti, ma la decisione di fabbricare ogni prodotto dipende dal numero di vendite potenziali. In questo scenario, si vuole prevedere quante volte ogni prodotto sarà acquistato (previsione del numero di vendite). Esistono diversi modi per definire questo problema utilizzando il machine learning. La scelta di come definire il problema varia a seconda del caso d'uso o delle esigenze aziendali.

Si desidera prevedere il numero di acquisti che i clienti effettueranno per ogni prodotto (nel qual caso il target è numerico e si sta risolvendo un problema di regressione)? Oppure si vuole prevedere quali prodotti avranno più di 10 acquisti (nel qual caso il target è binario e si sta risolvendo un problema di classificazione binaria)?

È importante evitare di complicare eccessivamente il problema e trovare la soluzione più semplice adatta alle esigenze. Tuttavia, è anche importante evitare di perdere informazioni, soprattutto le informazioni nella cronologia risposte. In questo caso, la conversione del numero di vendite precedenti effettive in una variabile binaria "oltre 10" invece di "meno" farebbe perdere informazioni preziose. Se si investe tempo nel decidere quale target abbia più senso prevedere, si eviterà la creazione di modelli che non rispondono alla domanda.

Raccolta di dati etichettati

I problemi con ML iniziano con i dati, preferibilmente molti dati (esempi o osservazioni) di cui si conosce già la risposta target. I dati per i quali si conosce già la risposta target vengono chiamati dati etichettati. Nel ML controllato, l'algoritmo insegna a se stesso ad apprendere dagli esempi etichettati forniti.

Ogni esempio/osservazione nei dati deve contenere due elementi:

- **Il target:** la risposta che si desidera prevedere. Si forniscono dati che vengono etichettati con il target (risposta corretta) per consentire all'algoritmo ML di apprendere. Quindi, si potrà utilizzare il modello ML addestrato per prevedere tale risposta sui dati di cui non si conosce la risposta target.
- **Variabili/caratteristiche:** si tratta di attributi dell'esempio che possono essere utilizzati per identificare pattern che consentano di prevedere la risposta target.

Ad esempio, per il problema della classificazione delle e-mail, il target è un'etichetta che indica se un'e-mail è spam o non spam. Esempi di variabili sono il mittente dell'e-mail, il testo nel corpo del messaggio e-mail, il testo dell'oggetto, l'ora in cui è stata inviata l'e-mail e la presenza di precedente corrispondenza tra il mittente e il destinatario.

Spesso i dati non sono immediatamente disponibili in forma etichettata. La raccolta e la preparazione delle variabili e del target sono spesso le fasi più importanti nella risoluzione di un problema ML. I dati di esempio devono essere rappresentativi dei dati che si avranno a disposizione durante l'utilizzo del modello per effettuare una previsione. Ad esempio, se si desidera prevedere se un'e-mail è o non è spam, è necessario raccogliere dati positivi (e-mail spam) e negativi (e-mail non spam) affinché l'algoritmo di machine learning sia in grado di trovare pattern che distinguano tra i due tipi di e-mail.

Una volta etichettati i dati, potrebbe essere necessario convertirli in un formato accettabile per l'algoritmo o il software. Ad esempio, per utilizzare Amazon ML occorre convertire i dati in formato CSV (elenco separato da virgole), dove ogni esempio costituisce una riga del file CSV, ogni colonna contiene una variabile di input e una singola colonna contiene la risposta target.

Analisi dei dati

Prima di fornire i dati etichettati a un algoritmo ML, è consigliabile ispezionare i dati per identificare i problemi e ottenere informazioni sui dati che si stanno utilizzando. La capacità predittiva del modello è elevata solo se anche la qualità dei dati forniti è elevata.

Quando si analizzano i dati, è necessario tenere presenti le seguenti considerazioni:

- **Riepiloghi delle variabili e dei dati di destinazione:** è utile comprendere i valori che le proprie variabili assumono e quali sono i valori dominanti nei dati. È possibile far controllare questi riepiloghi a un esperto del problema che si desidera risolvere. La domanda che si deve porre o che si deve fare all'esperto I dati sono all'altezza delle aspettative? Ci potrebbe essere un problema di raccolta dei dati? Una classe del target è più frequente rispetto alle altre classi? Vi sono più valori mancanti o dati non validi del previsto?
- **Correlazioni variabili-target:** è utile conoscere la correlazione tra ogni variabile e la classe target perché un'elevata correlazione implica l'esistenza di una relazione tra la variabile e la classe target. In generale, è preferibile includere variabili con elevata correlazione perché sono quelle con la più elevata capacità predittiva (segnale) ed escludere variabili con bassa correlazione, perché probabilmente sono irrilevanti.

In Amazon ML, è possibile analizzare i dati mediante la creazione di un'origine dati e rivedendo il report dei dati ottenuto.

Elaborazione delle caratteristiche

Dopo avere imparato a conoscere i propri dati tramite i riepiloghi e le visualizzazioni, è possibile trasformare ulteriormente le proprie variabili per renderle più significative. Questo processo è noto come elaborazione delle caratteristiche. Si supponga, ad esempio, di avere una variabile che acquisisce la data e l'ora in cui si è verificato un evento. Tale data e ora non si verificheranno mai più e, di conseguenza, non saranno utili per prevedere il target. Tuttavia, se questa variabile è trasformata in caratteristiche che rappresentano l'ora del giorno, il giorno della settimana e il mese, tali variabili potrebbero essere utili per scoprire se l'evento tende a verificarsi in una determinata ora, giorno feriale o mese. L'elaborazione delle caratteristiche per formare punti di dati più generalizzabili da cui apprendere può apportare notevoli miglioramenti ai modelli predittivi.

Altre esempi di elaborazione di funzionalità comuni:

- **Sostituzione dei dati mancanti o non validi con valori più significativi** (ad esempio, se si sa che il valore mancante di una variabile tipo prodotto significa effettivamente che si tratta di un libro, è possibile sostituire tutti i valori mancanti nel tipo prodotto con il valore per il libro). Una strategia comune utilizzata per imputare i valori mancanti consiste nel sostituire i valori mancanti con la media o il valore mediano. È importante comprendere i dati prima di scegliere una strategia per la sostituzione dei valori mancanti.
- **Formazione di prodotti cartesiani di una variabile con un'altra.** Ad esempio, se si dispone di due variabili, come la densità di popolazione (urban, suburban, rural) e lo stato (Washington,

Oregon, California), potrebbero esservi informazioni utili nelle caratteristiche costituite da un prodotto cartesiano di queste due variabili, che genera caratteristiche (urban_Washington, suburban_Washington, rural_Washington, urban_Oregon, suburban_Oregon, rural_Oregon, urban_California, suburban_California, rural_California).

- Trasformazioni non lineari come il binning di variabili numeriche alle categorie. In molti casi, il rapporto tra una caratteristica numerica e la destinazione è non lineare (il valore della caratteristica non aumenta né diminuisce monotonicamente con la destinazione). In questi casi, può essere utile effettuare il binning della caratteristica numerica in caratteristiche categoriche che rappresentano diversi intervalli della caratteristica numerica. Ogni caratteristica categorica (bin) può quindi essere modellata come se avesse la propria relazione lineare con la destinazione. Ad esempio, si supponga di sapere che la caratteristica numerica continua età è correlata non linearmente alla probabilità di acquistare un libro. È possibile effettuare il binning dell'età in caratteristiche categoriche che potrebbero essere in grado di acquisire in modo più preciso il rapporto con la destinazione. Il numero ottimale di bin per una variabile numerica dipende dalle caratteristiche della variabile e dal suo rapporto con la destinazione, e ciò si determina meglio attraverso la sperimentazione. Amazon ML suggerisce il numero ottimale di bin per una caratteristica numerica basata su dati statistici nella composizione suggerita. Consultare la Guida per gli sviluppatori per ulteriori informazioni sulla composizione suggerita.
- Caratteristiche specifiche di dominio (ad esempio, se si dispone di lunghezza, larghezza e altezza come variabili separate, è possibile creare una nuova caratteristica volume che sia il prodotto di queste tre variabili).
- Caratteristiche specifiche delle variabili. Alcuni tipi di variabili, come le caratteristiche del testo, che acquisiscono la struttura di una pagina Web o la struttura di una frase, hanno metodi di elaborazione generici che consentono di estrarre struttura e contesto. Ad esempio, la formazione di n-grammi con il testo "the fox jumped over the fence" può essere rappresentata con unigrammi: the, fox, jumped, over, fence o bigrammi: the fox, fox jumped, jumped over, over the, the fence.

L'inclusione delle caratteristiche più importanti aiuta a migliorare la capacità predittiva. Chiaramente, non è sempre possibile conoscere in anticipo le caratteristiche dotate di un "segnale" o di un'influenza predittiva. Pertanto è opportuno includere tutte le caratteristiche che potenzialmente sono correlate all'etichetta di destinazione e lasciare che l'algoritmo di addestramento del modello scelga la caratteristiche con le correlazioni più forti. In Amazon ML, l'elaborazione delle caratteristiche può essere specificata nella composizione durante la creazione di un modello. Consultare la Developer Guide per un elenco di processori di caratteristiche disponibili.

Divisione dei dati in dati di addestramento e di valutazione

L'obiettivo fondamentale di ML è generalizzare al di là delle istanze dei dati utilizzate per preparare i modelli. Si vuole valutare il modello per stimare la qualità della generalizzazione dei suoi pattern per i dati su cui il modello è stato addestrato. Tuttavia, poiché le istanze future hanno valori di destinazione sconosciuti e non è possibile verificare ora l'accuratezza delle previsioni per le istanze future, è necessario utilizzare alcuni dei dati di cui si conosce già la risposta come proxy per i dati futuri. Non è utile valutare un modello con gli stessi dati utilizzati per l'addestramento, perché in questo modo si premiano modelli in grado di "ricordare" i dati di addestramento, anziché utilizzarli per la generalizzazione.

Una strategia comune consiste nell'utilizzare tutti i dati etichettati disponibili e frazionarli in sottoinsiemi per l'addestramento e la valutazione, di solito con un rapporto del 70-80% per l'addestramento e del 20-30% per la valutazione. Il sistema ML impiega i dati di addestramento per addestrare i modelli a visualizzare i pattern e utilizza i dati di valutazione per valutare la qualità predittiva del modello di addestramento. Il sistema ML valuta le prestazioni predittive confrontando le previsioni sull'insieme dei dati di valutazione con i valori "true" (noti come valori acquisiti sul campo), utilizzando una serie di parametri. Di solito, è possibile utilizzare il modello "migliore" sul sottoinsieme di valutazione per fornire previsioni sulle istanze future di cui non si conosce la risposta target.

Amazon ML divide i dati inviati per l'addestramento di un modello tramite la console Amazon ML in un 70% per l'addestramento e in un 30% per la valutazione. Per impostazione predefinita, Amazon ML usa il primo 70% dei dati di input nell'ordine in cui appaiono nei dati di origine per l'origine dati di addestramento e il restante 30% dei dati per l'origine dati di valutazione. Amazon ML consente inoltre di selezionare a caso il 70% dei dati di origine per l'addestramento anziché utilizzare il primo 70%, e di utilizzare la parte restante di questo sottoinsieme casuale per la valutazione. È possibile utilizzare le API Amazon ML per specificare rapporti di divisione personalizzati e per fornire dati di addestramento e valutazione divisi al di fuori di Amazon ML. Amazon ML offre inoltre strategie per dividere i dati. Per ulteriori informazioni sulle strategie di divisione, consultare [Divisione dei dati](#).

Addestramento del modello

Ora è possibile fornire all'algoritmo ML (ovvero all'algoritmo di apprendimento) i dati di addestramento. L'algoritmo apprende dai dati di addestramento i pattern che mappano le variabili alla destinazione e genera un modello che consente di acquisire queste relazioni. Il modello ML può essere utilizzato per ottenere previsioni su nuovi dati di cui non si conosce la risposta target.

Modelli lineari

Esiste un numero elevato di modelli ML disponibili. Amazon ML apprende un tipo di modello ML: i modelli lineari. Il termine modello lineare presuppone che il modello sia specificato come combinazione lineare di caratteristiche. In base ai dati di addestramento, il processo di apprendimento calcola un peso per ogni caratteristica, per formare un modello in grado di prevedere o stimare il valore di destinazione. Ad esempio, se l'obiettivo è la quantità di assicurazione che un cliente acquisterà e le variabili sono l'età e il reddito, un modello lineare semplice sarebbe il seguente:

```
Estimated target = 0.2 + 5·age + 0.0003·income
```

Algoritmo di apprendimento

L'algoritmo di apprendimento ha il compito di apprendere il sistema di ponderazione per il modello. Il sistema di ponderazione descrive la probabilità che i pattern che il modello sta apprendendo riflettano le relazioni effettive nei dati. Un algoritmo di apprendimento è composto da una funzione di perdita e da una tecnica di ottimizzazione. La perdita è la sanzione che si subisce quando la stima della destinazione fornita dal modello ML non è esattamente uguale alla destinazione. Una funzione di perdita quantifica tale sanzione come un singolo valore. Una tecnica di ottimizzazione cerca di ridurre al minimo la perdita. In Amazon Machine Learning, si utilizzano tre funzioni di perdita, una per ciascuno dei tre tipi di problemi di previsione. La tecnica di ottimizzazione utilizzata in Amazon ML è la discesa stocastica del gradiente (Stochastic Gradient Descent - SGD) online. La SGD effettua passate sequenziali sui dati di addestramento e durante ogni passata aggiorna le ponderazioni delle caratteristiche, un esempio alla volta, con l'obiettivo di raggiungere le ponderazioni ottimali in grado di ridurre al minimo la perdita.

Amazon ML usa i seguenti algoritmi di apprendimento:

- Per la classificazione binaria Amazon ML impiega la regressione logistica (funzione di perdita logistica +SGD).
- Per la classificazione multiclasse Amazon ML impiega la regressione logistica multinomiale (perdita logistica multinomiale +SGD).
- Per la regressione Amazon ML impiega la regressione lineare (funzione di perdita al quadrato +SGD).

Parametri di addestramento

L'algoritmo di apprendimento Amazon ML accetta parametri, chiamati iperparametri o parametri di addestramento, che consentono di controllare la qualità del modello risultante. A seconda

dell'iperparametro, Amazon ML seleziona automaticamente le impostazioni o fornisce valori statici di default per gli iperparametri. Anche se le impostazioni di default degli iperparametri generalmente producono modelli utili, si potrebbero migliorare le prestazioni predittive dei modelli modificando i valori degli iperparametri. Le seguenti sezioni descrivono l'iper-prezzo comuneparametri associati ad algoritmi di apprendimento per modelli lineari, come ad esempio quelli creati da Amazon ML.

Velocità di apprendimento

La velocità di apprendimento è un valore costante utilizzato nell'algoritmo di discesa stocastica del gradiente (Stochastic Gradient Descent - SGD). La velocità di apprendimento influenza la velocità con cui l'algoritmo raggiunge (converge su) le ponderazioni ottimali. L'algoritmo SGD aggiorna le ponderazioni del modello lineare per ogni esempio di dati rilevato. La dimensione di questi aggiornamenti è controllata dalla velocità di apprendimento. Una velocità di apprendimento troppo elevata potrebbe impedire alle ponderazioni di raggiungere la soluzione ottimale. Un valore troppo basso fa sì che l'algoritmo richieda troppe passate per avvicinarsi alle ponderazioni ottimali.

In Amazon ML, la velocità di apprendimento è selezionata in base ai dati dell'utente.

Dimensione del modello

Se si dispone di numerose caratteristiche di input, il numero di pattern possibili per i dati può portare a un modello di grandi dimensioni. I modelli di grandi dimensioni hanno implicazioni pratiche in quanto, ad esempio, richiedono più RAM per contenere il modello durante l'addestramento e quando vengono generate le previsioni. In Amazon ML, è possibile ridurre la dimensione del modello utilizzando la regolarizzazione L1 o limitando specificamente la dimensione del modello attraverso l'indicazione delle dimensioni massime. Si noti che se si riduce eccessivamente la dimensione del modello, si potrebbe ridurre anche la capacità predittiva di quest'ultimo.

Per ulteriori informazioni sulla dimensione del modello predefinito, consultare [parametri di addestramento: Tipi e valori predefiniti](#). Per ulteriori informazioni sulla regolarizzazione, consultare [Regolarizzazione](#).

Numero di passate

L'algoritmo SGD effettua passate sequenziali sui dati di addestramento. Il parametro `Number of passes` controlla il numero di passate che l'algoritmo effettua sui dati di addestramento. Ulteriori passate risultano in un modello che si adatta meglio ai dati (se la velocità di apprendimento non è troppo elevata), ma il vantaggio diminuisce con un numero crescente di passate. Per insiemi di dati di dimensioni inferiori, è possibile aumentare in modo significativo il numero di passate, consentendo in

tal modo all'algoritmo di apprendimento di adattarsi maggiormente e in modo efficiente ai dati. In caso di insiemi di dati di grandissime dimensioni, una singola passata potrebbe essere sufficiente.

Per maggiori informazioni sul numero predefinito di passate, consultare [parametri di addestramento: Tipi e valori predefiniti](#).

Mischiare i dati

In Amazon ML, è necessario mischiare i dati, in quanto l'algoritmo SGD è influenzato dall'ordine delle righe nei dati di addestramento. Mischiando i dati di addestramento è possibile ottenere modelli ML migliori, in quanto tale operazione aiuta l'algoritmo SGD a evitare soluzioni ottimali per il primo tipo di dati che incontra, ma non per l'intera gamma di dati. Il mescolamento consente di mischiare l'ordine dei dati in modo che l'algoritmo SGD non incontri un solo tipo di dati per troppe osservazioni in successione. Se si considera solo un tipo di dati per molti aggiornamenti successivi delle ponderazioni, l'algoritmo potrebbe non essere in grado di correggere le ponderazioni del modello per un nuovo tipo di dati perché l'aggiornamento potrebbe essere troppo grande. Inoltre, quando i dati non si presentano in modo casuale, l'algoritmo ha difficoltà a trovare rapidamente la soluzione ottimale per tutti i tipi di dati; in alcuni casi, l'algoritmo potrebbe non trovare mai la soluzione ottimale. Il mescolamento dei dati di addestramento aiuta l'algoritmo a convergere sulla soluzione ottimale in tempi più brevi.

Supponiamo di volere addestrare, ad esempio, un modello ML per prevedere un tipo di prodotto e che i dati di addestramento includano tra i prodotti tipi di film, giocattolo e videogiochi. Se si ordinano i dati in base alla colonna del tipo di prodotto prima di caricare i dati in Amazon S3, l'algoritmo vede i dati alfabeticamente in base al tipo di prodotto. L'algoritmo vede per primi tutti i dati dei film e il modello ML inizia ad apprendere i pattern relativi ai film. Quindi, quando il modello rileva i dati sui giocattoli, ogni aggiornamento che l'algoritmo effettua adatta il modello al tipo di prodotto giocattoli, anche se gli aggiornamenti degradano i pattern che si adattano ai film. Questo improvviso passaggio dal tipo film al tipo giocattoli può produrre un modello che non apprende come prevedere in modo accurato i tipi di prodotto.

Per ulteriori informazioni sul tipo di mescolamento predefinito, consultare [parametri di addestramento: Tipi e valori predefiniti](#).

Regolarizzazione

La regolarizzazione aiuta a evitare l'overfitting dei modelli lineari riguardo agli esempi di dati di addestramento (ovvero memorizzare i pattern invece di generalizzarli) penalizzando i valori di ponderazione estremi. La regolarizzazione L1 ha l'effetto di ridurre il numero di caratteristiche

utilizzate nel modello spingendo a zero le ponderazioni delle caratteristiche che in caso contrario avrebbero piccole ponderazioni. Di conseguenza, la regolarizzazione L1 consente di ottenere modelli di tipo sparse e riduce la quantità di disturbo nel modello. La regolarizzazione L2 comporta valori di ponderazione globale più piccoli e stabilizza le ponderazioni quando vi è un'elevata correlazione tra le caratteristiche di input. È possibile controllare la quantità di regolarizzazione L1 o L2 applicata utilizzando i parametri `Regularization type` e `Regularization amount`. Un valore molto grande di regolarizzazione potrebbe causare l'azzeramento delle ponderazioni di tutte le caratteristiche, impedendo a un modello di apprendere i pattern.

Per maggiori informazioni sui valori di regolarizzazione, consultare [parametri di addestramento: Tipi e valori predefiniti](#).

Valutazione dell'accuratezza del modello

L'obiettivo del modello ML è apprendere i pattern che possono essere generalizzati correttamente per i dati invisibili, invece di memorizzare soltanto i dati che gli sono stati mostrati durante l'addestramento. Una volta che si dispone di un modello, è importante controllare se il modello sia performante sugli esempi non osservati che non sono stati utilizzati per l'addestramento del modello. Per eseguire questa operazione, è possibile utilizzare il modello per prevedere la risposta sul set di dati di valutazione (dati offerti) e quindi confrontare il target previsto con la risposta effettiva (dati acquisiti sul campo).

In ML vengono utilizzati numerosi parametri per misurare l'accuratezza predittiva di un modello. La scelta del parametro dell'accuratezza dipende dall'attività di ML. È importante esaminare questi parametri per decidere se il modello è performante.

Classificazione binaria

L'output effettivo di molti algoritmi di classificazione binaria è un punteggio di previsione. Il punteggio indica la certezza del sistema che una data osservazione appartenga alla classe positiva. Per decidere se l'osservazione debba essere classificata come positiva o negativa, in quanto consumatore di questo punteggio, si interpreterà il punteggio scegliendo una soglia di classificazione (interruzione) e si confronterà il punteggio con tale soglia. Qualsiasi osservazione con punteggi superiori alla soglia viene quindi prevista come classe positiva mentre i punteggi inferiori alla soglia vengono previsti come classe negativa.

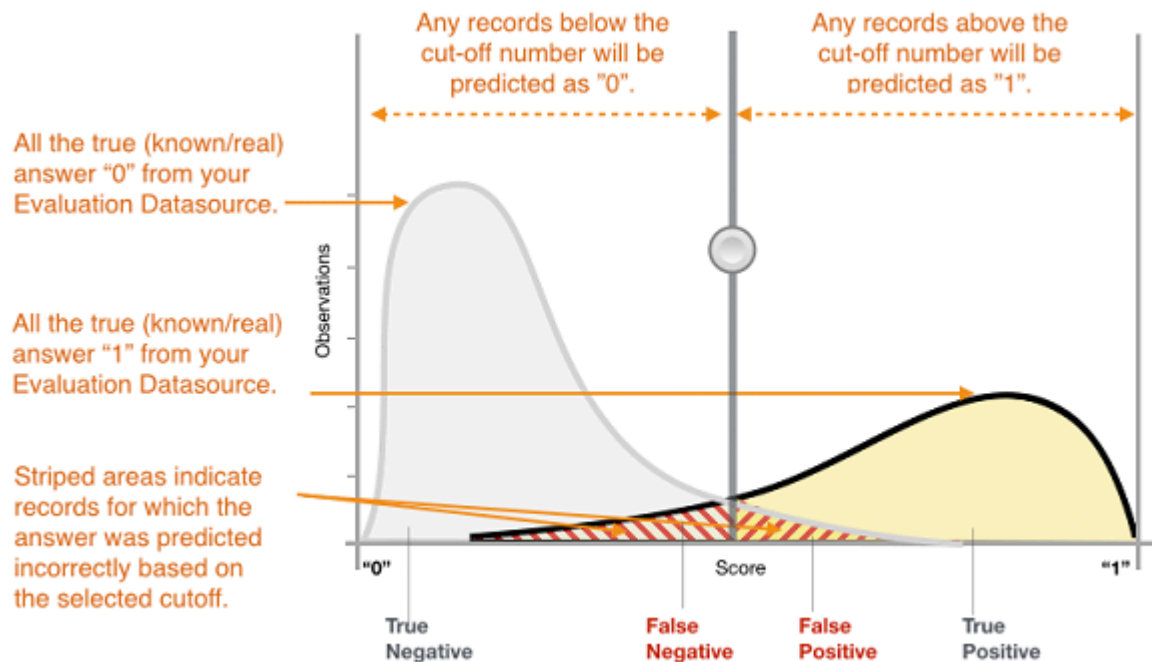


Figura 1: distribuzione dei punteggi per un modello di classificazione binaria

Le previsioni ora rientrano in quattro gruppi basati sulla risposta effettiva nota e sulla risposta prevista: previsioni positive esatte (true positive), previsioni negative esatte (true negative), previsioni positive errate (false positive) e previsioni negative errate (false negative).

Il parametro di accuratezza della classificazione binaria quantifica i due tipi di previsioni corrette e i due tipi di errori. I parametri più comuni sono accuratezza (ACC), precisione, recall, percentuale di falsi positivi, misura F1. Ogni parametro misura un aspetto diverso del modello predittivo. L'accuratezza (ACC) misura la percentuale di previsioni corrette. La precisione misura la percentuale di positivi effettivi tra gli esempi previsti come positivi. Il recall misura quanti positivi effettivi sono stati previsti come positivi. La misura F1 è la media armonica tra precisione e recall.

L'AUC è un tipo diverso di parametro. Misura la capacità del modello ML di prevedere un punteggio più elevato per gli esempi positivi rispetto agli esempi negativi. Poiché l'AUC è indipendente dalla soglia selezionata, è possibile farsi un'idea delle prestazioni in termini di previsioni del modello attraverso il parametro AUC, senza selezionare una soglia.

A seconda del problema aziendale, si potrebbe essere più interessati a un modello che esegua correttamente uno specifico sottoinsieme di questi parametri. Ad esempio, due applicazioni aziendali potrebbe avere requisiti molto diversi per i loro modelli ML:

- Un'applicazione potrebbe essere molto sicura che le previsioni positive siano effettivamente positive (elevata precisione) e potersi permettere di classificare erroneamente alcuni esempi positivi come negativi (recall moderato).
- Un'altra applicazione potrebbe dover prevedere correttamente tutti gli esempi positivi possibile (recall elevato) e accetterà l'errata classificazione di alcuni esempi negativi come positivi (precisione moderata).

In Amazon ML, le osservazioni ottengono un punteggio previsto nell'intervallo $[0, 1]$. La soglia del punteggio per prendere la decisione di classificare gli esempi come 0 o 1 è impostata per impostazione predefinita su 0,5. Amazon ML consente di esaminare le implicazioni legate alla scelta di soglie di punteggio differenti e consente di scegliere una soglia adeguata in grado di soddisfare le esigenze aziendali.

Classificazione multiclasse

A differenza del processo per i problemi di classificazione binaria, non è necessario scegliere un punteggio soglia per fare previsioni. La risposta prevista è la classe (ad esempio, l'etichetta) con il miglior punteggio previsto. In alcuni casi, può essere opportuno utilizzare la risposta prevista solo se il suo punteggio è elevato. In questo caso, è possibile scegliere una soglia per i punteggi previsti in base alla quale si accetta o meno la risposta prevista.

I parametri tipici utilizzati nella multiclasse sono gli stessi utilizzati nel caso della classificazione binaria. Il parametro viene calcolato per ogni classe trattandolo come un problema di classificazione binaria, dopo avere raggruppato tutte le altre classi come appartenenti alla seconda classe. Quindi il parametro binario viene calcolato come media di tutte le classi, per ottenere un parametro di macro media (ogni classe è trattata allo stesso modo) o di media ponderata (ponderazione in base alla frequenza della classe). In Amazon ML, la misura F1 del punteggio macro-medio viene utilizzata per valutare il successo predittivo di un classificatore multiclasse.

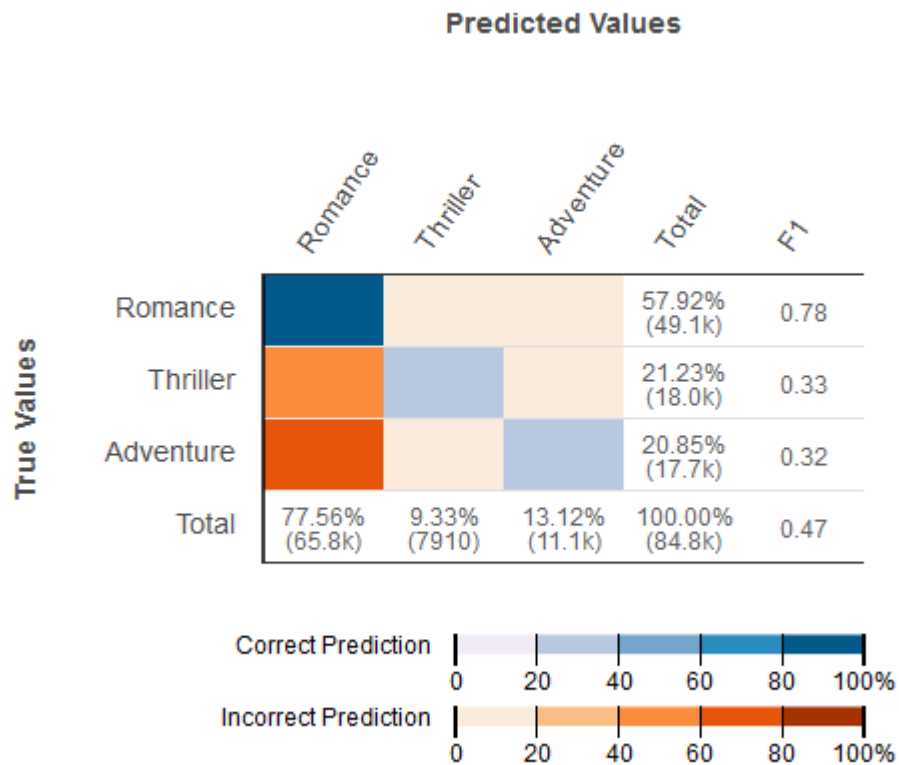


Figura 2: Matrice di confusione per un modello di classificazione multiclasse

È utile esaminare la matrice di confusione per i problemi multiclasse. La matrice di confusione è una tabella che mostra ogni classe dei dati di valutazione e il numero o la percentuale di previsioni esatte e di previsioni errate.

Regressione

Per le attività di regressione, i parametri abituali dell'accuratezza sono la radice dell'errore quadratico medio (RMSE - Root Mean Square Error) e l'errore assoluto medio percentuale (MAPE - Mean Absolute Percentage Error). Questi parametri misurano la distanza tra il target numerico previsto e la risposta numerica effettiva (dati acquisiti sul campo). In Amazon ML, la metrica RMSE viene utilizzata per valutare l'accuratezza predittiva di un modello di regressione.

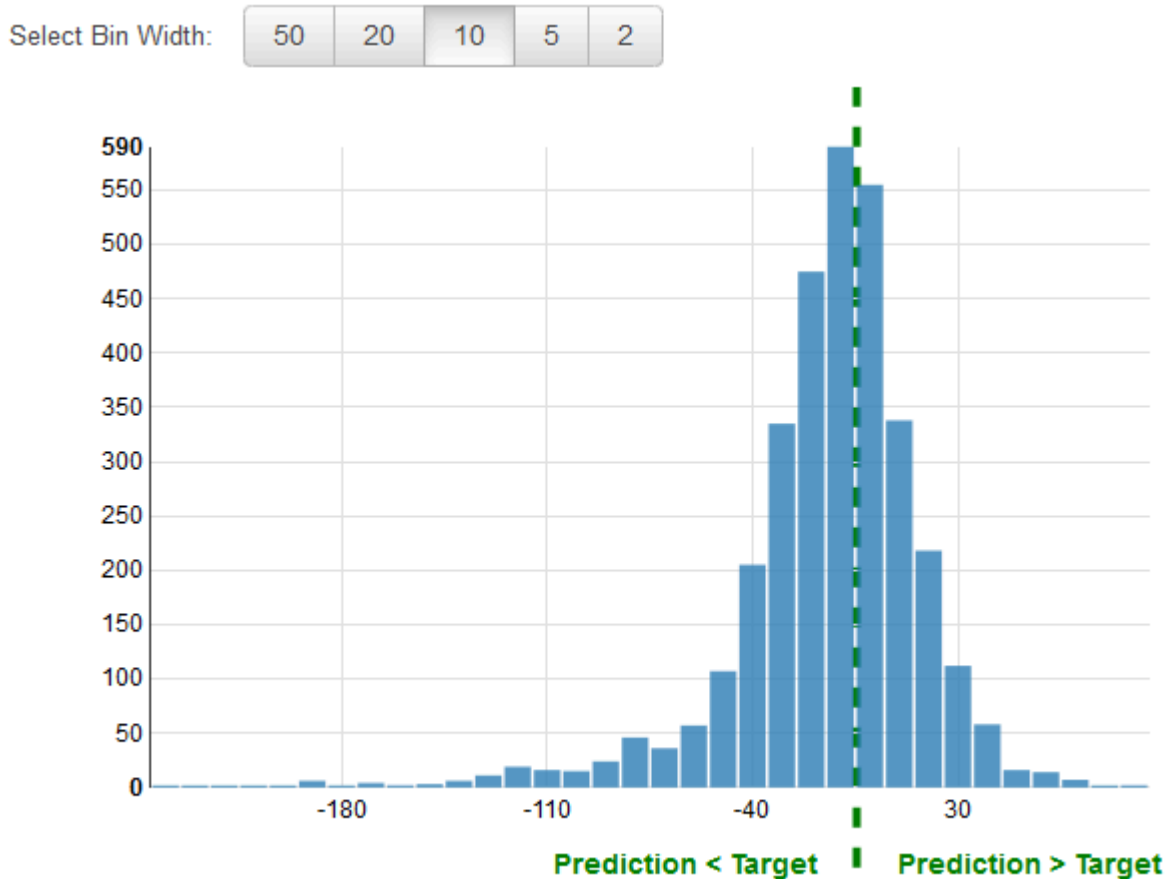


Figura 3: distribuzione dei residui per un modello di regressione

È prassi comune esaminare i residui per cercare problemi di regressione. Un residuo di un'osservazione nei dati di valutazione è la differenza tra il target reale e il target previsto. I residui rappresentano quella parte del target che il modello non è in grado di prevedere. Un residuo positivo indica che il modello sta sottovalutando il target (il target effettivo è più grande del target previsto). Un residuo negativo indica una sopravvalutazione (il target effettivo è più piccolo del target previsto). L'istogramma dei residui relativi ai dati di valutazione, ove distribuito con una forma a campana e centrato sullo zero, indica che il modello compie errori in modo aleatorio e non sovra-prevede o sotto-prevede sistematicamente una determinata gamma di valori target. Se i residui non assumono una forma a campana centrata sullo zero, è presente una struttura nell'errore di previsione del modello. L'aggiunta di ulteriori variabili al modello potrebbero aiutarlo ad acquisire il pattern che non viene acquisito dal modello attuale.

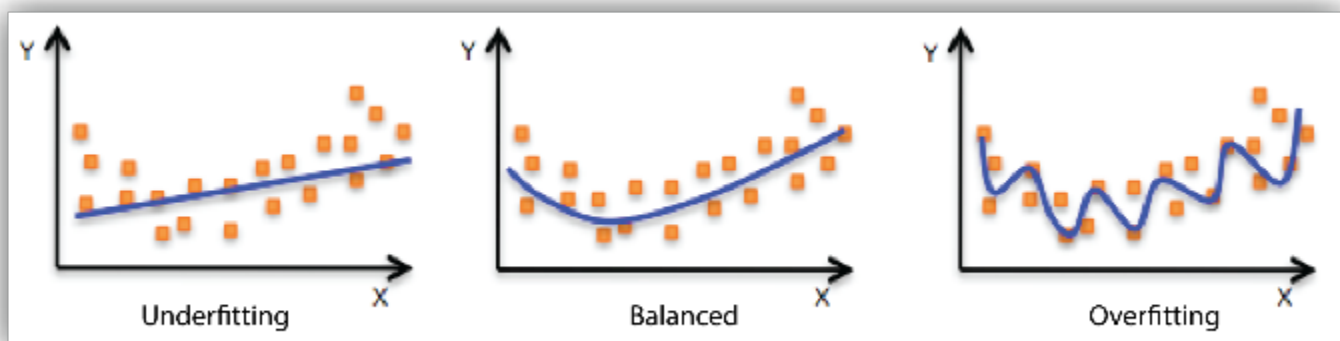
Miglioramento dell'accuratezza del modello

L'ottenimento di un modello ML che soddisfi le proprie esigenze richiede di solito l'iterazione del processo ML e l'esecuzione di tentativi con alcune variazioni. È possibile che nella prima iterazione non si ottenga un modello di previsione altamente predittivo, oppure di potrebbe voler migliorare il modello per ottenere previsioni migliori. Per migliorare le prestazioni, è possibile effettuare l'iterazione di queste fasi:

1. Raccolta dei dati: Si aumenta il numero di esempi di addestramento
2. Elaborazione delle caratteristiche: Aggiunta di variabili e una migliore elaborazione delle caratteristiche
3. Regolazione dei parametri del modello: Considera valori alternativi per i parametri di addestramento utilizzati dall'algorithmo di apprendimento

Fitting del modello: Underfitting vs. overfitting

La comprensione di quale sia il modello adatto è importante per capire la causa principale della scarsa accuratezza del modello. La comprensione di tale aspetto consentirà di trovare misure correttive. È possibile determinare se un modello predittivo è soggetto a underfitting o overfitting dei dati di addestramento esaminando l'errore di previsione sui dati di addestramento e sui dati di valutazione.



Il modello è soggetto a underfitting dei dati di addestramento quando ha prestazioni scarse sui dati di addestramento. Questo avviene perché il modello non è in grado di acquisire il rapporto tra gli esempi di input (spesso chiamati X) e i valori target (spesso chiamati Y). Il modello è soggetto a overfitting dei dati di addestramento quando si vede che il modello di funziona bene con i dati di addestramento, ma

non con i dati di valutazione. Questo avviene perché il modello memorizza i dati che ha visto e non è in grado di generalizzare gli esempi che non ha visto.

Le scarse prestazioni sui dati di addestramento potrebbero essere dovute al fatto che il modello è troppo semplice (le caratteristiche in ingresso non sono abbastanza espressive) per descrivere correttamente il target. È possibile migliorare le prestazioni aumentando la flessibilità del modello. Per aumentare la flessibilità del modello si può provare quanto segue:

- Aggiungere nuove caratteristiche specifiche per il dominio e più prodotti cartesiani delle caratteristiche e modificare il tipo di elaborazione delle caratteristiche utilizzato (ad esempio, aumentando la dimensione degli n-grammi)
- Diminuire la quantità di regolarizzazione utilizzata

Se il modello effettua l'overfitting dei dati di addestramento, è opportuno intervenire per ridurre la flessibilità del modello. Per ridurre la flessibilità del modello si può provare quanto segue:

- Selezione delle caratteristiche: si può considerare l'utilizzo di un numero inferiore di combinazioni delle caratteristiche, la diminuzione della dimensione degli n-grammi e la riduzione del numero di bin di attributi numerici.
- Aumentare la quantità di regolarizzazione utilizzata.

L'accuratezza sui dati di addestramento e sui dati di prova potrebbe essere scarsa perché l'algoritmo di apprendimento non ha avuto a disposizione abbastanza dati da cui apprendere. È possibile migliorare le prestazioni nel seguente modo:

- Aumentare la quantità di esempi di addestramento.
- Aumentare il numero di passate sui dati di addestramento esistenti.

Utilizzo del modello per effettuare previsioni

Ora che si dispone di un modello ML che funziona correttamente, lo si può utilizzare per effettuare previsioni. In Amazon Machine Learning, vi sono due modi per utilizzare un modello per effettuare previsioni:

Previsioni in batch

La previsione in batch è utile quando si desidera generare previsioni per un insieme di osservazioni in una volta sola e si vuole intervenire successivamente su una determinata percentuale o su un determinato numero di osservazioni. Di solito, non si dispone di un requisito di bassa latenza per tale applicazione. Ad esempio, se si vuole decidere quali clienti dovranno essere i destinatari di una campagna pubblicitaria per un prodotto, si possono ottenere i punteggi di previsione per tutti i clienti, per poi ordinare le previsioni del modello al fine di identificare i clienti che hanno maggiore probabilità di acquistare, utilizzando come target il primo 5% dei clienti che più probabilmente acquisteranno.

Previsioni online

Gli scenari di previsione online sono per i casi in cui si vogliono generare previsioni, una a una, per ogni esempio, in modo indipendente rispetto agli altri esempi, in un ambiente a bassa latenza. Ad esempio, è possibile utilizzare le previsioni per decidere subito se una determinata transazione potrebbe essere una delle transazioni fraudolente.

Riqualficazione dei modelli sui nuovi dati

Affinché un modello effettui previsioni accurate, i dati su cui effettua le previsioni devono avere una distribuzione simile a quella dei dati su cui il modello è stato addestrato. Poiché le distribuzioni dei dati possono cambiare nel tempo, la distribuzione di un modello non è un evento una tantum, ma un processo continuo. È buona norma monitorare costantemente i dati in entrata e riqualficare il modello su dati più recenti se si accerta che la distribuzione dei dati si è discostata in modo significativo dall'originale distribuzione dei dati di addestramento. Se i dati di monitoraggio per rilevare una modifica nella distribuzione dei dati hanno un overhead elevato, una strategia più semplice consiste nell'addestrare il modello periodicamente, ad esempio con cadenza giornaliera, settimanale o mensile. Per riqualficare i modelli in Amazon ML, è necessario creare un nuovo modello basato sui nuovi dati di addestramento.

Il processo Amazon Machine Learning

La tabella seguente descrive come utilizzare la console di Amazon ML per eseguire il processo ML illustrato in questo documento.

Processo ML	Attività di Amazon ML
Analisi dei dati	Per analizzare i dati in Amazon ML, creare un'origine dati e rivedere la pagina delle informazioni sui dati.
Divisione dei dati in origini dati di addestramento e di valutazione	<p>Amazon ML è in grado di dividere l'origine dati per utilizzare il 70% dei dati per l'addestramento del modello e il 30% per la valutazione delle prestazioni predittive del modello.</p> <p>Quando si utilizza la procedura guidata Crea un modello ML con le impostazioni predefinite, Amazon ML divide i dati.</p> <p>Se si utilizza la procedura guidata Crea un modello ML con le impostazioni personalizzate e si sceglie di valutare il modello ML, verrà visualizzata un'opzione per consentire ad Amazon ML di frazionare i dati ed eseguire una valutazione sul 30% dei dati.</p>
Mischiare i dati di addestramento	Quando si utilizza la procedura guidata Crea un modello ML con le impostazioni predefinite, Amazon ML meschia i dati. È anche possibile mischiare i dati prima di importarli in Amazon ML.
Elaborazione delle caratteristiche	<p>Il processo di unire i dati di addestramento in un formato ottimale per l'apprendimento e la generalizzazione è noto come trasformazione delle caratteristiche. Quando si utilizza la procedura guidata Crea un modello ML con le impostazioni predefinite, Amazon ML suggerisce le impostazioni per l'elaborazione delle caratteristiche per i dati.</p> <p>Per specificare le impostazioni di elaborazione delle caratteristiche, è possibile usare l'opzione Custom (Personalizza) della procedura guidata Crea un modello ML e fornire una composizione per l'elaborazione delle caratteristiche.</p>
Addestramento del modello	Quando si utilizza la procedura guidata Crea un modello ML per creare un modello in Amazon ML, Amazon ML addestra il modello.
Selezione dei parametri del modello	In Amazon ML, è possibile ottimizzare quattro parametri che influenzano le prestazioni predittive del modello: la dimensione del modello, il numero di passate, il tipo di mescolamento (shuffling) e la regolarizzazione. È possibile impostare questi parametri quando si utilizza la

Processo ML	Attività di Amazon ML
	procedura guidata Crea un modello ML per creare un modello ML e scegliere l'opzione Custom (Personalizzato).
Valutazione delle prestazioni del modello	Utilizzare la procedura guidata Crea valutazione per valutare le prestazioni predittive del modello.
Selezione delle caratteristiche	L'algoritmo di apprendimento di Amazon ML è in grado di eliminare le caratteristiche che contribuiscono molto al processo di apprendimento. Per indicare che si desidera eliminare queste caratteristiche, scegliere il parametro <code>L1 regularization</code> quando si crea il modello ML.
Impostazione di un punteggio soglia per l'accuratezza predittiva	Rivedere le prestazioni predittive del modello nella relazione di valutazione in corrispondenza di diverse soglie di punteggio, quindi impostare il punteggio soglia in base alle applicazioni aziendali. Il punteggio soglia determina il modo in cui il modello definisce una corrispondenza predittiva. Modificare il numero per controllare i falsi positivi e i falsi negativi.
Uso del modello	Utilizzare il modello per ottenere previsioni per un batch di osservazioni utilizzando la procedura guidata Crea previsione in batch. In alternativa, si possono ottenere previsioni per singole osservazioni on demand abilitando il modello ML a elaborare previsioni in tempo reale utilizzando l'API <code>Predict</code> .

Impostazione di Amazon Machine Learning

È necessario disporre di un account AWS prima di poter utilizzare Amazon Machine Learning per la prima volta. Se non si dispone di un account, consultare Registrati ad AWS.

Registrati ad AWS

Quando si effettua la registrazione ad Amazon Web Services (AWS), l'account AWS viene automaticamente registrato per tutti i servizi AWS, tra cui Amazon ML. Ti vengono addebitati solo i servizi che utilizzi. Se si dispone di un account AWS, saltare questa fase. Se non si dispone di un account AWS, utilizzare la seguente procedura per crearne uno.

Registrazione per creare un account AWS

1. Visitare il sito <http://aws.amazon.com> e scegliere Sign Up (Registrati).
2. Seguire le istruzioni su schermo.

Come parte della procedura di registrazione si riceverà una telefonata, durante la quale si dovrà inserire un PIN usando la tastiera del telefono.

Tutorial: Utilizzare Amazon ML per prevedere le risposte a un'offerta di marketing

Con Amazon Machine Learning (Amazon ML), è possibile creare e addestrare modelli predittivi e ospitare le applicazioni in una soluzione cloud scalabile. In questo tutorial, illustreremo come utilizzare la console di Amazon ML per creare un'origine dati, creare un modello di machine learning (ML) e utilizzare il modello per generare previsioni utilizzabili nelle applicazioni.

L'esercitazione di esempio illustra come identificare i potenziali clienti per una campagna di marketing mirata, ma è possibile applicare gli stessi principi per creare e utilizzare una vasta gamma di modelli ML. Per completare l'esercitazione di esempio si utilizzeranno i set di dati di banking e marketing pubblici della [University of California a Irvine \(UCI\) Machine Learning Repository](#). Questi set di dati contengono dati generali sui clienti e informazioni su come hanno reagito ai precedenti contatti di marketing. Si possono utilizzare questi dati per identificare quali clienti hanno maggiori probabilità di effettuare una sottoscrizione al nuovo prodotto, un deposito bancario a termine, noto anche come certificato di deposito (CD).

Warning

Questo tutorial non è incluso nel piano gratuito AWS. Per ulteriori informazioni sui prezzi di Amazon ML, consulta [Prezzi di Amazon Machine Learning](#).

Prerequisito

Per eseguire il tutorial, è necessario disporre di un account AWS. Se non si dispone di un account AWS, consultare [Impostazione di Amazon Machine Learning](#).

Fasi

- [Fase 1: Preparazione dei dati](#)
- [Fase 2: Creare un'origine dati di addestramento](#)
- [Fase 3: Creare un modello ML](#)
- [Fase 4: Esaminare le prestazioni predittive del modello ML e impostare un punteggio soglia](#)
- [Fase 5: Utilizzare il modello ML per generare previsioni](#)

- [Fase 6: Eliminare](#)

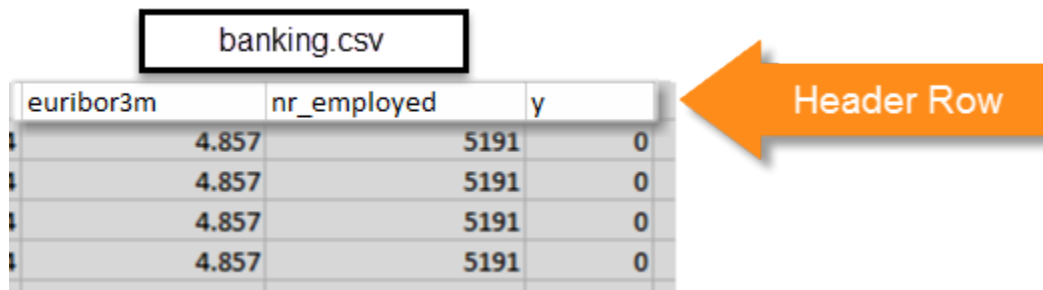
Fase 1: Preparazione dei dati

Nel machine learning, in genere è possibile ottenere i dati e assicurarsi che siano correttamente formattati prima di avviare il processo di addestramento. Ai fini del presente tutorial, abbiamo ottenuto un set di dati di esempio dal [Repository UCI Machine Learning](#), formattandolo conformemente alle linee guida di Amazon ML e rendendolo disponibile per il download. Scaricare il set di dati dal percorso di storage Amazon Simple Storage Service (Amazon S3) e caricarlo nel proprio bucket S3 seguendo le procedure indicate di seguito.

Per i requisiti di formattazione Amazon ML, consultare [Comprendere il formato dei dati per Amazon ML](#).

Per scaricare i set di dati

1. Scaricare il file che contiene i dati cronologici dei clienti che hanno acquistato prodotti simili al deposito bancario a termine facendo clic su [banking.zip](#). Decomprimere la cartella e salvare il file banking.csv sul computer.
2. Scarica il file che userai per prevedere se i potenziali clienti risponderanno alla tua offerta facendo clic su [banking-batch.zip](#). Decomprimere la cartella e salvare il file banking-batch.csv sul computer.
3. Aprire banking.csv. Verranno visualizzate righe e colonne di dati. La riga di intestazione contiene i nomi degli attributi di ogni colonna. Un attributo è una proprietà denominata in modo univoco che descrive una determinata caratteristica di ciascun cliente; ad esempio, nr_employed indica lo stato lavorativo del cliente. Ogni riga rappresenta la raccolta delle osservazioni relative a un singolo cliente.



euribor3m	nr_employed	y
4.857	5191	0
4.857	5191	0
4.857	5191	0
4.857	5191	0

Si vuole che il modello ML risponda alla domanda "Il cliente effettuerà la sottoscrizione al mio nuovo prodotto?". Nel set di dati banking.csv, la risposta a questa domanda è l'attributo y,

che contiene i valori 1 (per sì) o 0 (per no). L'attributo che si vuole che Amazon ML apprenda a prevedere è noto come attributo di destinazione.

Note

L'attributo y è un attributo binario. Può contenere solo uno di due valori, in questo caso 0 o 1. Nel set di dati UCI originale, l'attributo y è Sì o No. Abbiamo modificato il set di dati originale per l'utente. Tutti i valori dell'attributo y che significano sì sono ora 1 e tutti i valori che significano no sono ora 0. Se si utilizzano propri dati, è possibile impiegare altri valori per un attributo binario. Per ulteriori informazioni sui valori validi, consultare [Utilizzo del campo AttributeType](#).

I seguenti esempi mostrano i dati prima e dopo la modifica dei valori dell'attributo y in attributi binari 0 e 1.

Before transformation

banking.csv			Target
euribor3m	nr_employed	y	
4.857	5191	no	
4.857	5191	no	
4.857	5191	yes	
4.857	5191	yes	
4.857	5191	no	

After transformation

banking.csv			Target
euribor3m	nr_employed	y	
4.857	5191	0	
4.857	5191	0	
4.857	5191	1	
4.857	5191	1	
4.857	5191	0	

Il file `banking-batch.csv` non contiene l'attributo `y`. Dopo aver creato un modello ML, si utilizzerà il modello per prevedere `y` per ciascun record di quel file.

Quindi, caricare `ilbanking.csv` e `banking-batch.csv` in Amazon S3.

Caricare i file in un percorso Amazon S3

1. Accedi alla AWS Management Console e apri la console di Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Nell'elenco All Buckets (Tutti i bucket), creare un bucket o scegliere il percorso in cui si vogliono caricare i file.
3. Nella barra di navigazione, scegliere Upload (Carica).
4. Seleziona Aggiungi file.
5. Nella finestra di dialogo, passare al desktop, scegliere `banking.csv` e `banking-batch.csv`, quindi Open (Apri).

Ora è possibile [creare la propria origine dati di addestramento](#).

Fase 2: Creare un'origine dati di addestramento

Dopo aver caricato il file `banking.csv` Set di dati nella propria posizione Amazon Simple Storage Service (Amazon S3), lo si può utilizzare per creare un'origine dati di addestramento. Un'origine dati è un oggetto Amazon Machine Learning (Amazon ML) che contiene il percorso dei dati di input e importanti metadata riguardo ai dati di input. Amazon ML utilizza l'origine dati per operazioni quali l'addestramento e la valutazione del modello ML.

Per creare un'origine dati, è necessario fornire quanto segue:

- La posizione di Amazon S3 dei dati e l'autorizzazione per accedere ai dati
- Lo schema, che include i nomi degli attributi dei dati e il tipo di ogni attributo (Numeric, Text, Categorical o Binary)
- Il nome dell'attributo che contiene la risposta che si desidera che Amazon ML impari a prevedere, l'attributo di destinazione

Note

Nell'origine dati non vengono memorizzati i dati, ma solo i riferimenti ad essi. Evitare di spostare o modificare i file archiviati in Amazon S3. Se si spostano o modificano tali file, Amazon ML non riesce ad accedervi per creare un modello ML, generare le valutazioni o generare le previsioni.

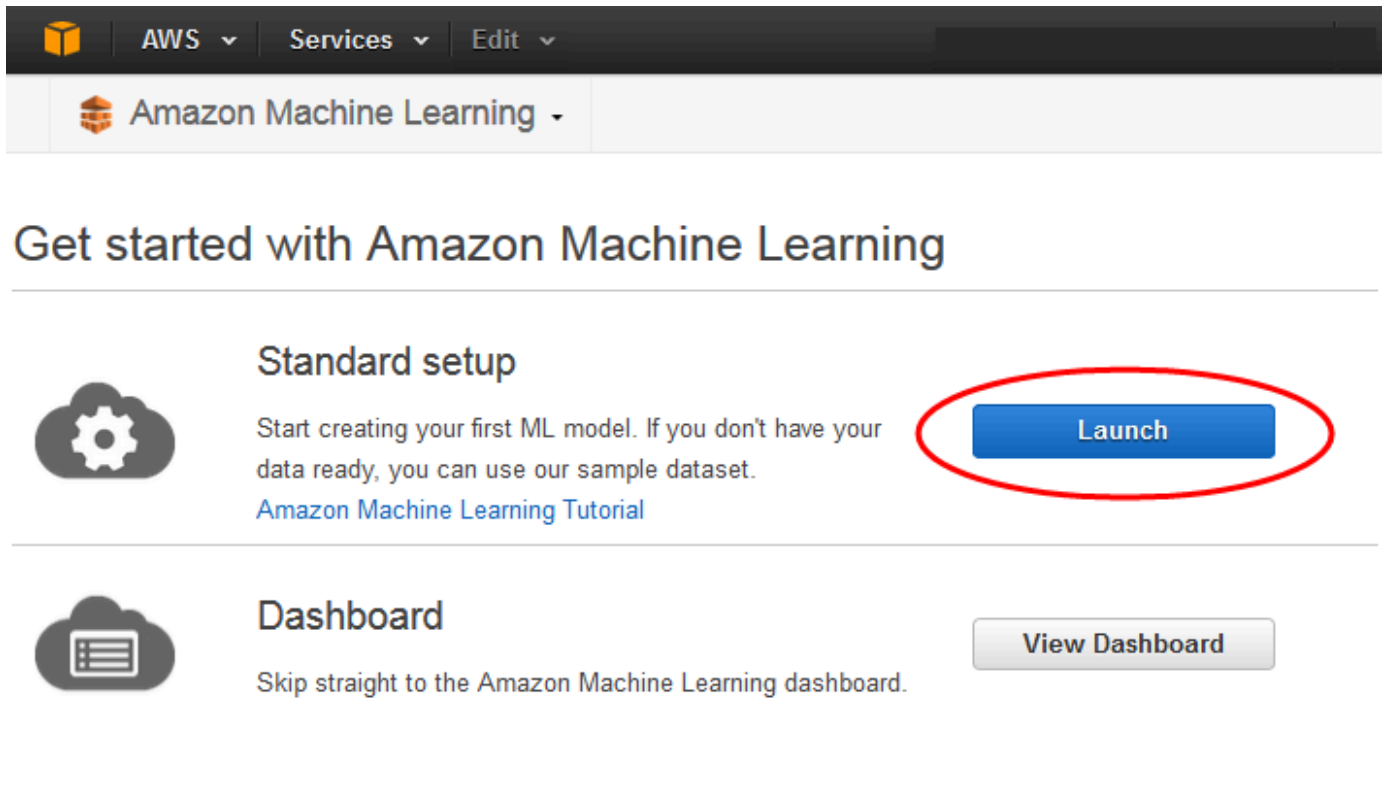
Per creare un'origine dati di addestramento

1. Aprire la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Scegliere Get Started (Inizia).

Note

Questo tutorial presuppone che si tratti del primo utilizzo di Amazon ML. Se hai già utilizzato Amazon ML, puoi usare il Creazione di un nuovo... Un elenco a discesa nel pannello di controllo Amazon ML per creare una nuova origine dati.

3. Sull'inizio ad usare Amazon Machine Learning pagina, scegliere Avvio di.



Standard setup

Start creating your first ML model. If you don't have your data ready, you can use our sample dataset.
[Amazon Machine Learning Tutorial](#)

Launch

Dashboard

Skip straight to the Amazon Machine Learning dashboard.

View Dashboard

4. Nella pagina Input Data (Dati di input), per Where is your data located? (Dove si trovano i tuoi dati?), assicurarsi che sia stato selezionato S3.


Where is your data located? S3 Redshift

5. Per Percorso S3, digitare la posizione completa del `banking.csv` dal passaggio 1: Preparazione dei dati. Ad esempio: `bucket/banking.csv`. Amazon ML anteduce ad `s3://al` nome del bucket per l'utente.
6. Per Datasource name (Nome origine dati), digitare **Banking Data 1**.

S3 location *

s3:// aml-sample-data/banking.csv

Enter the path to a single file or folder in Amazon S3. You need to grant Amazon ML permission to read this data. [Learn more](#).

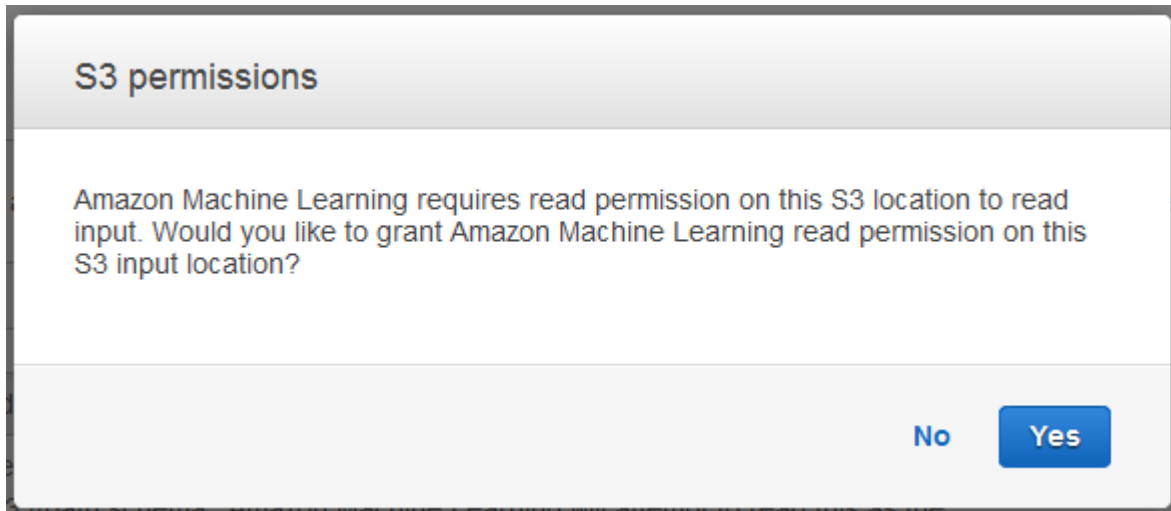
If you already have a schema for this data, provide it in a file at `s3://<path-of-input-data>.schema`. If you don't have a schema, Amazon ML will help you create one on the next page. 

Datasource name

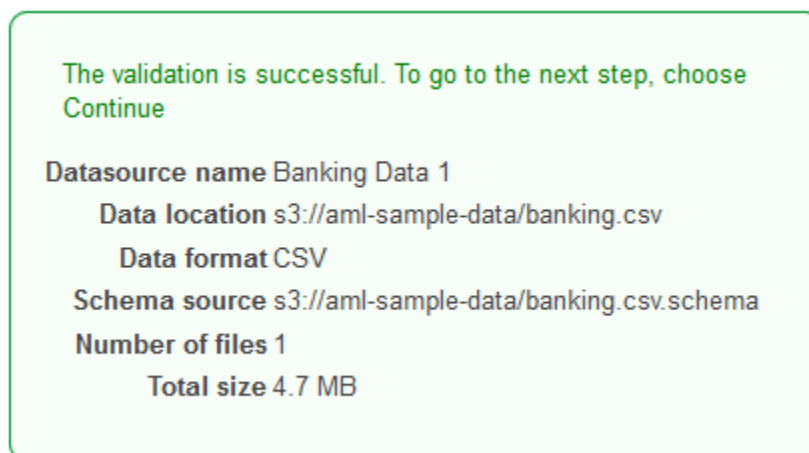
Banking Data 1

7. Selezionare Verify (Verifica).

8. Nella finestra di dialogo S3 permissions (Autorizzazioni S3), scegliere Yes (Sì).



9. Se Amazon ML riesce ad accedere e a leggere il file di dati nella posizione S3, verrà visualizzata una pagina simile alla seguente. Verificare le proprietà e scegliere Continue (Continua).



Dopo occorre stabilire uno schema. UNschema è costituito dalle informazioni che servono ad Amazon ML per interpretare i dati di input per un modello ML, inclusi i nomi degli attributi speciali. Sono disponibili due modi per fornire uno schema ad Amazon ML:

- Fornire un file di schema separato quando si caricano i dati Amazon S3.
- Consentire ad Amazon ML di dedurre i tipi di attributo e di creare uno schema per l'utente.

In questo tutorial, chiederemo ad Amazon ML di dedurre lo schema.

Per informazioni sulla creazione di un file di schema separato, consultare [Creazione di uno schema di dati per Amazon ML](#).

Per consentire ad Amazon ML di dedurre lo schema

1. SulSchema, Amazon ML mostra lo schema che ha dedotto. Esaminare i tipi di dati che Amazon ML ha dedotto per gli attributi. È importante che agli attributi sia assegnato il tipo di dati corretto, per aiutare Amazon ML ad acquisire correttamente i dati e abilitare la corretta elaborazione delle caratteristiche sugli attributi.
 - Gli attributi che hanno solo due stati possibili, come ad esempio sì o no, devono essere contrassegnati come Binary (Binario).
 - Gli attributi che sono numeri o stringhe utilizzati per denotare una categoria devono essere contrassegnati come Categorical (Categorico).
 - Gli attributi che sono quantità numeriche per le quali l'ordine è significativo devono essere contrassegnati come Numeric (Numerico).
 - Gli attributi che sono stringhe che si desidera trattare come parole delimitate da spazi devono essere contrassegnati come Text (Testo).

<input type="checkbox"/>	Name	Data Type	Sample Field Value 1
<input type="checkbox"/>	age	Numeric ▼	56
<input type="checkbox"/>	campaign	Numeric ▼	1
<input type="checkbox"/>	cons_conf_idx	Numeric ▼	-36.4
<input type="checkbox"/>	cons_price_idx	Numeric ▼	93.994
<input type="checkbox"/>	contact	Categorical ▼	telephone
<input type="checkbox"/>	day_of_week	Categorical ▼	mon
<input type="checkbox"/>	default	Categorical ▼	no
<input type="checkbox"/>	duration	Numeric ▼	261
<input type="checkbox"/>	education	Categorical ▼	basic.4y
<input type="checkbox"/>	emp_var_rate	Numeric ▼	1.1

- In questo tutorial, Amazon ML ha identificato correttamente i tipi di dati per tutti gli attributi, quindi scegliere Continua.

Quindi, selezionare un attributo di destinazione.

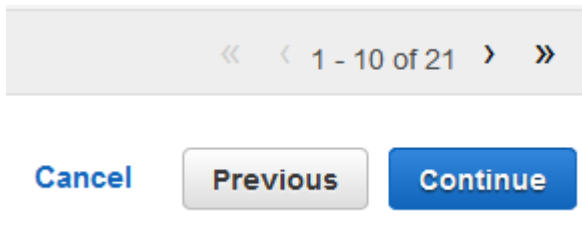
Ricordare che la destinazione è l'attributo che il modello ML deve imparare a prevedere. L'attributo y indica se un singolo ha aderito a una campagna in passato: 1 (sì) o 0 (no).

Note

Scegliere un attributo di destinazione solo se si utilizza l'origine dati per l'addestramento e la valutazione di modelli ML.

Per selezionare y come attributo di destinazione

- In basso a destra nella tabella, scegliere la freccia singola per passare all'ultima pagina della tabella, dove è visualizzato l'attributo denominato y.



- Nella colonna Target (Destinazione), selezionare y.



Amazon ML conferma che y è selezionato come destinazione.

- Scegliere Continue (Continua).

4. Nella pagina Row ID (ID riga) per Does your data contain an identifier? (I dati contengono un identificatore?), assicurarsi che sia selezionata l'impostazione predefinita No.
5. Selezionare Review (Rivedi), quindi Continue (Continua).

Ora che si dispone di un'origine dati di addestramento, è possibile [creare il proprio modello](#).

Fase 3: Creare un modello ML

Dopo aver creato l'origine dati di addestramento, la si utilizza per creare un modello ML, addestrare il modello e quindi valutare i risultati. Il modello ML è una raccolta di modelli che Amazon ML individua nei dati durante l'addestramento. È possibile utilizzare il modello per creare previsioni.

Creazione di un modello ML

1. Poiché la procedura guidata per le attività iniziali crea sia un'origine dati di addestramento sia un modello, Amazon Machine Learning (Amazon ML), utilizza automaticamente l'origine dati di addestramento appena creata e fa accedere direttamente al programmaImpostazioni del modello ML(Certificato creato). Nella pagina ML model settings (Impostazioni modello ML) per ML model name (Nome modello ML), occorre verificare che sia visualizzato il modello ML predefinito **ML model: Banking Data 1**.


L'utilizzo di un nome descrittivo, come quello dell'impostazione predefinita, consente di identificare e gestire facilmente il modello ML.

2. Per Training and evaluation settings (Impostazioni di addestramento e valutazione), accertarsi che sia stato selezionata l'opzione Default.

Select training and evaluation settings

Recipes and training parameters control the ML model training process. You can select these settings for your ML model or use the defaults provided by Amazon ML. In either case, you can choose to have Amazon ML reserve a portion of the input data for evaluation. [Learn more.](#)

Default (Recommended)

Choose this option if you want to use Amazon ML's recommended recipe, training parameters, and evaluation settings. 

Name this evaluation (Optional)

Evaluation: ML model: Banking Data 1

3. Per Name this evaluation (Denomina questa valutazione) accettare l'impostazione predefinita **Evaluation: ML model: Banking Data 1**.
4. Scegliere Review (Rivedi), rivedere le impostazioni e scegliere Finish (Fine).

Dopo aver scelto Termina, Amazon ML aggiunge il modello alla coda di elaborazione. Quando Amazon ML crea il modello, applica le impostazioni predefinite ed esegue le operazioni seguenti:

- Divide l'origine dati per l'addestramento in due sezioni, una che contiene il 70% dei dati e l'altra che contiene il restante 30%
- Forma il modello ML nella sezione che contiene il 70% dei dati di input
- Valuta il modello utilizzando il restante 30% dei dati di input

Mentre il modello è in coda, Amazon ML comunica lo stato come Pending (In attesa). Mentre Amazon ML crea il modello, comunica lo stato come In Progress (In corso). Quando tutte le operazioni sono state completate, comunica lo stato come Completed (Completato). Attendere il completamento della valutazione prima di continuare.

Ora è possibile [esaminare le prestazioni del modello e impostare un punteggio limite](#).

Per ulteriori informazioni sull'addestramento e la valutazione dei modelli, vedere [Addestramento dei modelli ML](#) e [evaluate an ML model](#).

Fase 4: Esaminare le prestazioni predittive del modello ML e impostare un punteggio soglia

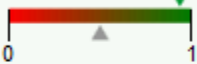
Ora che è stato creato il modello ML e Amazon Machine Learning (Amazon ML) lo ha valutato, è necessario capire se è idoneo all'uso. Durante la valutazione, Amazon ML ha calcolato un parametro di qualità standard del settore, chiamato Area Under a Curve ovvero area sotto la curva (AUC), che esprime la qualità delle prestazioni del modello ML. Amazon ML, inoltre, interpreta il parametro AUC per determinare se la qualità del modello ML sia adeguata alla maggior parte delle applicazioni di machine learning. (Ulteriori informazioni su AUC sono disponibili in [Misurazione dell'accuratezza del modello ML](#)). Rivediamo il parametro AUC e quindi regoliamo il punteggio soglia o limite per ottimizzare le prestazioni predittive del modello.

Esame del parametro AUC per il modello ML

1. Sul Riepilogo del modello ML pagina, nel Rapporto del modello ML riquadro di navigazione, scegliere Valutazioni, scegli Valutazione: modello di ML Il modello bancario 1 e quindi scegliere Riepilogo.
2. Nella pagina Evaluation summary (Riepilogo della valutazione) esaminare il riepilogo della valutazione, compreso il parametro AUC delle prestazioni del modello.

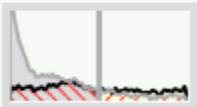
ML model performance metric

On your most recent evaluation, **ev-3fF6uP2W5VL**, the ML model's quality score is considered **extremely good** for most machine learning applications. ⓘ



AUC: 0.94
Baseline AUC: 0.50
Difference: 0.44

Next step: If you want to use this ML model to generate predictions, explore trade-offs to optimize the performance of your ML model first. ⓘ



Score threshold: 0.5



[Adjust score threshold](#)

Il modello ML genera punteggi di previsione numerici per ogni record in un'origine dati di previsione e quindi applica una soglia per convertire questi punteggi in etichette binarie di 0 (per no) o 1 (per sì). Modificando il punteggio soglia, è possibile regolare il modo in cui il modello ML assegna queste etichette. Ora passiamo all'impostazione del punteggio soglia.

Per impostare un punteggio soglia per il modello ML

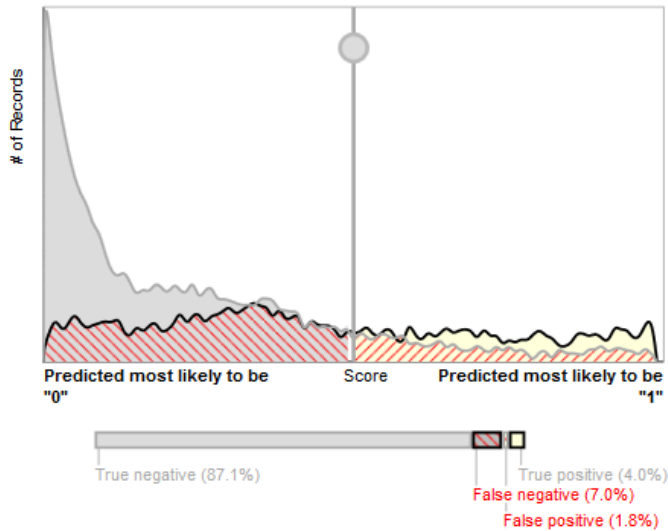
1. Nella pagina Evaluation Summary (Riepilogo della valutazione), scegliere Adjust Score Threshold (Adeguo punteggio soglia).

ML model performance

This chart shows the distributions of your predicted answers for the actual "1" and "0" records in your evaluation data. Any overlap of the actual "1"  & "0"  is where your ML model guesses wrong. [Learn more](#).

Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.

[Explain this chart](#)



Trade-off based on score threshold

[Reset score threshold \(0.5\)](#)

- **91% are correct**
500 true positive
10,766 true negative
- **9% are errors**
226 false positive
863 false negative

- 6% of the records are predicted as "1"
- 94% of the records are predicted as "0"

[Save score threshold at 0.50](#)

Advanced metrics



Accuracy 0.9119	0	<input type="range"/>	1
False positive rate 0.0206	0	<input type="range"/>	1
Precision 0.6887	0	<input type="range"/>	1
Recall 0.3668	0	<input type="range"/>	1

È possibile ottimizzare i parametri delle prestazioni del modello ML regolando il punteggio soglia. La regolazione di questo valore cambia il livello di fiducia che il modello deve avere in una previsione prima di ritenere che la previsione sia positiva. Inoltre, modifica il numero di falsi negativi e falsi positivi che si è disposti a tollerare nelle previsioni.

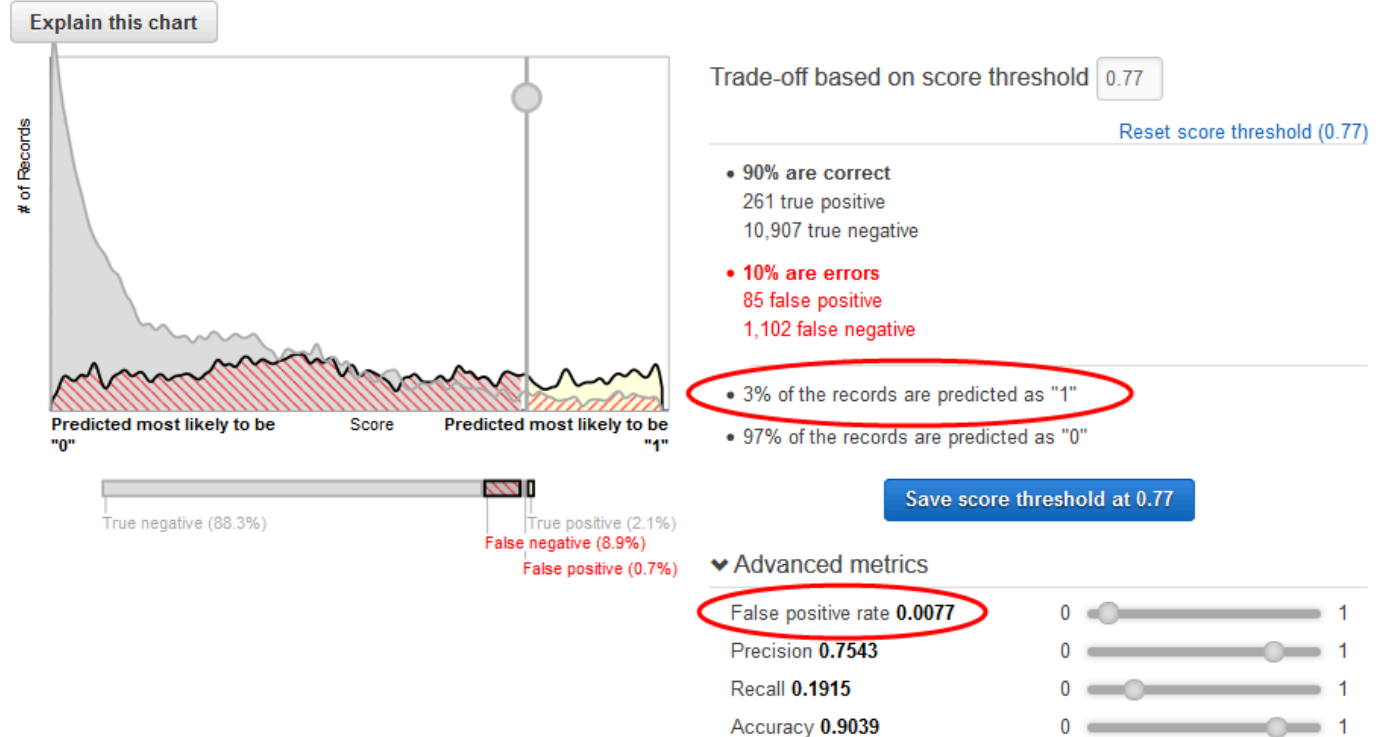
È possibile controllare il livello limite di ciò che il modello considera una previsione positiva aumentando il punteggio soglia fino a considerare positive solo le previsioni con la massima probabilità di essere tali. È anche possibile ridurre il punteggio soglia finché non avrà più falsi negativi. Si può scegliere il proprio cutoff in base alle esigenze aziendali. Ai fini di questo tutorial, ogni falso positivo comporta costi per la campagna, perciò vogliamo un rapporto elevato di veri positivi rispetto ai falsi positivi.

2. Supponiamo di avere come target il primo 3% dei clienti che effettueranno la sottoscrizione al prodotto. Far scorrere il selettore verticale per impostare il punteggio soglia su un valore che corrisponda a 3% of the records are predicted as "1" (3% dei record previsti come "1").

ML model performance

This chart shows the distributions of your predicted answers for the actual "1" and "0" records in your evaluation data. Any overlap of the actual "1"  & "0"  is where your ML model guesses wrong. [Learn more](#).

Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.



Si noti l'impatto di questo punteggio soglia sulle prestazioni del modello ML: la percentuale di falsi positivi è 0,007. Supponiamo che tale percentuale di falsi positivi sia accettabile.

3. Scegliere [Save score threshold at 0.77](#) (Salva punteggio soglia a 0,77).

Ogni volta che si utilizzerà questo modello ML per fare previsioni, il modello sarà in grado di prevedere i record con punteggi superiori a 0,77 come "1" e il resto dei record come "0".

Per ulteriori informazioni sul punteggio soglia, consultare [Classificazione binaria](#).

Ora è possibile [creare previsioni utilizzando il modello](#).

Fase 5: Utilizzare il modello ML per generare previsioni

Amazon Machine Learning (Amazon ML) è in grado di generare due tipi di previsioni: in batch e in tempo reale.

UNprevisione in tempo reale è una previsione per una singola osservazione che Amazon ML genera on demand. Le previsioni in tempo reale sono ideali per applicazioni per dispositivi mobili, siti Web e altre applicazioni che devono utilizzare i risultati in modo interattivo.

UNprevisione in batch è un set di previsioni per un gruppo di osservazioni. Amazon ML elabora i record di una previsione in batch insieme, perciò l'elaborazione potrebbe richiedere un po' di tempo. Usare le previsioni in batch per le applicazioni che richiedono previsioni per un set di osservazioni o previsioni che non utilizzano i risultati in modo interattivo.

Per questo tutorial, verrà generata una previsione in tempo reale per prevedere se un potenziale cliente effettuerà la sottoscrizione al nuovo prodotto. Verranno inoltre generate previsioni per un batch di grandi dimensioni di potenziali clienti. Per la previsione in batch, si utilizzerà il file `banking-batch.csv` che è stato caricato in [Fase 1: Preparazione dei dati](#).

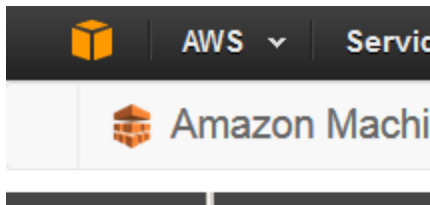
Iniziamo con una previsione in tempo reale.

Note

Per le applicazioni che richiedono previsioni in tempo reale, è necessario creare un endpoint in tempo reale per il modello ML. Durante la disponibilità di un endpoint in tempo reale sono previsti addebiti di costi. Prima di decidere di utilizzare le previsioni in tempo reale e iniziare a sostenere i costi associati, è possibile provare a usare la funzione di previsione in tempo reale in un browser Web, senza creare un endpoint in tempo reale. È quanto faremo per questo tutorial.

Per provare una previsione in tempo reale

1. Nel riquadro di navigazione ML model report (Report del modello ML), scegliere Try real-time predictions (Prova le previsioni in tempo reale).



ML model report

Summary

Settings

Monitoring

Tools

Try real-time predictions

- Scegliere Paste a record (Incolla un record).

Try real-time predictions

Try generating real-time predictions for free using the web browser on this page. To request a real-time prediction, complete the following form or provide a single data record in CSV format. To provide a data record, choose the **Paste a record** button.

Paste a record

Name	Type	Value
1	age	Numeric

- Nella finestra di dialogo Paste a record (Incolla un record), incollare la seguente osservazione:

```
32, services, divorced, basic.9y, no, unknown, yes, cellular, dec, mon, 110, 1, 11, 0, nonexistent, -1.8, 9
```

- Nella finestra di dialogo Incolla un record, scegliere Inviare per confermare che vuoi generare una previsione per questa osservazione. Amazon ML popola i valori nel modulo della previsione in tempo reale.

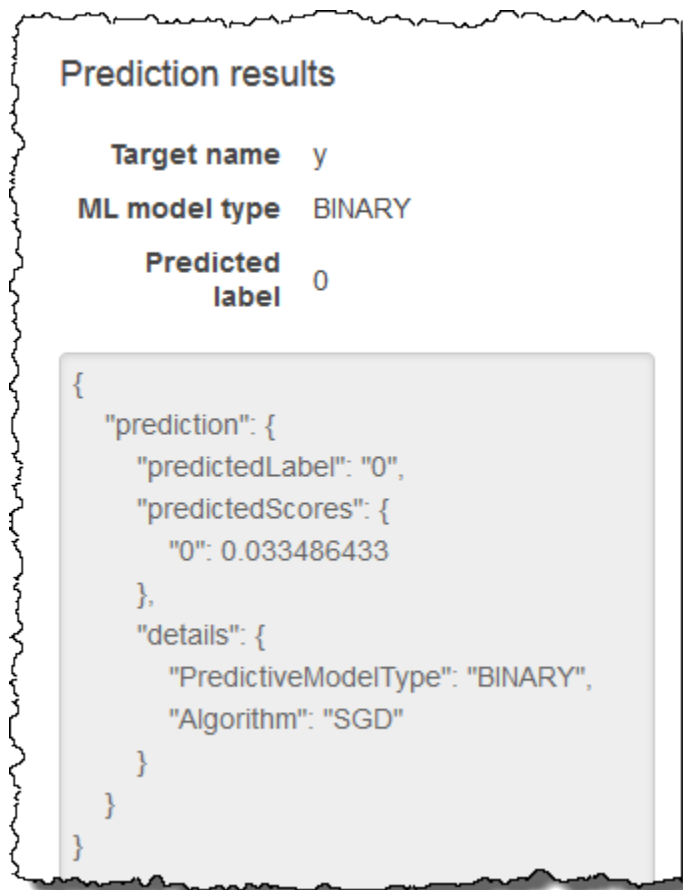
Name	Type	Value
1	age	Numeric
		32.0

Note

È inoltre possibile popolare i campi Value (Valore) digitando i singoli valori. Indipendentemente dal metodo scelto, si deve fornire un'osservazione che non è stata utilizzata per addestrare il modello.

5. Nella parte inferiore della pagina, scegliere Create prediction (Crea previsione).

La previsione appare nel riquadro Prediction results (Risultati delle previsioni) a destra. Questa previsione ha una Predicted label (Etichetta prevista) corrispondente a 0, il che significa che è improbabile che il potenziale cliente reagisca alla campagna. Una Predicted label (Etichetta prevista) uguale a 1 significherebbe che è probabile che il cliente reagisca alla campagna.

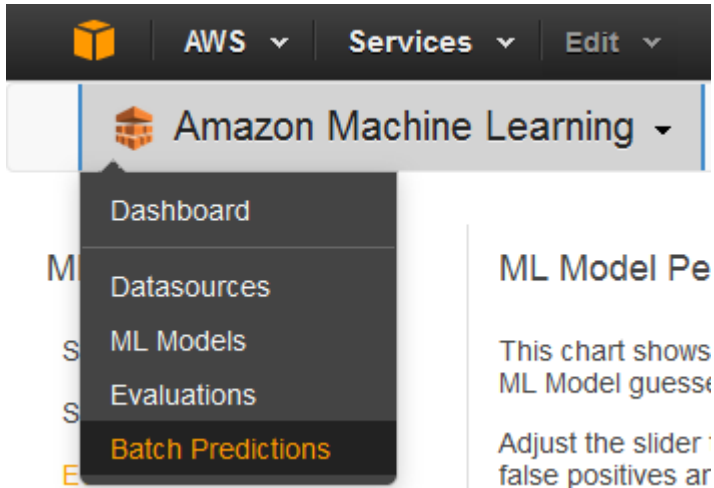


Ora si procede alla creazione di una previsione in batch. È possibile fornire ad Amazon ML il nome del modello ML in uso; la posizione Amazon Simple Storage Service (Amazon S3) dei dati di input

per i quali si desidera generare le previsioni (Amazon ML creerà un'origine dati per la previsione in batch da questi dati) e la posizione Amazon S3 per archiviare i risultati.

Per creare una previsione in batch.

1. Scegliere Amazon Machine Learning, quindi Batch Predictions (Previsioni in batch).



2. Scegliere Create new batch prediction (Crea nuova previsione in batch).
3. Sul Modello ML per previsioni batch pagina, scegliere modello ML: Dati bancari 1.

Amazon ML mostra il nome del modello ML, l'ID, l'ora di creazione e l'ID dell'origine dati associata.

4. Scegliere Continue (Continua).
5. Per generare le previsioni, è necessario fornire ad Amazon ML i dati per i quali servono le previsioni. Si chiamano dati di input. In primo luogo, inserire i dati di input in un'origine dati, in modo che Amazon ML vi possa accedere.

Per Locate the input data (Individuare i dati di input), scegliere My data is in S3, and I need to create a datasource (I miei dati sono in S3 e devo creare un'origine dati).

Locate the input data I already created a datasource pointing to my S3 data
 My data is in S3, and I need to create a datasource

6. Per Datasource name (Nome origine dati), digitare **Banking Data 2**.
7. Per S3 Location (Posizione S3), digitare il percorso completo del file `banking-batch.csv`:
proprio-bucket/banking-batch.csv.

8. Per Does the first line in your CSV contain the column names? (La prima riga del CSV contiene i nomi di colonna?), scegliere Yes (Sì).
9. Selezionare Verify (Verifica).

Amazon ML convalida il percorso dei dati.

10. Scegliere Continue (Continua).
11. Per Destinazione S3), digitare il nome del percorso Amazon S3 in cui sono stati caricati i file nella Fase 1. Preparazione dei dati. Amazon ML carica i risultati delle previsioni in tale posizione.
12. Per Nome della previsione Batch accettare l'impostazione predefinita, **Batch prediction: ML model: Banking Data 1**. Amazon ML sceglie il nome di default sulla base del modello che utilizzerà per creare le previsioni. In questo tutorial, il modello e le previsioni sono denominati sulla base dell'origine dati Banking Data 1.
13. Scegliere Review (Rivedi).
14. Nella finestra di dialogo S3 permissions (Autorizzazioni S3), scegliere Yes (Sì).

S3 permissions

Amazon Machine Learning requires write permission on this S3 location to write output.
Would you like to grant Amazon Machine Learning write permission on this S3 location?

No

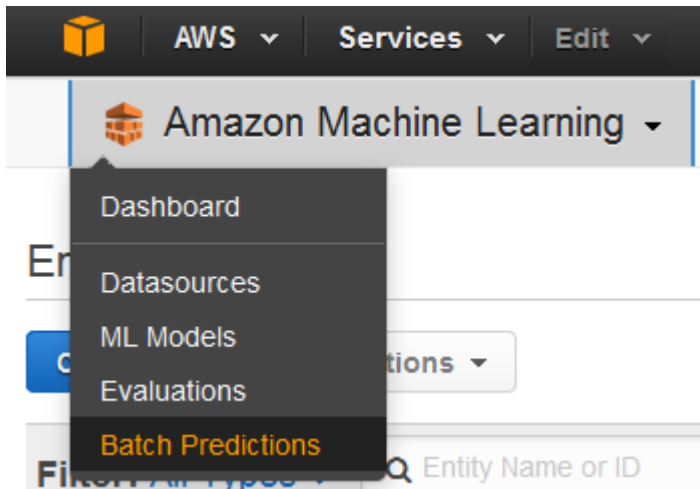
Yes

15. Nella pagina Review (Rivedi), scegliere Finish (Fine).


La richiesta di previsione in batch viene inviata ad Amazon ML e inserita in una coda. Il tempo impiegato da Amazon ML per elaborare una previsione in batch dipende dalla dimensione dell'origine dati e dalla complessità del modello ML. Mentre Amazon ML elabora la richiesta, indica lo stato di Amazon ML In Progress (In corso). Dopo aver completato la previsione in batch, lo stato della richiesta diventa Completed (Completato). A quel punto è possibile visualizzare i risultati.

Per visualizzare le previsioni

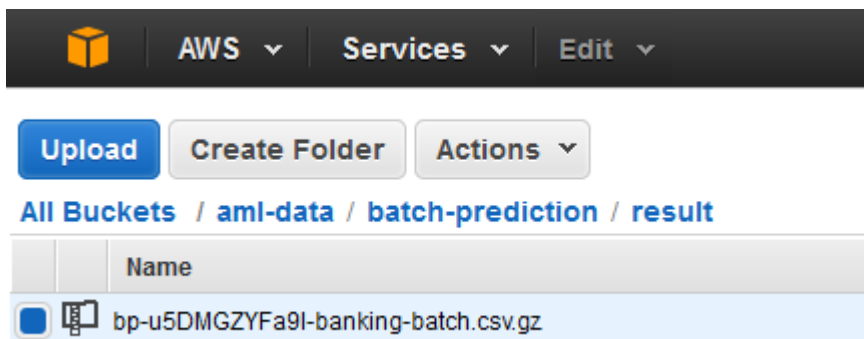
1. Scegliere Amazon Machine Learning, quindi Batch Predictions (Previsioni in batch).



2. Nell'elenco delle previsioni, scegliere la previsione Batch: modello ML: Dati bancari 1. Viene visualizzata la pagina Batch prediction info (Informazioni sulla previsione in batch).

Name	Subscription propensity Predictions 
ID	bp-u5DMGZYFa9I
Creation Time	Mar 5, 2015 3:28:33 PM
Status	Completed
Log	Download Log
Datasource ID	ds-33Rqgz9w3ee
ML Model ID	ml-u7ljoShX2kX
Input S3 URL	s3://aml-data/banking-batch.csv
Output S3 URL	s3://aml-data/

3. Per visualizzare i risultati della previsione in batch, vai alla console Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/> e navigare fino alla posizione Amazon S3 a cui si fa riferimento nella URL di output S3 URL. Da qui, accedere alla cartella dei risultati, che ha un nome simile a s3://aml-data/batch-prediction/result.



La previsione è memorizzata in un file compresso .gzip con estensione .gz.

4. Scaricare il file di previsione nel desktop, decomprimerlo e aprirlo.

bestAnswer	score
0	0.06046
0	0.00507
0	0.01410
0	0.00170
0	0.00184
0	0.07133
0	0.30811

Il file ha due colonne, bestAnswer (migliore risposta) e score (punteggio) e una riga per ogni osservazione all'interno dell'origine dati. I risultati nella colonna bestAnswer (migliore risposta) sono basati sul punteggio soglia di 0,77 impostato in [Fase 4: Esaminare le prestazioni predittive del modello ML e impostare un punteggio soglia](#). Uno score (punteggio) maggiore di 0,77 dà origine a una bestAnswer (migliore risposta) di 1, che costituisce una risposta o previsione positiva, e uno score (punteggio) inferiore a 0,77 dà origine a una bestAnswer (migliore risposta) di 0, che costituisce una risposta o previsione negativa.

I seguenti esempi mostrano previsioni positive e negative basate sul punteggio soglia di 0,77.

Previsione positiva:

bestAnswer	score
1	0.8228876

In questo esempio, il valore di bestAnswer (migliore risposta) è 1 e il valore di score (punteggio) è 0,8228876. Il valore di bestAnswer (migliore risposta) è 1 perché score (punteggio) è superiore al punteggio soglia di 0,77. Un valore bestAnswer (migliore risposta) 1 indica che è probabile che il cliente acquisterà il prodotto e, pertanto, è considerata una previsione positiva.

Previsione negativa:

bestAnswer	score
0	0.7695356

In questo esempio, il valore di bestAnswer (migliore risposta) è 0 perché il valore di score (punteggio) è 0,7695356, che è inferiore al punteggio soglia di 0,77. Un valore bestAnswer (migliore risposta) 0

indica che è improbabile che il cliente acquisterà il prodotto e, pertanto, è considerata una previsione negativa.

Ogni riga del risultato in batch corrisponde a una riga nell'input in batch (un'osservazione nell'origine dati).

Dopo aver analizzato le previsioni, è possibile avviare la campagna di marketing mirata, ad esempio inviando volantini a tutti gli utenti con un punteggio previsto di 1.

Dopo aver creato, rivisto e utilizzato il modello, [ripulire i dati e le risorse AWS creati](#) per evitare di incorrere in costi inutili e mantenere il workspace ordinato.

Fase 6: Eliminare

Per evitare costi Amazon Simple Storage Service (Amazon S3) aggiuntivi, eliminare i dati archiviati in Amazon S3. Non è previsto alcun addebito per le risorse Amazon ML inutilizzate, ma si consiglia di eliminarle per mantenere il workspace pulito.

Per eliminare i dati di input archiviati in Amazon S3

1. Apri la console Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Accedere al percorso Amazon S3 in cui è stato archiviato `ilbanking.csvbanking-batch.csvfile`.
3. Selezionare i file `banking.csvbanking-batch.csv` e `.writePermissionCheck.tmp`.
4. Scegli Actions (Operazioni), quindi Delete (Elimina).
5. Quando viene richiesta la conferma, selezionare OK.

Anche se non sono previsti addebiti per mantenere il record della previsione in batch che ha eseguito Amazon ML o le origini dati, il modello e la valutazione creati durante il tutorial, consigliamo di eliminarli per evitare di occupare troppo spazio nel workspace.

Per eliminare le previsioni in batch

1. Accedere al percorso Amazon S3 in cui è stato archiviato l'output della previsione in batch.
2. Scegliere la cartella `batch-prediction`.
3. Scegli Actions (Operazioni), quindi Delete (Elimina).
4. Quando viene richiesta la conferma, selezionare OK.

Per eliminare le risorse Amazon ML

1. Sul pannello di controllo Amazon ML, selezionare le risorse seguenti.
 - L'origine dati Banking Data 1
 - L'origine dati Banking Data 1_[percentBegin=0, percentEnd=70, strategy=sequential]
 - L'origine dati Banking Data 1_[percentBegin=70, percentEnd=100, strategy=sequential]
 - L'origine dati Banking Data 2
 - Il modello ML model: Banking Data 1
 - La valutazione Evaluation: ML model: Banking Data 1
2. Scegli Actions (Operazioni), quindi Delete (Elimina).
3. Nella finestra di dialogo, scegliere Delete (Elimina) per eliminare tutte le risorse selezionate.

Il tutorial è stato completato con successo. Per continuare a utilizzare la console per creare origini dati, modelli e previsioni consultare la [Guida per sviluppatori Amazon Machine Learning](#). Per ulteriori informazioni su come utilizzare le API, consultare la [documentazione di riferimento delle API di Amazon Machine Learning](#).

Creazione e utilizzo delle origini dati

È possibile utilizzare origini dati Amazon ML per addestrare un modello ML, valutare un modello ML e generare previsioni in batch utilizzando un modello ML. Gli oggetti origine dati contengono metadata sui dati di input. Quando si crea un'origine dati, Amazon ML legge i dati di input, calcola le statistiche descrittive sui relativi attributi e memorizza le statistiche, insieme a uno schema e ad altre informazioni, come parte dell'oggetto origine dati. Dopo aver creato un'origine dati, è possibile utilizzare l'opzione [Informazioni sui dati di Amazon ML](#) per esplorare le proprietà statistiche di dati di input e utilizzare l'origine dati per [addestrare un modello ML](#).

Note

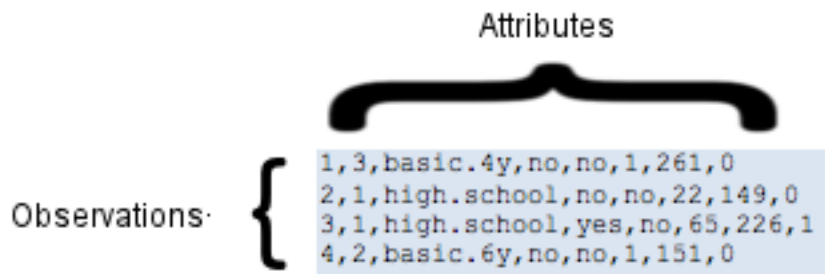
Questa sezione presuppone che l'utente abbia familiarità con [concetti Amazon Machine Learning](#).

Argomenti

- [Comprendere il formato dei dati per Amazon ML](#)
- [Creazione di uno schema di dati per Amazon ML](#)
- [Divisione dei dati](#)
- [Informazioni sui dati](#)
- [Uso di Amazon S3 con Amazon ML](#)
- [Creazione di un'origine dati Amazon ML dai dati in Amazon Redshift](#)
- [Utilizzo dei dati da un database Amazon RDS per creare un'origine dati Amazon ML](#)

Comprendere il formato dei dati per Amazon ML

I dati di input sono i dati che è possibile utilizzare per creare un'origine dati. È necessario salvare i dati di input nel formato .csv (valori separati da virgola). Ogni riga del file .csv è un singolo record di dati o un'osservazione. Ogni colonna del file .csv contiene un attributo dell'osservazione. Ad esempio, la figura riportata di seguito mostra i contenuti di un file .csv che dispone di quattro osservazioni, ciascuna nella propria riga. Ogni osservazione contiene otto attributi, separati da una virgola. Gli attributi rappresentano le seguenti informazioni su ogni individuo rappresentato da un'osservazione: customerId, jobId, education, housing, loan, campaign, duration, willRespondToCampaign.



Attributes

Amazon ML richiede i nomi per ogni attributo. È possibile specificare i nomi degli attributi:

- Includendo i nomi degli attributi nella prima riga (nota anche come riga di intestazione) del file .csv utilizzato per i dati di input
- Includendo i nomi degli attributi in uno schema separato che si trova nello stesso bucket S3 dei dati di input

Per ulteriori informazioni sull'utilizzo dei file dello schema, consultare la pagina relativa alla [creazione di uno schema di dati](#).

L'esempio seguente di un file .csv include i nomi degli attributi nella riga di intestazione.

```
customerId, jobId, education, housing, loan, campaign, duration, willRespondToCampaign
1, 3, basic.4y, no, no, 1, 261, 0
2, 1, high.school, no, no, 22, 149, 0
3, 1, high.school, yes, no, 65, 226, 1
4, 2, basic.6y, no, no, 1, 151, 0
```

Requisiti relativi al formato dei file di input

Il file.csv che contiene i dati di input deve soddisfare i seguenti requisiti:

- Deve essere in testo normale che utilizza un set di caratteri come ASCII, Unicode o EBCDIC.
- Essere costituito da osservazioni, una sola osservazione per riga.

- Per ogni osservazione, i valori degli attributi devono essere separati da virgole.
- Se il valore di un attributo contiene una virgola (delimitatore), l'intero valore di attributo deve essere racchiuso tra virgolette doppie.
- Ogni osservazione deve terminare con un carattere di fine riga, ossia un carattere speciale o una sequenza di caratteri che indica la fine di una riga.
- I valori degli attributi non possono includere caratteri di fine riga, anche se il valore dell'attributo è racchiuso tra virgolette doppie.
- Ogni osservazione deve avere lo stesso numero di attributi e la stessa sequenza di attributi.
- Ogni osservazione non deve avere dimensioni superiori a 100 KB. Amazon ML rifiuta qualsiasi osservazione di dimensioni superiori a 100 KB durante l'elaborazione. Se Amazon ML rifiuta più di 10.000 osservazioni, rifiuta l'intero file .csv.

Utilizzo di più file come input di dati per Amazon ML

È possibile fornire il proprio input ad Amazon ML come un singolo file oppure come una raccolta di file. Le raccolte devono soddisfare queste condizioni:

- Tutti i file devono avere lo stesso schema di dati.
- Tutti i file devono risiedere nello stesso prefisso Amazon Simple Storage Service (Amazon S3) e il percorso indicato per la raccolta deve terminare con una barra (/).

Ad esempio, se i file di dati vengono denominati input1.csv, input2.csv e input3.csv e il nome del bucket S3 è s3://examplebucket, i percorsi dei file potrebbero avere questo aspetto:

```
s3://examplebucket/path/to/data/input1.csv
```

```
s3://examplebucket/path/to/data/input2.csv
```

```
s3://examplebucket/path/to/data/input3.csv
```

È possibile fornire la seguente posizione S3 come input per Amazon ML:

```
's3://examplebucket/path/to/data/'
```

Caratteri di fine riga nel formato CSV

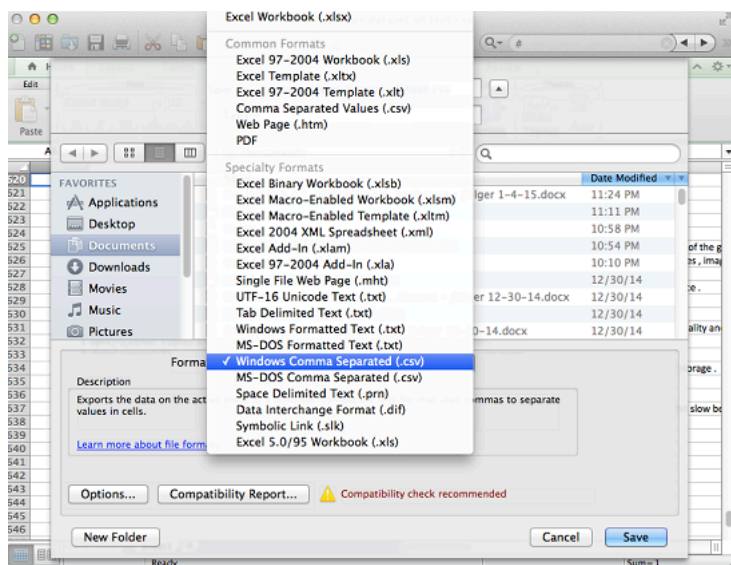
Quando si crea il proprio file .csv, ogni osservazione terminerà con un carattere speciale di fine riga. Questo carattere non è visibile, ma è incluso automaticamente alla fine di ogni osservazione

premendo il tasto Invio o Ritorno a capo. I caratteri speciali che rappresentano la fine della riga variano a seconda del sistema operativo in uso. I sistemi Unix, come ad esempio Linux o OS X, utilizzano un carattere di avanzamento riga indicato da "\n" (codice ASCII decimale 10 o esadecimale 0x0a). Microsoft Windows utilizza due caratteri chiamati ritorno a capo e avanzamento riga che vengono indicati da "\r\n" (codici ASCII decimali 13 e 10 o esadecimali 0x0d e 0x0a).

Se si desidera utilizzare OS X e Microsoft Excel per creare il proprio file .csv, eseguire la procedura seguente. Assicurarsi di scegliere il formato corretto.

Salvare un file .csv se si utilizza OS X ed Excel

1. Quando si salva il file .csv, scegliere Format (Formato) e Windows Comma Separated (.csv).
2. Scegli Salva.



⚠ Important

Non salvare il file.csv utilizzando il file Valori separati da virgola (.csv) o Separato da virgola MS-DOS (.csv) formati perché Amazon ML non è in grado di leggerli.

Creazione di uno schema di dati per Amazon ML

Uno schema è composto da tutti gli attributi dei dati di input e dai relativi tipi di dati. Permette ad Amazon ML di comprendere i dati dell'origine dati. Amazon ML utilizza le informazioni nello schema per leggere e interpretare i dati di input, calcolare le statistiche, applicare le trasformazioni corrette

degli attributi e ottimizzare i propri algoritmi di apprendimento. Se non si fornisce uno schema, Amazon ML ne ricava uno dai dati.

Esempio di schema

Per consentire ad Amazon ML di leggere correttamente i dati di input e generare previsioni accurate, a ogni attributo deve essere assegnato il tipo di dati corretto. Prendiamo in esame un esempio per vedere come i tipi di dati sono assegnati agli attributi e in che modo gli attributi e i tipi di dati sono inclusi in uno schema. Chiameremo il nostro esempio "Customer Campaign" (Campagna clienti) perché vogliamo prevedere quali clienti risponderanno alla nostra campagna e-mail. Il file di input è un file .csv con nove colonne:

```
1,3,web developer,basic.4y,no,no,1,261,0
2,1,car repair,high.school,no,no,22,149,0
3,1,car mechanic,high.school,yes,no,65,226,1
4,2,software developer,basic.6y,no,no,1,151,0
```

Questo è lo schema per questi dati:

```
{
  "version": "1.0",
  "rowId": "customerId",
  "targetAttributeName": "willRespondToCampaign",
  "dataFormat": "CSV",
  "dataFileContainsHeader": false,
  "attributes": [
    {
      "attributeName": "customerId",
      "attributeType": "CATEGORICAL"
    },
    {
      "attributeName": "jobId",
      "attributeType": "CATEGORICAL"
    },
    {
      "attributeName": "jobDescription",
      "attributeType": "TEXT"
    },
    {
```

```
        "attributeName": "education",
        "attributeType": "CATEGORICAL"
    },
    {
        "attributeName": "housing",
        "attributeType": "CATEGORICAL"
    },
    {
        "attributeName": "loan",
        "attributeType": "CATEGORICAL"
    },
    {
        "attributeName": "campaign",
        "attributeType": "NUMERIC"
    },
    {
        "attributeName": "duration",
        "attributeType": "NUMERIC"
    },
    {
        "attributeName": "willRespondToCampaign",
        "attributeType": "BINARY"
    }
]
}
```

Nel file dello schema di questo esempio, il valore di `rowId` è `customerId`:

```
"rowId": "customerId",
```

L'attributo `willRespondToCampaign` viene definito come attributo di destinazione:

```
"targetAttributeName": "willRespondToCampaign ",
```

L'attributo `customerId` e il tipo di dati `CATEGORICAL` sono associati alla prima colonna, l'attributo `jobId` e il tipo di dati `CATEGORICAL` sono associati alla seconda colonna, l'attributo `jobDescription` e il tipo di dati `TEXT` sono associati alla terza colonna, l'attributo `education` e il tipo di dati `CATEGORICAL` sono associati alla quarta colonna e così via. La nona colonna è associata

a un attributo `willRespondToCampaign` con un tipo di dati `BINARY` e questo attributo, inoltre, viene definito come l'attributo di destinazione.

Utilizzo del campo `targetAttributeName`

Il valore `targetAttributeName` è il nome dell'attributo che si desidera prevedere. È necessario assegnare un `targetAttributeName` durante la creazione o la valutazione di un modello.

Quando si sta allenando o valutando un modello ML, `targetAttributeName` identifica il nome dell'attributo nei dati di input che contengono le risposte «corrette» per l'attributo di destinazione. Amazon ML utilizza il target, che include le risposte corrette, per rilevare i pattern e generare un modello ML.

Quando si sta valutando il modello, Amazon ML utilizza il target per verificare l'accuratezza delle previsioni. Dopo aver creato e valutato il modello ML, è possibile utilizzare i dati con un `targetAttributeName` non assegnato per generare previsioni con il modello ML.

È possibile definire l'attributo di destinazione nella console di Amazon ML al momento della creazione di un'origine dati o in un file di schema. Se si crea il proprio file di schema, utilizzare la sintassi seguente per definire l'attributo di destinazione:

```
"targetAttributeName": "exampleAttributeTarget",
```

In questo esempio, `exampleAttributeTarget` è il nome dell'attributo nel file di input che costituisce l'attributo di destinazione.

Utilizzo del campo `rowID`

`row ID` (ID riga) è un flag facoltativo associato a un attributo dei dati di input. Se specificato, l'attributo contrassegnato come `row ID` è incluso nell'output di previsione. Questo attributo semplifica l'associazione tra una previsione e la corrispondente osservazione. Un esempio di ottimo `row ID` è un ID cliente o un attributo univoco simile.

Note

L'ID riga è solo per riferimento. Amazon ML non lo utilizza durante l'addestramento di un modello ML. La selezione di un attributo come ID riga lo esclude dall'uso per l'addestramento di un modello ML.

È possibile definire il `row ID` nella console di Amazon ML al momento della creazione di un'origine dati o in un file di schema. Se si crea il proprio file di schema, utilizzare la sintassi seguente per definire il `row ID`:

```
"rowId": "exampleRow",
```

Nell'esempio precedente, `exampleRow` è il nome dell'attributo nel file di input definito come ID riga.

Quando si generano previsioni in batch, è possibile ottenere l'output seguente:

```
tag,bestAnswer,score
55,0,0.46317
102,1,0.89625
```

In questo esempio, `RowID` rappresenta l'attributo `customerID`. Ad esempio, si prevede che il `customerID` 55 risponderà alla campagna e-mail con bassa fiducia (0,46317), mentre si prevede che `customerID` 102 risponderà alla campagna e-mail con elevata fiducia (0,89625).

Utilizzo del campo `AttributeType`

In Amazon ML, esistono quattro tipi di dati per gli attributi:

Binary

Scegliere `BINARY` per un attributo che ha solo due stati possibili, ad esempio `yes` o `no`.

Ad esempio, l'attributo `isNew`, per verificare se una persona è un nuovo cliente, potrà avere un valore `true` per indicare che la persona è un nuovo cliente e un valore `false` per indicare che non è un nuovo cliente.

I valori negativi validi sono `0`, `n`, `no`, `f` e `false`.

I valori positivi validi sono `1`, `y`, `yes`, `t` e `true`.

Amazon ML ignora il caso di input binari e rimuove lo spazio bianco circostante. Ad esempio, `"FaLSe"` è un valore binario valido. È possibile combinare i valori binari utilizzati nella stessa origine dati, ad esempio utilizzando `true,no`, e `1`. Solo uscite Amazon ML `0e1` per gli attributi binari.

Categorico

Scegliere `CATEGORICAL` (categorico) per un attributo che utilizza un numero limitato di valori univoci di stringa. Ad esempio, un ID utente, il mese e un codice postale sono valori categorici.

Gli attributi categorici vengono trattati come una singola stringa e non sono sottoposti a ulteriore tokenizzazione.

Numerici

Scegliere NUMERIC (numerico) per un attributo che utilizza una quantità come un valore.

Ad esempio, temperatura, peso e percentuale di clic sono valori numerici.

Non tutti gli attributi che contengono numeri sono numerici. Gli attributi categorici, come ad esempio i giorni del mese e gli ID, sono spesso rappresentati da numeri. Per essere considerato numerico, un numero deve essere paragonabile a un altro numero. Ad esempio, l'ID cliente 664727 non dice nulla riguardo all'ID cliente 124552, ma una ponderazione di 10 indica che tale attributo è più pesante di un attributo con una ponderazione di 5. I giorni del mese non sono numerici, perché il primo giorno di un mese può verificarsi prima o dopo il secondo giorno di un altro mese.

Note

Quando si utilizza Amazon ML per creare il proprio schema, assegna il `NUMERIC` tipo di dati per tutti gli attributi che utilizzano i numeri. Se Amazon ML crea lo schema, verificare la presenza di assegnazioni errate e impostare tali attributi su `CATEGORICAL`.

Text (Testo)

Scegliere TEXT per un attributo che è una stringa di parole. Quando legge gli attributi di testo, Amazon ML li converte in token, delimitati da spazi.

Ad esempio, `email subject` diventa `email` e `subject`, mentre `email-subject here` diventa `email-subject` e `here`.

Se il tipo di dati di una variabile nello schema di addestramento non corrisponde al tipo di dati per tale variabile nello schema di valutazione, Amazon ML modifica il tipo di dati di valutazione in base al tipo di dati di addestramento. Ad esempio, se lo schema dei dati di allenamento assegna un tipo di dati di `TEXT` alla variabile `age`, ma lo schema di valutazione assegna un tipo di dati di `NUMERIC` a `age`, quindi Amazon ML tratta le età dei dati di valutazione come `TEXT` variabili invece di `NUMERIC`.

Per informazioni sulle statistiche associate a ogni tipo di dati, consultare [Statistiche descrittive](#).

Fornire uno schema ad Amazon ML

Ogni origine dati richiede uno schema. È possibile scegliere tra due modi per fornire uno schema ad Amazon ML:

- Consentire ad Amazon ML di dedurre i tipi di dati di ogni attributo dal file di dati di input e di creare automaticamente uno schema per l'utente.
- Fornire un file di schema al caricamento dei dati Amazon Simple Storage Service (Amazon S3).

Consentire ad Amazon ML di creare uno schema

Quando si utilizza la console di Amazon ML per creare un'origine dati, Amazon ML utilizza regole semplici, basate sui valori delle variabili, per creare lo schema. È consigliabile rivedere lo schema creato da Amazon ML e correggere i tipi di dati, se non sono corretti.

Fornire uno schema

Dopo aver creato il file di schema, è necessario renderlo disponibile per Amazon ML. Sono disponibili due opzioni:

1. Fornire lo schema utilizzando la console Amazon ML.

Utilizzare la console per creare l'origine dati e includere il file di schema aggiungendo l'estensione `.schema` al nome del file dei dati di input. Ad esempio, se l'URI di Amazon Simple Storage Service (Amazon Simple Storage Service) per i dati di input è `s3://my-bucket-name/data/input.csv`, l'URI dello schema sarà `s3://my-bucket-name/data/input.csv.schema`. Amazon ML individua automaticamente il file di schema fornito invece di tentare di dedurre lo schema dai dati.

Per usare una directory di file come dati di input per Amazon ML, aggiungere l'estensione `.schema` al percorso della directory. Ad esempio, se i file di dati si trovano nel percorso `s3://examplebucket/path/to/data/`, l'URI dello schema sarà `s3://examplebucket/path/to/data/.schema`.

2. Fornire lo schema utilizzando l'API Amazon ML API.

Se si prevede di richiamare l'API Amazon ML per creare l'origine dati, è possibile caricare il file di schema in Amazon S3 e quindi fornire l'URI a tale file nel `DataSchemaLocationS3` attributo del `CreateDataSourceFromS3API`. Per ulteriori informazioni, consultare

[CreateDataSourceFromS3](#).

È possibile fornire lo schema direttamente nel payload di `CreateDataSource*API` invece di salvarlo prima in Amazon S3. Per farlo è sufficiente collocare l'intera stringa dello schema nell'attributo `DataSchema` delle API `CreateDataSourceFromS3`, `CreateDataSourceFromRDS` o `CreateDataSourceFromRedshift`. Per ulteriori informazioni, consultare la pagina relativa alla [documentazione di riferimento delle API di Amazon Machine Learning](#).

Divisione dei dati

L'obiettivo fondamentale di un modello ML è fare previsioni accurate su future istanze di dati al di là di quelle usate per addestrare i modelli. Prima di utilizzare un modello ML per fare previsioni, è necessario valutare le prestazioni predittive del modello. Per stimare la qualità delle previsioni di un modello ML con dati che non abbia visto, è possibile prenotare o dividere una parte dei dati per i quali si conosce già la risposta come proxy per i dati futuri e valutare quanto sono precise le previsioni del modello ML riguardo alle risposte corrette per tali dati. È possibile ripartire l'origine dati tra un'origine dati di addestramento e un'origine dati di valutazione.

Amazon ML offre tre opzioni per frazionare i dati:

- **Pre-dividere i dati**- È possibile dividere i dati in due ubicazioni di input, prima di caricarli su Amazon Simple Storage Service (Amazon S3) e creare con loro due diverse origini dati.
- **Divisione sequenziale Amazon ML**- È possibile dire ad Amazon ML di dividere i dati sequenzialmente durante la creazione delle origini dati di addestramento e di valutazione.
- **Divisione casuale Amazon ML**- È possibile dire ad Amazon ML di dividere i dati con un metodo inizializzato in modo casuale durante la creazione delle origini dati di addestramento e di valutazione.

Pre-divisione dei dati

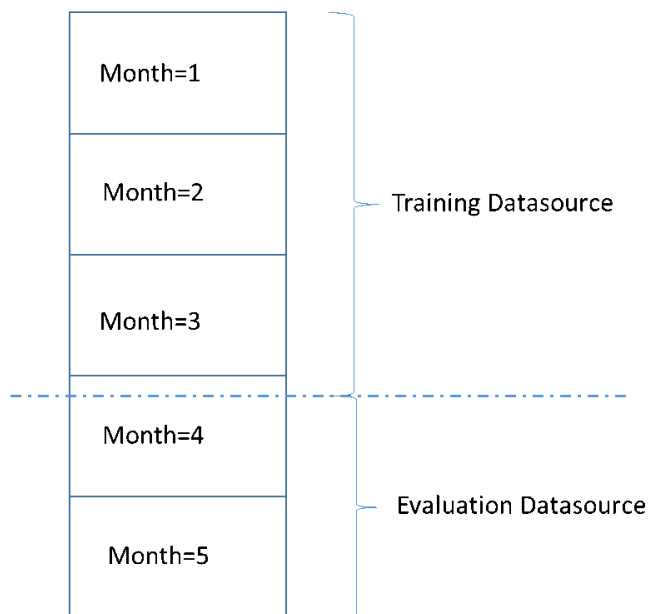
Se si desidera un controllo esplicito sui dati delle origini dati di addestramento e di valutazione, è possibile dividere i dati in ubicazioni dati distinte e creare origini dati separate per l'ubicazione di input e quella di valutazione.

Divisione sequenziale dei dati

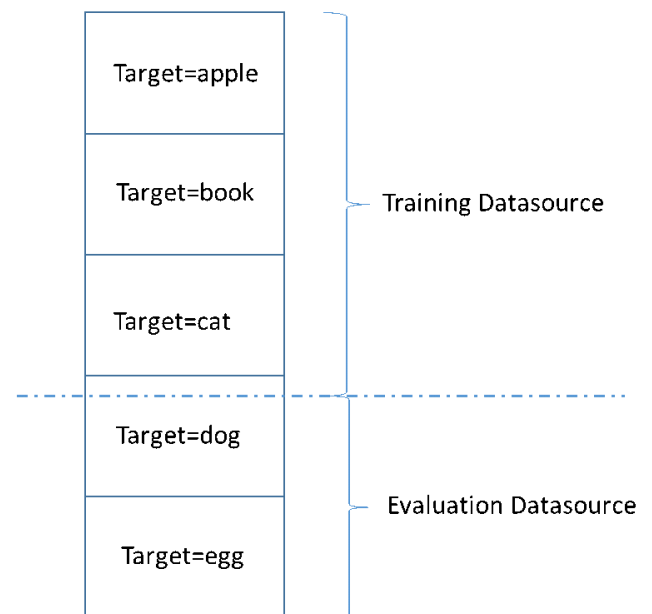
Un modo semplice per dividere i dati di input per l'addestramento e la valutazione è selezionare sottoinsiemi non sovrapposti dei dati mantenendo l'ordine dei record dei dati. Questo approccio è utile se si desidera valutare i modelli ML sui dati per una determinata data o all'interno di un determinato intervallo di tempo. Ad esempio, supponiamo di avere a disposizione i dati relativi al coinvolgimento dei clienti negli ultimi cinque mesi e di volere utilizzare questi dati storici per prevedere il coinvolgimento dei clienti nel mese successivo. L'uso dell'inizio dell'intervallo per l'addestramento e dei dati dalla fine dell'intervallo per la valutazione potrebbe produrre una stima più precisa della qualità del modello rispetto all'utilizzo di record di dati provenienti dall'intero intervallo di dati.

La figura seguente mostra esempi di quando è preferibile utilizzare una strategia di divisione sequenziale rispetto a una strategia casuale.

Case 1: Sequential split is the **correct** strategy



Case 2: Sequential split is the **wrong** strategy

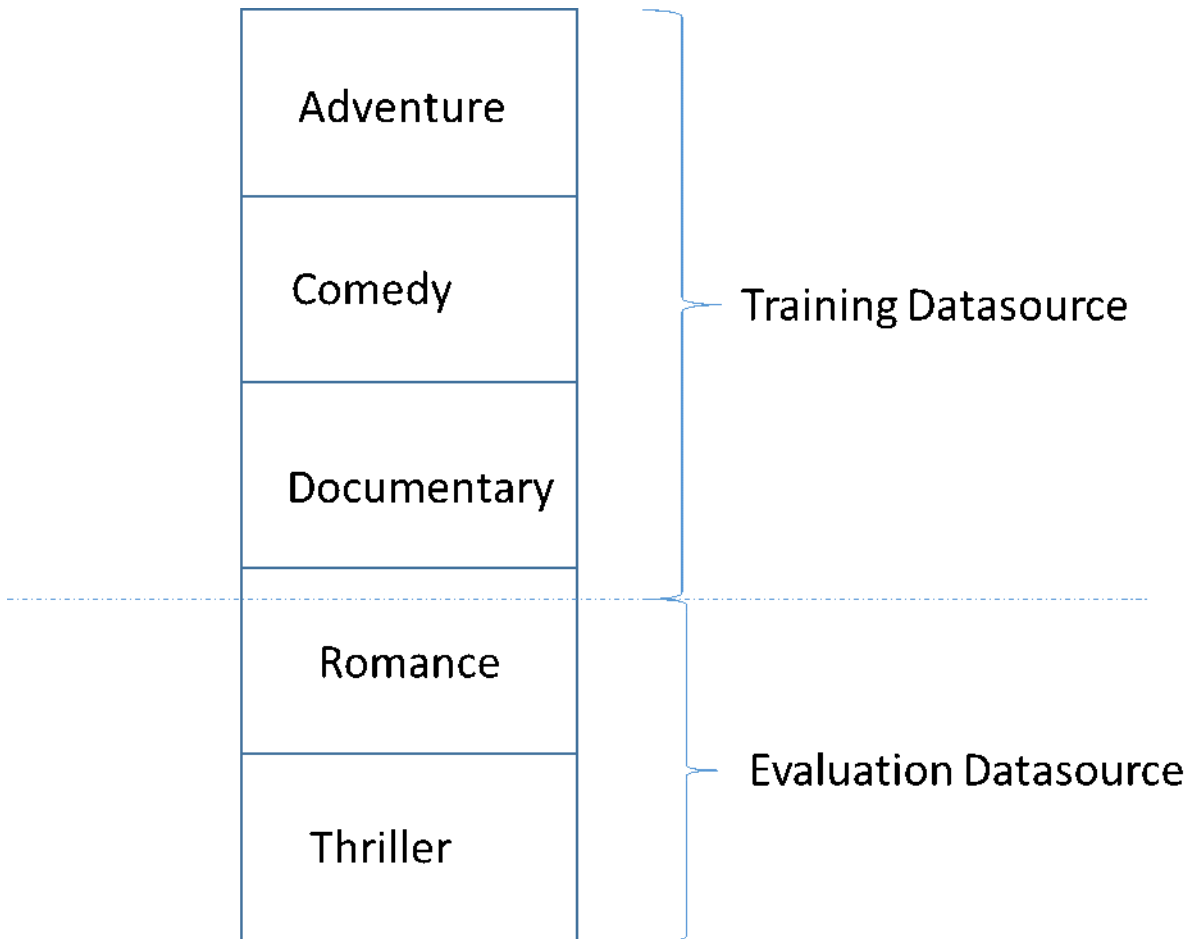


Quando si crea un'origine dati, è possibile scegliere di dividere le origini dati sequenzialmente cosicché Amazon ML usa il primo 70% dei dati per l'addestramento e il restante 30% dei dati per la valutazione. Questa è l'impostazione di default quando si utilizza la console Amazon ML per dividere i dati.

Divisione casuale dei dati

La divisione casuale dei dati di input in origini dati per l'addestramento e per la valutazione garantisce che la distribuzione dei dati sia simile per le origini dati di addestramento e di valutazione. Si sceglie questa opzione quando non è necessario conservare l'ordine dei dati di input.

Amazon ML utilizza un metodo di generazione di numeri pseudocasuale per dividere i dati. La partenza si basa parzialmente su un valore di stringa di input e parzialmente sul contenuto degli stessi dati. Per impostazione predefinita, la console Amazon ML usa la posizione S3 dei dati di input come stringa. Gli utenti API possono fornire una stringa personalizzata. Ciò significa che se si utilizzano lo stesso bucket e gli stessi dati S3, Amazon ML divide i dati nello stesso modo. Per modificare il modo in cui Amazon ML divide i dati, è possibile utilizzare `CreateDataSourceFromS3`, `CreateDataSourceFromRedshift`, oppure `CreateDataSourceFromRDSAPI` e fornisce un valore per la stringa seed. Quando si utilizzano queste API per creare origini dati separate per l'addestramento e per la valutazione, è importante utilizzare lo stesso valore di stringa di partenza per entrambe le origini dati e il flag complementare per un'origine dati, per garantire che non vi siano sovrapposizioni tra i dati di addestramento e quelli di valutazione.



Un problema comune nello sviluppo di un modello di ML di alta qualità è valutare il modello ML su dati che non siano simili a quelli utilizzati per l'addestramento. Ad esempio, supponiamo di utilizzare ML per prevedere il genere dei film e che i dati di addestramento contengano film dei generi Avventura, Commedia e Documentario. Tuttavia, i dati di valutazione contengono solo i dati dei generi Romantico e Thriller. In questo caso, il modello ML non apprende le informazioni sui generi Romantico e Thriller e la valutazione non consente di capire quanto è elevata la qualità delle prestazioni di apprendimento del modello per i generi Avventura, Commedia e Documentario. Di conseguenza, le informazioni sul genere sono inutili e la qualità delle previsioni del modello ML per tutti i generi è compromessa. Il modello e la valutazione sono troppo diversi (hanno statistiche descrittive estremamente diverse) per essere utili. Questo può accadere quando i dati di input sono ordinati per una delle colonne del set di dati e quindi divisi sequenzialmente.

Se le origini dati di addestramento e di valutazione hanno diverse distribuzioni dei dati, verrà visualizzato un avviso di valutazione nel modello di valutazione. Per ulteriori informazioni sugli avvisi di valutazione, consultare [Avvisi relativi alla valutazione](#).

Non è necessario utilizzare la divisione casuale in Amazon ML se i dati di input sono già randomizzati, ad esempio mescolando casualmente i dati di input in Amazon S3 o utilizzando una query SQL Amazon Redshift `random()` funzione o query MySQL `SQLrand()` funzione durante la creazione delle origini dati. In questi casi, è possibile utilizzare l'opzione di divisione sequenziale per creare origini dati di addestramento e valutazione con distribuzioni simili.

Informazioni sui dati

Amazon ML calcola statistiche descrittive sui dati di input che si possono utilizzare per comprendere i dati.

Statistiche descrittive

Amazon ML calcola le seguenti statistiche descrittive per diversi tipi di attributi:

Numerici:

- Istogrammi di distribuzione
- Numero di valori non validi
- Valori minimo, mediano, medio e massimo

Binari e categorici:

- Conteggio (di valori distinti per categoria)
- Istogramma della distribuzione dei valori
- Valori più frequenti
- Conteggi dei valori univoci
- Percentuale del valore "true" (solo binari)
- Parole più importanti
- Parole più frequenti

Testo:

- Nome dell'attributo
- Correlazione con il target (se il target è impostato)
- Totale parole

- Parole univoche
- Intervallo del numero di parole in una riga
- Intervallo di lunghezze delle parole
- Parole più importanti

Accesso alle informazioni sui dati nella console di Amazon ML

Nella console di Amazon ML è possibile scegliere il nome o l'ID di un'origine dati qualsiasi per visualizzare la relativa informazioni sui dati (Certificato creato). Questa pagina fornisce i parametri e le visualizzazioni che consentono di conoscere i dati di input associati all'origine dati, incluse le informazioni riportate di seguito:

- Riepilogo dei dati
- Distribuzioni di destinazione
- Valori mancanti
- Valori non validi
- Statistiche di riepilogo delle variabili in base al tipo di dati
- Distribuzioni delle variabili in base al tipo di dati

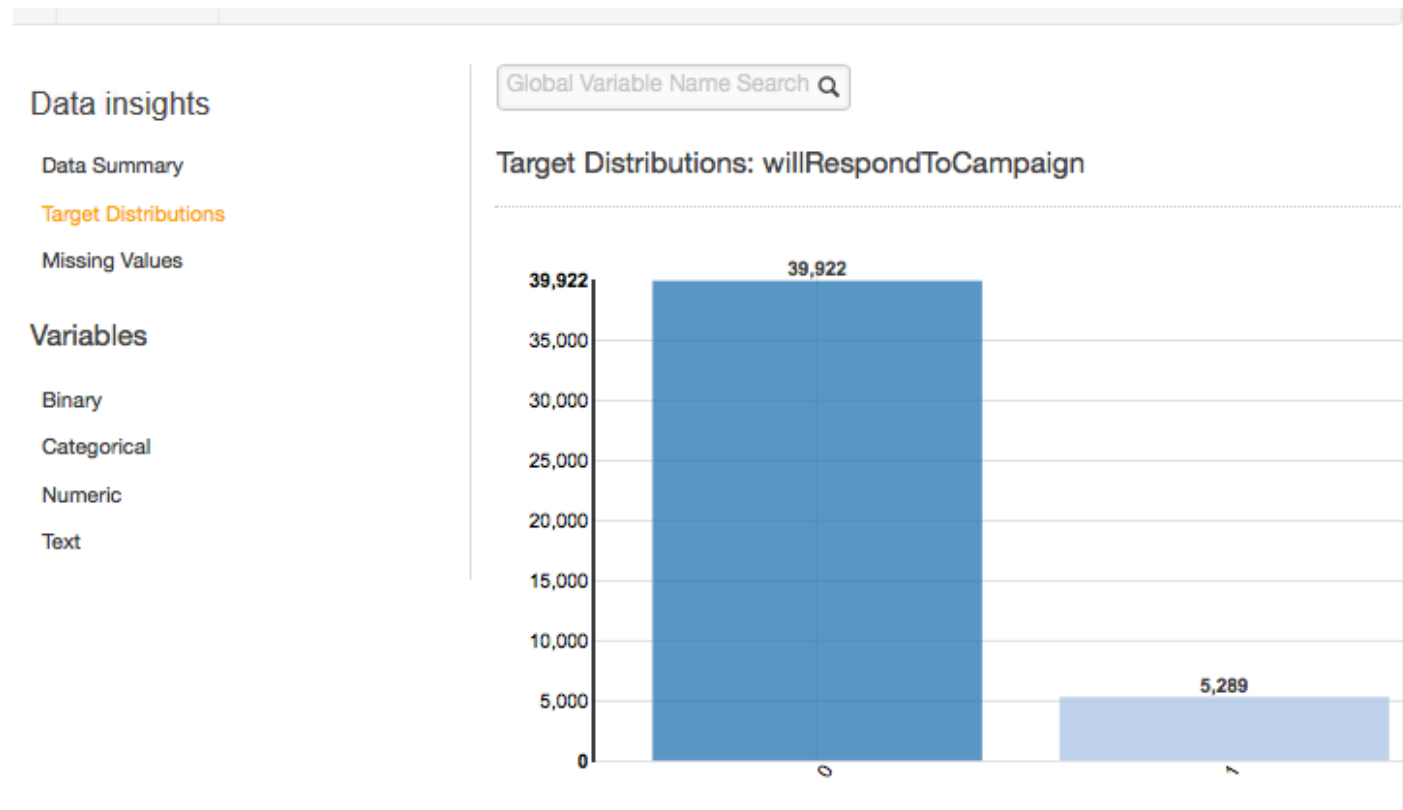
Le seguenti sezioni descrivono i parametri e le visualizzazioni nel dettaglio.

Riepilogo dei dati

Il report di riepilogo dei dati di un'origine dati visualizza le informazioni di riepilogo, tra cui ID dell'origine dati, nome, dove è stata completata, stato attuale, attributo di destinazione, informazioni sui dati di input (posizione del bucket S3, formato dei dati, numero di record elaborati e numero di record danneggiata riscontrati durante l'elaborazione) e numero di variabili in base al tipo di dati.

Distribuzioni di destinazione

Il report delle distribuzioni di destinazione mostra la distribuzione dell'attributo target dell'origine dati. In questo esempio, vi sono 39.922 osservazioni in cui l'attributo target `willRespondToCampaign` è uguale a 0. Questo è il numero di clienti che non hanno risposto alla campagna e-mail. Vi sono 5.289 osservazioni dove `willRespondToCampaign` è uguale a 1. Questo è il numero di clienti che hanno risposto alla campagna e-mail.



Valori mancanti

Il report dei valori mancanti elenca gli attributi dei dati di input per i quali mancano i valori. Solo gli attributi con tipi di dati numerici possono avere valori mancanti. Poiché i valori mancanti possono influenzare la qualità dell'addestramento di un modello ML, è consigliabile fornire i valori mancanti, se possibile.

Durante l'addestramento del modello ML, se l'attributo di destinazione è mancante, Amazon ML rifiuta il record corrispondente. Se l'attributo di destinazione è presente nel record, ma manca il valore di un altro attributo numerico, Amazon ML ignorerà il valore mancante. In questo caso, Amazon ML crea un attributo sostitutivo e lo imposta su 1 per indicare che l'attributo è mancante. In questo modo Amazon ML può apprendere pattern da fatto che vi sono valori mancanti.

Valori non validi

I valori non validi possono verificarsi solo con tipi di dati numerici e binari. Si possono trovare i valori non validi visualizzando le statistiche di riepilogo delle variabili nei report dei tipi di dati. Negli esempi seguenti vi è un solo valore non valido nell'attributo di durata Numeric e due valori non validi nel tipo di dati Binary (uno nell'attributo housing e uno nell'attributo loan).

Numeric Variables

Variables ^	Correlations to Target ⇅	Missing Values ⇅	Invalid Values ⇅	Range ⇅	Mean ⇅	Median ⇅	Preview
duration	0.05165	2 (0%)	1 (0%)	0 - 4918	258.1618	180	

Binary Variables

Variables ^	Correlations to Target ⇅	Percent True ⇅	Invalid Values ⇅	Preview
campaign	NA	100%	27667 (61%)	
housing	0.01842	56%	1 (0%)	
loan	0.00656	16%	1 (0%)	
willRespondToCampaign	NA	12%	0 (0%)	

Correlazione variabile-destinazione

Dopo aver creato un'origine dati, Amazon ML è in grado di valutare l'origine dati e identificare la correlazione, o impatto, tra le variabili e la destinazione. Ad esempio, il prezzo di un prodotto potrebbe avere un notevole impatto sulle vendite, mentre le dimensioni del prodotto potrebbero avere scarsa valenza predittiva.

In generale, è consigliabile includere quante più variabili possibile nei dati di addestramento. Tuttavia, il disturbo introdotto dall'inclusione di molte variabili con scarsa capacità predittiva potrebbe influire negativamente sulla qualità e l'accuratezza del modello ML.

È possibile migliorare le prestazioni predittive del modello rimuovendo le variabili che hanno scarso impatto durante l'addestramento del modello. È possibile definire quali variabili sono rese disponibili per il processo di machine learning in unricetta, che è un meccanismo di trasformazione di Amazon ML. Per ulteriori informazioni sulle composizioni, consultare [Trasformazioni dei dati per il Machine Learning](#).

Statistiche di riepilogo degli attributi in base al tipo di dati

Nel report delle informazioni sui dati, è possibile visualizzare gli attributi nelle statistiche di riepilogo per i seguenti tipi di dati:

- Binario
- Categorici
- Numerico
- Text

Le statistiche di riepilogo in base al tipo di dati Binary (binari) mostrano tutti gli attributi binari. La colonna Correlations to target (Correlazioni con il target) mostra le informazioni condivise tra la colonna di destinazione e la colonna degli attributi. La colonna Percent true (Percentuale true) mostra la percentuale di osservazioni che hanno il valore 1. La colonna Invalid values (Valori non validi) mostra il numero di valori non validi, nonché la percentuale di valori non validi per ogni attributo. La colonna Preview (Anteprima) fornisce un collegamento a una distribuzione grafica di ogni attributo.

Binary Variables

Variables	Correlations to Target	Percent True	Invalid Values	Preview
campaign	NA	100%	27667 (61%)	
housing	0.01842	56%	1 (0%)	
loan	0.00656	16%	1 (0%)	
willRespondToCampaign	NA	12%	0 (0%)	

Le statistiche di riepilogo per il tipo di dati Categorical (categorici) mostrano tutti gli attributi di categoria con il numero di valori univoci, il valore più frequente e il valore meno frequente. La colonna Preview (Anteprima) fornisce un collegamento a una distribuzione grafica di ogni attributo.

Categorical Variables

Variables	Correlations to Target	Unique Values	Most Frequent	Least Frequent	Preview
campaign	0.00433	49	1	39	
customerid	NA	45211	45211	1	
education	0.00355	5	secondary		
housing	0.01846	4	1		
jobid	0.00671	13	blue-collar		
willRespondToCampaign	NA	3	0		

Le statistiche di riepilogo per il tipo di dati Numeric (numerici) mostrano tutti gli attributi numerici con il numero di valori mancanti, valori non validi, intervallo di valori, media e valori mediani. La colonna Preview (Anteprima) fornisce un collegamento a una distribuzione grafica di ogni attributo.

Numeric Variables

Variables	Correlations to Target	Missing Values	Invalid Values	Range	Mean	Median	Preview
duration	0.05165	2 (0%)	1 (0%)	0 - 4918	258.1618	180	

Le statistiche di riepilogo per il tipo di dati Text (testo) mostrano tutti gli attributi di testo, il numero totale di parole in tale attributo, il numero di parole univoche in tale attributo, la gamma di parole in un attributo, la gamma di lunghezze delle parole e le parole più importanti. La colonna Preview (Anteprima) fornisce un collegamento a una distribuzione grafica di ogni attributo.

Text attributes

Attributes	Correlations to target *	Total words	Unique words	Words in attribute (range)	Word length (range)	Most prominent words
Phrase	0.07118	751741	12811	0 - 48	1 - 18	enters, trust ...

« < 1 - 1 of 1 Attributes > »

* Correlations to Target is an approximate statistic for text attributes.

L'esempio seguente mostra le statistiche relative al tipo di dati Text (testo) per una variabile di testo denominata review (revisione), con quattro record.

1. The fox jumped over the fence.
2. This movie is intriguing.
- 3.
4. Fascinating movie.

Le colonne di questo esempio mostrerebbero le seguenti informazioni.

- La colonna Attributes (Attributi) mostra il nome della variabile. In questo esempio, la colonna direbbe "review".
- La colonna Correlations to target (Correlazioni con il target) esiste solo se è specificato un target. La correlazione misura la quantità di informazioni che questo attributo fornisce riguardo al target. Più è elevata la correlazione, maggiori sono le informazioni che l'attributo fornisce sul target. La correlazione si misura in termini di informazione reciproca tra una rappresentazione semplificata dell'attributo di testo e il target.
- La colonna Total words (Totale parole) mostra il numero di parole generate dalla tokenizzazione di ogni record, delimitando le parole con uno spazio vuoto. In questo esempio, la colonna direbbe "12".
- La colonna Unique words (Parole univoche) mostra il numero di parole univoche per un attributo. In questo esempio, la colonna direbbe "10".
- La colonna Words in attribute (range) (Parole nell'attributo (intervallo)) mostra il numero di parole in una singola riga nell'attributo. In questo esempio, la colonna direbbe "0-6".
- La colonna Word length (range) (Lunghezza parole (intervallo)) mostra l'intervallo del numero di caratteri che si trovano nelle parole. In questo esempio, la colonna direbbe "2-11".
- La colonna Most prominent words (Parole più importanti) mostra una graduatoria delle parole che appaiono nell'attributo. Se è presente un attributo di destinazione, le parole sono classificate in base alla loro correlazione con il target, il che significa che le parole che hanno la massima correlazione appaiono per prime nell'elenco. Se non è presente una destinazione nei dati, le parole sono classificate in base alla loro entropia.

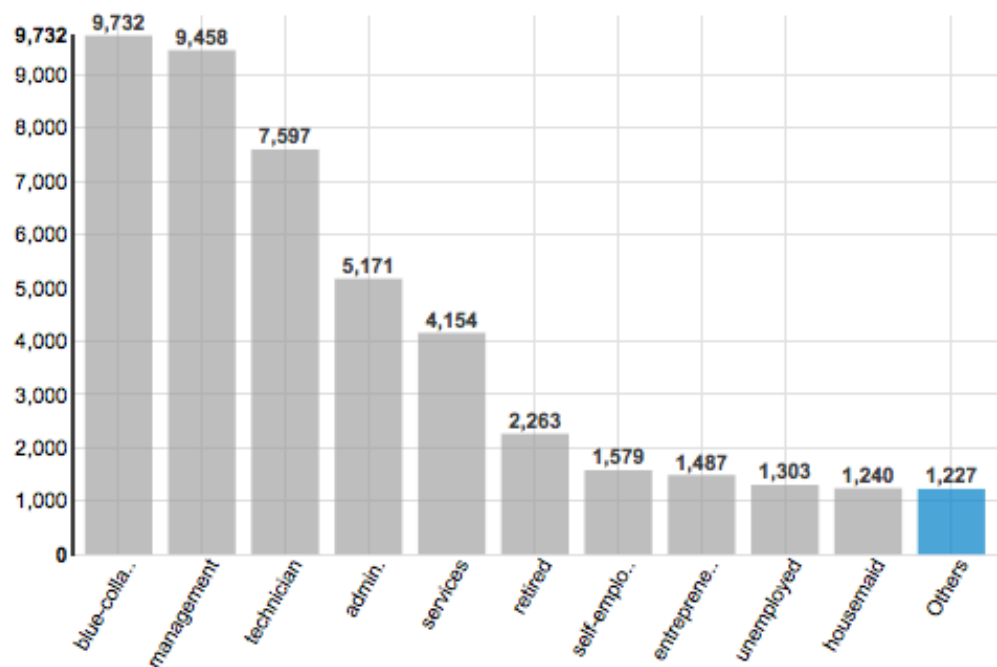
Comprendere la distribuzione degli attributi Categorical e Binary

Facendo clic sul collegamento Preview (Anteprima) associato a un attributo categorico o binario, è possibile visualizzare la distribuzione dell'attributo e i dati di esempio del file di input per ogni valore categorico dell'attributo.

Ad esempio, la seguente screenshot mostra la distribuzione per l'attributo categorico jobId. La distribuzione visualizza i primi 10 valori categorici, con tutti gli altri valori raggruppati come "others" (altri). Classifica ciascuno dei primi 10 valori categorici con il numero di osservazioni nel file di input che contengono tale valore, nonché con un link per visualizzare le osservazioni di esempio dal file dei dati di input.

Categorical Variables: jobId

Top 10 jobId



All Categories

Ranking	Category	Count	
1	blue-collar	9732	Sample data
2	management	9458	Sample data
3	technician	7597	Sample data

Comprendere la distribuzione degli attributi Numeric

Per visualizzare la distribuzione di un attributo numerico, fare clic sul collegamento [Preview \(Anteprima\)](#) dell'attributo. Quando si visualizza la distribuzione di un attributo numerico, è possibile scegliere bin con dimensioni 500, 200, 100, 50 o 20. Più grande è il bin, minore è il numero di barre del grafico che saranno visualizzate. Inoltre, la risoluzione della distribuzione sarà più grossolana con bin di grandi dimensioni. Al contrario, se si impostano le dimensioni del bucket a 20, aumenta la risoluzione della distribuzione visualizzata.

I valori minimo, medio e massimo vengono anche visualizzati, come mostrato nella screenshot seguente.

Numeric Variables: duration

Select Bin Width:

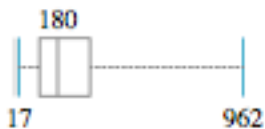
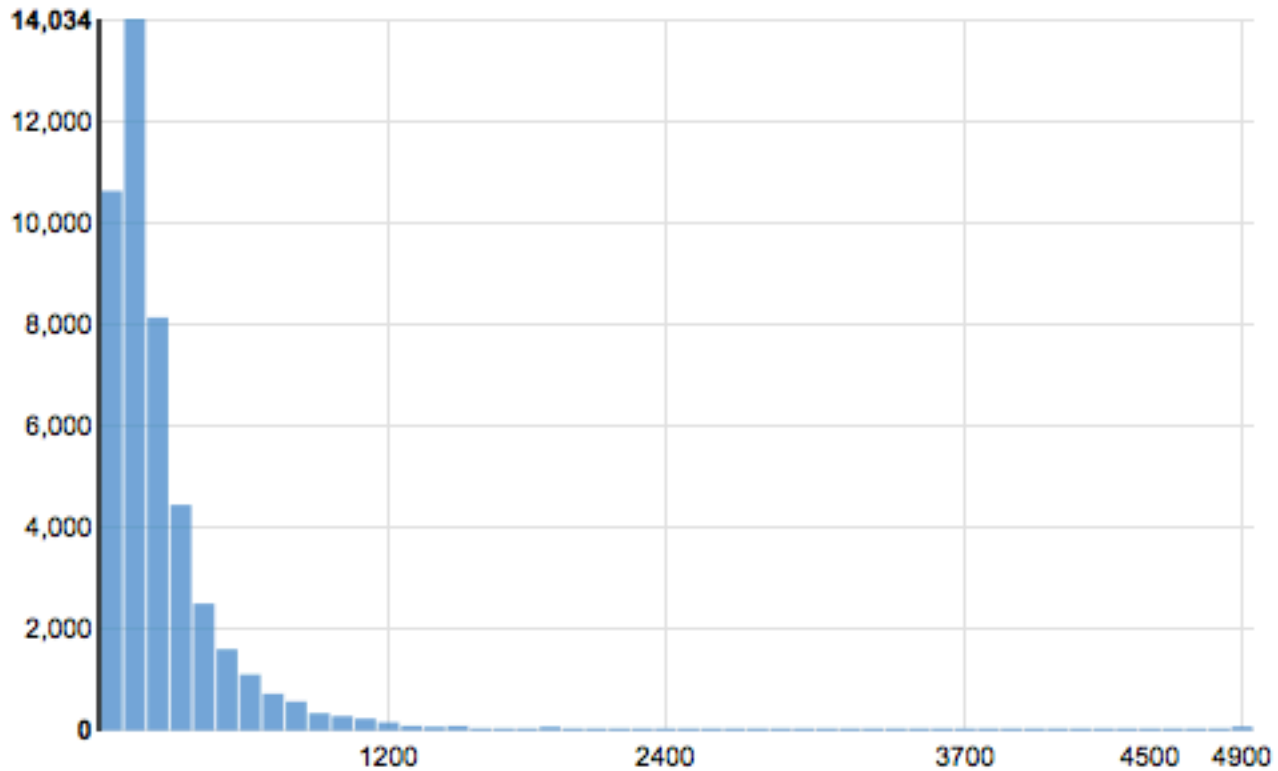
500

200

100

50

20

**Min: 0 Mean: 258.1618 Max: 4918**

Comprendere la distribuzione degli attributi Text

Per visualizzare la distribuzione di un attributo di testo, fare clic sul collegamento Preview (Anteprima) dell'attributo. Quando si visualizza la distribuzione di un attributo di testo, si possono vedere le seguenti informazioni.

Text attributes: Phrase

Ranking	Token	Word prominence	Count	
1	enters	0.01105	7	0.0%
2	trust	0.00884	28	0.0%
3	bad	0.00735	833	0.2%
4	film	0.00669	4747	1.3%
5	movie	0.00611	4242	1.2%
6	unwieldy	0.00605	11	0.0%
7	good	0.00574	1620	0.5%
8	ashamed	0.00551	7	0.0%
9	funny	0.00550	1078	0.3%
10	wankery	0.00498	9	0.0%

« < 1 - 10 of 11091 > »

Ranking (classificazione)

I token di testo sono classificati in base alla quantità di informazioni che forniscono, dai più informativi ai meno informativi.

Token

Il token mostra la parola del testo di input oggetto della riga di statistiche.

Word prominence (Importanza delle parole)

Se è presente un attributo di destinazione, le parole sono classificate in base alla loro correlazione con il target, pertanto le parole che hanno la massima correlazione appaiono per prime nell'elenco. Se nei dati non è presente un target, le parole sono classificate in base alla loro entropia, ossia alla quantità di informazioni che riescono a comunicare.

Count number (Numero conteggio)

Il numero conteggio mostra il numero di record di input in cui è apparso il token.

Count percentage (Percentuale conteggio)

La percentuale conteggio mostra la percentuale di righe di dati di input in cui è apparso il token.

Uso di Amazon S3 con Amazon ML

Amazon Simple Storage Service (Amazon S3) è un servizio di storage su Internet. È possibile utilizzare Amazon S3 per memorizzare e recuperare qualsiasi volume di dati, in qualunque momento e da qualunque luogo tramite il Web. Amazon ML usa Amazon S3 come repository di dati principale per le attività seguenti:

- Accedere ai file di input per creare oggetti origini dati per l'addestramento e la valutazione dei modelli ML.
- Accedere ai file di input per generare previsioni in batch.
- Quando si generano previsioni in batch utilizzando i modelli ML, per genere il file di previsione in un bucket S3 specificato dall'utente.
- Copiare i dati archiviati su Amazon Redshift o Amazon Relational Database Service (Amazon RDS) in un file .csv e caricarlo in Amazon S3.

Per consentire ad Amazon ML di eseguire queste operazioni, è necessario autorizzare ad Amazon ML le autorizzazioni per accedere ai dati Amazon S3.

Note

Non è possibile generare file di previsione in batch in un bucket S3 che accetta solo file crittografati lato server. Verificare che la policy del bucket permetta il caricamento di file non crittografati confermando che la policy non include un effetto Deny per l'azione `s3:PutObject` quando non è presente alcuna intestazione `s3:x-amz-server-side-encryption` nella richiesta. Per ulteriori informazioni sulle policy dei bucket S3 lato server, consultare [Protezione dei dati con la crittografia lato server](#) nella [Uso di Amazon Simple Storage Service](#).

Caricamento di dati in Amazon S3

Si devono caricare i dati di input in Amazon Simple Storage Service (Amazon S3) perché Amazon ML legge i dati dalle posizioni di Amazon S3. Si possono caricare i dati direttamente in Amazon S3 (ad esempio dal computer), oppure Amazon ML può copiare i dati archiviati su Amazon Redshift o Amazon Relational Database Service (RDS) in un file .csv e caricarlo in Amazon S3.

Per ulteriori informazioni sulla copia dei dati da Amazon Redshift o Amazon RDS, consultare rispettivamente [Utilizzo di Amazon Redshift con Amazon ML](#) o [Utilizzo di Amazon RDS con Amazon ML](#).

Il resto di questa sezione descrive come caricare i dati di input direttamente dal computer su Amazon S3. Prima di iniziare le procedure di questa sezione, è necessario disporre dei dati in un file .csv. Per informazioni su come formattare correttamente il file .csv affinché Amazon ML possa utilizzarlo, consultare [Comprendere il formato dei dati per Amazon ML](#).

Per caricare i dati dal computer su Amazon S3

1. Accedere alla Console di gestione AWS e aprire la console Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Creare un bucket o sceglierne uno esistente.
 - a. Per creare un bucket, scegliere Create Bucket (Crea bucket). Denominare il bucket, scegliere una regione (è possibile scegliere qualsiasi regione disponibile), quindi Create (Crea). Per ulteriori informazioni, consultare [Creare un bucket](#) nella Guida alle operazioni di base di Amazon Simple Storage.
 - b. Per usare un bucket esistente, eseguire la ricerca del bucket scegliendolo nell'elenco All Buckets (Tutti i bucket). Quando viene visualizzato il nome del bucket, selezionarlo e scegliere Upload (Carica).
3. Nella finestra di dialogo Upload (Carica), scegliere Add files (Aggiungi file).
4. Accedere alla cartella che contiene il file di input .csv e scegliere Open (Apri).

Autorizzazioni

Per concedere ad Amazon ML le autorizzazioni per accedere a uno dei bucket S3, è necessario modificare la policy dei bucket.

Per informazioni sulla concessione ad Amazon ML dell'autorizzazione per leggere i dati dal bucket in Amazon S3, consultare [Concessione ad Amazon ML delle autorizzazioni per leggere i dati da Amazon S3](#).

Per informazioni sulla concessione ad Amazon ML dell'autorizzazione per l'output dei risultati delle previsioni in batch nel proprio bucket in Amazon S3, consultare [Concessione ad Amazon ML delle autorizzazioni per generare previsioni in Amazon S3](#).

Per ulteriori informazioni sulla gestione delle autorizzazioni di accesso alle risorse Amazon S3, consultare la [Guida per gli sviluppatori Amazon S3](#).

Creazione di un'origine dati Amazon ML dai dati in Amazon Redshift

Se disponi di dati archiviati in Amazon Redshift, è possibile utilizzare la Creazione di un'origine dati nella console di Amazon Machine Learning (Amazon ML) per creare un oggetto origine dati. Quando si crea un'origine dati dai dati Amazon Redshift, è necessario specificare il cluster che contiene i dati e le query SQL per recuperare i dati. Amazon ML esegue la query richiamando Amazon RedshiftUnLoad sul cluster. Amazon ML archivia i risultati nella posizione Amazon Simple Storage Service (Amazon S3) scelta e quindi utilizza i dati archiviati in Amazon S3 per creare l'origine dati. L'origine dati, il cluster Amazon Redshift e il bucket S3 devono essere tutti nella stessa regione.

Note

Amazon ML non supporta la creazione di origini dati da cluster Amazon Redshift in VPC privati. Il cluster deve disporre di un indirizzo IP pubblico.

Argomenti

- [Parametri obbligatori per la procedura guidata Crea origine dati](#)
- [Creazione di un'origine dati con Amazon Redshift Data \(console\)](#)
- [Risoluzione dei problemi di Amazon Redshift](#)

Parametri obbligatori per la procedura guidata Crea origine dati

Affinché Amazon ML si connetta al database Amazon Redshift e legga i dati per conto dell'utente, è necessario fornire quanto segue:

- `Amazon RedshiftClusterIdentifier`
- Il nome del database Amazon Redshift
- Le credenziali del database Amazon Redshift (nome utente e password)
- Amazon ML Amazon RedshiftAWS Identity and Access Management(IAM) ruolo
- La query SQL Amazon Redshift
- (Facoltativo) La posizione dello schema Amazon ML
- Il percorso di gestione temporanea di Amazon S3 (dove Amazon ML colloca i dati prima di creare l'origine dati)

Inoltre, è necessario accertarsi che gli utenti o i ruoli IAM che creano le origini dati Amazon Redshift (tramite la console o utilizzando il `CreateDataSourceFromRedshift` azione) avere `iam:PassRole` autorizzazione.

Amazon Redshift `ClusterIdentifier`

Utilizzare questo parametro che distingue tra maiuscole e minuscole per consentire ad Amazon ML di trovare e connettersi al cluster. È possibile ottenere l'identificatore del cluster (nome) dalla console Amazon Redshift. Per ulteriori informazioni sui cluster, consulta [Cluster Amazon Redshift](#).

Nome del database Amazon Redshift

Utilizzare questo parametro per indicare ad Amazon ML quale database del cluster Amazon Redshift contiene i dati che si desidera utilizzare come origine dati.

Credenziali del database Amazon Redshift

Utilizzare questi parametri per specificare il nome utente e la password dell'utente del database Amazon Redshift nel cui contesto sarà eseguita la query di sicurezza.

Note

Amazon ML richiede un nome utente e una password Amazon Redshift per connettersi al database Amazon Redshift. Dopo aver scaricato i dati in Amazon S3, Amazon ML non riutilizza mai la password, né la memorizza.

Ruolo Amazon ML Amazon Redshift

Utilizzare questo parametro per specificare il nome del ruolo IAM che Amazon ML deve utilizzare per configurare i gruppi di sicurezza per il cluster Amazon Redshift e il bucket della policy per il percorso di gestione temporanea di Amazon S3.

Se non si dispone di un ruolo IAM in grado di accedere ad Amazon Redshift, Amazon ML può crearne uno. Quando Amazon ML crea un ruolo, crea e attribuisce una policy gestita dal cliente a un ruolo IAM. La policy che Amazon ML crea concede ad Amazon ML l'autorizzazione per accedere solo al cluster specificato dall'utente.

Se si dispone già di un ruolo IAM per accedere ad Amazon Redshift, è possibile digitare l'ARN del ruolo oppure scegliere il ruolo dall'elenco a discesa. I ruoli IAM con accesso Amazon Redshift sono elencati nella parte superiore dell'elenco a discesa.

Il ruolo IAM deve avere i seguenti contenuti:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "machinelearning.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": { "aws:SourceAccount": "123456789012" },
        "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-east-1:123456789012:datasource/*" }
      }
    }
  ]
}
```

Per ulteriori informazioni sulle policy gestite dal cliente, consulta [Policy gestite dal cliente](#) nella IAM User Guide.

Query SQL Amazon Redshift

Utilizzare questo parametro per specificare la query SQL SELECT che Amazon ML esegue nel database Amazon Redshift per selezionare i dati. Amazon ML utilizza Amazon

Redshift [SCARICARE](#) azione per copiare in modo sicuro i risultati della query in una posizione Amazon S3.

Note

Amazon ML funziona meglio quando i record di input sono in ordine casuale (mischiat). È possibile mischiare i risultati della query SQL Amazon Redshift utilizzando Amazon Redshift.casuale ()funzione. Ad esempio, supponiamo che questa sia la query originale:

```
"SELECT col1, col2, ... FROM training_table"
```

È possibile incorporare un mescolamento casuale aggiornando la query in questo modo:

```
"SELECT col1, col2, ... FROM training_table ORDER BY random()"
```

Posizione Posizione dello schema (facoltativa)

Utilizzare questo parametro per specificare il percorso Amazon S3 dello schema per i dati Amazon Redshift che Amazon ML esporterà.

Se non si fornisce uno schema per l'origine dati, la console di Amazon ML crea automaticamente uno schema Amazon ML in base allo schema dati della query SQL Amazon Redshift. Gli schemi Amazon ML dispongono di un minor numero di tipi di dati rispetto agli schemi Amazon Redshift, perciò non si tratta di una conversione uno-a-uno. La console Amazon ML converte i tipi di dati Amazon Redshift in tipi di dati Amazon ML utilizzando il seguente schema di conversione.

Tipi di dati Amazon Redshift	Alias di Amazon Redshift	Tipo di dati Amazon ML
SMALLINT	INT2	NUMERIC
INTEGER	INT, INT4	NUMERIC
BIGINT	INT8	NUMERIC
DECIMAL	NUMERIC	NUMERIC
REAL	FLOAT4	NUMERIC

Tipi di dati Amazon Redshift	Alias di Amazon Redshift	Tipo di dati Amazon ML
DOUBLE PRECISION	FLOAT8, FLOAT	NUMERIC
BOOLEAN	BOOL	BINARY
CHAR	CHARACTER, NCHAR, BPCHAR	CATEGORICAL
VARCHAR	CHARACTER VARYING, NVARCHAR, TEXT	TEXT
DATE		TEXT
TIMESTAMP	TIMESTAMP WITHOUT TIME ZONE	TEXT

Da convertire in Amazon ML `Binary` tipi di dati, i valori `Booleans` Amazon Redshift nei dati devono essere valori `Binary` Amazon ML `Binary` supportati. Se il tipo di dati `Boolean` ha valori non supportati, Amazon ML li converte nel tipo di dati più specifico che riesce a trovare. Ad esempio, se un `booleano` Amazon Redshift ha i valori `0,1, e2`, Amazon ML converte il valore booleano in un `Numeric` tipo di dati. Per ulteriori informazioni sui valori binari supportati, consultare [Utilizzo del campo `AttributeType`](#).

Se Amazon ML non è in grado di individuare un tipo di dati, l'impostazione predefinita è `Text`.

Dopo che Amazon ML ha convertito lo schema, è possibile esaminare e correggere i tipi di dati assegnati da Amazon ML nella procedura guidata Crea origine dati e rivedere lo schema prima che Amazon ML crei l'origine dati.

Posizione temporanea di Amazon S3

Utilizzare questo parametro per specificare il nome del percorso di gestione temporanea di Amazon S3 in cui Amazon ML archivia i risultati della query SQL Amazon Redshift. Dopo aver creato l'origine dati, Amazon ML utilizza i dati nel percorso di gestione temporanea anziché tornare ad Amazon Redshift.

Note

Poiché Amazon ML assume il ruolo IAM definito dal ruolo Amazon Redshift Amazon ML, Amazon ML dispone delle autorizzazioni per accedere a eventuali oggetti nel percorso specificato di gestione temporanea di Amazon S3. Per questo motivo, consigliamo di archiviare nel percorso di gestione temporanea di Amazon S3 solo i file che non contengono informazioni sensibili. Ad esempio, se il bucket root è `s3://mybucket/`, consigliamo di creare una posizione per archiviare solo i file a cui si desidera che Amazon ML abbia accesso, come `s3://mybucket/AmazonMLInput/`.

Creazione di un'origine dati con Amazon Redshift Data (console)

La console Amazon ML offre due modi per creare un'origine dati utilizzando i dati Amazon Redshift. È possibile creare un'origine dati completando la procedura guidata Crea origine dati oppure, se si dispone già di un'origine dati creata con i dati Amazon Redshift, è possibile copiare l'origine dati originale e modificarne le impostazioni. Se si copia un'origine dati, è possibile creare più origini dati simili.

Per ulteriori informazioni sulla creazione di un'origine dati utilizzando l'API, consultare la pagina [CreateDataSourceFromRedshift](#).

Per ulteriori informazioni sui parametri nelle seguenti procedure, consultare la pagina [Parametri obbligatori per la procedura guidata Crea origine dati](#).

Argomenti

- [Creazione di un'origine dati \(console\)](#)
- [Copia di un'origine dati \(console\)](#)

Creazione di un'origine dati (console)

Per scaricare i dati da Amazon Redshift in un'origine dati Amazon ML, utilizzare la procedura guidata Crea origine dati.

Per creare un'origine dati dai dati in Amazon Redshift

1. Aprire la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.

2. Sul pannello di controllo Amazon ML, sotto Entità, scegli Creazione di un nuovo... e quindi scegliere L'origine dati.
 3. Sul Dati di input, scegliere Amazon Redshift.
 4. Nella procedura guidata Crea origine dati, per Cluster identifier (Identificatore cluster), digitare il nome del cluster.
 5. Per Nome del database digitare il nome del database Amazon Redshift.
 6. Per Database user name (Nome utente database), digitare il nome utente del database.
 7. Per Database password (Password database), digitare la password del database.
 8. Per IAM role (Ruolo IAM), scegliere il proprio ruolo IAM. Se non ne hai ancora uno, scegliere Crea un nuovo ruolo. Amazon ML crea un ruolo IAM Amazon Redshift per te.
 9. Per testare le impostazioni di Amazon Redshift, scegli Accesso ai test (accanto a Ruolo IAM). Se Amazon ML non è in grado di connettersi ad Amazon Redshift con le impostazioni fornite, non è possibile proseguire nella creazione di un'origine dati. Per la risoluzione dei problemi, consultare [Risoluzione degli errori](#).
 10. Per SQL query (Query SQL), digitare la query SQL.
 11. Per Posizione dello schema, scegliere se si vuole che Amazon ML crei uno schema per l'utente. Se è già stato creato una schema, digitare il percorso Amazon S3 per i file dello schema.
 12. Per Posizione temporanea di Amazon S3, digitare il percorso Amazon S3 al bucket in cui si desidera che Amazon ML archivi i dati che scarica da Amazon Redshift.
 13. (Facoltativo) Per Datasource name (Nome origine dati), digitare un nome per l'origine dati.
 14. Selezionare Verify (Verifica). Amazon ML verifica se riesce a connettersi al database Amazon Redshift.
 15. Nella pagina Schema, esaminare i tipi di dati per tutti gli attributi e correggerli se necessario.
 16. Scegliere Continue (Continua).
 17. Se si vuole utilizzare questa origine dati per creare o valutare un modello ML, in Do you plan to use this dataset to create or evaluate an ML model? (Utilizzare questo set di dati per creare o valutare un modello ML?), scegliere Yes (Sì). Se è stato selezionato Yes (Sì), scegliere la riga di destinazione. Per informazioni sulle destinazioni, consultare la pagina [Utilizzo del campo targetAttributeName](#).
- Se si intende utilizzare questa origine dati insieme a un modello già creato per generare previsioni, scegliere No.
18. Scegliere Continue (Continua).

19. Per Does your data contain an identifier? (I dati contengono un identificatore?), se i tuoi dati non contengono un identificatore riga, scegliere No.

Se i dati contengono un identificatore di riga, scegliere Yes (Sì). Per ulteriori informazioni sugli identificatori di riga, consultare la pagina [Utilizzo del campo rowID](#).

20. Scegliere Review (Rivedi).

21. Nella pagina Review (Rivedi), rivedere le impostazioni e scegliere Finish (Fine).

Dopo aver creato un'origine dati, è possibile utilizzarla per [create an ML model](#). Se è già stato creato un modello, è possibile utilizzare l'origine dati per [evaluate an ML model](#) o [generate predictions](#).

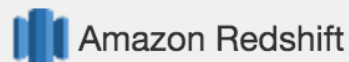
Copia di un'origine dati (console)

Quando si desidera creare un'origine dati simile a una già esistente, è possibile utilizzare la console di Amazon ML per copiare l'origine dati originale e modificarne le impostazioni. Ad esempio, è possibile scegliere di iniziare con un'origine dati esistente e quindi modificare lo schema di dati perché si abbini meglio ai propri dati; oppure specificare una query SQL utilizzata per scaricare i dati da Amazon Redshift oppure specificare un'altra AWS Identity and Access Management (IAM) utente per accedere al cluster Amazon Redshift.

Per copiare e modificare un'origine dati di Amazon Redshift

1. Aprire la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Sul pannello di controllo Amazon ML, sotto Entità, scegli Creazione di un nuovo... e quindi scegliere L'origine dati.
3. Sul Dati di input, per Dove sono i tuoi dati?, scegli Amazon Redshift. Se si dispone già di un'origine dati creata con i dati Amazon Redshift, è possibile copiare le impostazioni da un'altra origine dati.

Where is your data?



Do you want to copy the settings from another Amazon Redshift datasource to create a new datasource? To copy settings, choose [Find a datasource](#).

Se non si dispone già di un'origine dati creata con i dati Amazon Redshift, questa opzione non viene visualizzata.

4. Scegliere Find a datasource (Trova un'origine dati).
5. Selezionare l'origine dati che si desidera copiare e scegliere Copia delle impostazioni. Amazon ML popola automaticamente la maggior parte delle impostazioni dell'origine dati con le impostazioni dell'origine dati originale. Non copia la password del database, la posizione dello schema o il nome dell'origine dati dall'origine dati originale.
6. È possibile modificare le impostazioni inserite automaticamente. Ad esempio, se si desidera modificare i dati che Amazon ML scarica da Amazon Redshift, modificare la query SQL.
7. Per Database password (Password database), digitare la password del database. Amazon ML non archivia né riutilizza la password, perciò è sempre necessario fornirla.
8. (Opzionale) Per Posizione dello schema, Amazon ML preseleziona Voglio che Amazon ML generi uno schema consigliato per l'utente. Se è già stato creato uno schema, scegliere Voglio utilizzare lo schema che ho creato e archiviato in Amazon S3 e digitare il percorso per il file dello schema in Amazon S3.
9. (Facoltativo) Per Datasource name (Nome origine dati), digitare un nome per l'origine dati. In caso contrario, Amazon ML genera un nuovo nome di origine dati per l'utente.
10. Selezionare Verify (Verifica). Amazon ML verifica se riesce a connettersi al database Amazon Redshift.
11. (Facoltativo) Se Amazon ML ha dedotto lo schema per te, sul Schema, esaminare i tipi di dati per tutti gli attributi e correggerli se necessario.
12. Scegliere Continue (Continua).
13. Se si vuole utilizzare questa origine dati per creare o valutare un modello ML, in Do you plan to use this dataset to create or evaluate an ML model? (Utilizzare questo set di dati per creare o valutare un modello ML?), scegliere Yes (Sì). Se è stato selezionato Yes (Sì), scegliere la riga di destinazione. Per informazioni sulle destinazioni, consultare la pagina [Utilizzo del campo targetAttributeName](#).

Se si intende utilizzare questa origine dati insieme a un modello già creato per generare previsioni, scegliere No.

14. Scegliere Continue (Continua).
15. Per Does your data contain an identifier? (I dati contengono un identificatore?), se i tuoi dati non contengono un identificatore riga, scegliere No.

Se i dati contengono un identificatore riga, scegliere Yes (Sì) e selezionare la riga che si desidera utilizzare come un identificatore. Per ulteriori informazioni sugli identificatori di riga, consultare la pagina [Utilizzo del campo rowID](#).

16. Scegliere Review (Rivedi).
17. Esaminare le impostazioni e scegliere Finish (Fine).

Dopo aver creato un'origine dati, è possibile utilizzarla per [create an ML model](#). Se è già stato creato un modello, è possibile utilizzare l'origine dati per [evaluate an ML model](#) o [generate predictions](#).

Risoluzione dei problemi di Amazon Redshift

Quando crei la sorgente dati, i modelli ML e la valutazione di Amazon Redshift, Amazon Machine Learning (Amazon ML) riporta lo stato dei tuoi oggetti Amazon ML nella console Amazon ML. Se Amazon ML restituisce messaggi di errore, utilizza le seguenti informazioni e risorse per risolvere i problemi.

Per risposte alle domande generali relative ad Amazon ML, consulta la [Domande frequenti di Amazon Machine Learning](#). Puoi anche cercare risposte e pubblicare domande nel [Forum Amazon Machine Learning](#).

Argomenti

- [Risoluzione degli errori](#)
- [Contattare AWS Support](#)

Risoluzione degli errori

Il formato del ruolo non è valido. Fornire un ruolo IAM valido. Ad esempio `arn:aws:aws:aws:aws:aws:aws:awYourAccountID:ruolo/YourRedshiftRole`.

Causa

Il formato dell'ARN (Amazon Resource Name) del ruolo IAM non è corretto.

Soluzione

Nella procedura guidata Crea origine dati, correggere l'ARN per il ruolo. Per informazioni sulla formattazione degli ARN dei ruoli, consulta [ARN IAM](#) nell'IAM User Guide. L'area è facoltativa per gli ARN dei ruoli IAM.

Il ruolo non è valido. Amazon ML non può assumere il <role ARN>ruolo IAM. Fornisci un ruolo IAM valido e rendilo accessibile ad Amazon ML.

Causa

Il tuo ruolo non è impostato per consentire ad Amazon ML di assumerlo.

Soluzione

Nel [Consolle IAM](#), modifica il tuo ruolo in modo che abbia una politica di attendibilità che consenta ad Amazon ML di assumere il ruolo ad esso associato.

Questo utente <user ARN> non è autorizzato a trasferire il ruolo IAM <role ARN>.

Causa

Il tuo utente IAM non dispone di una politica di autorizzazioni che gli consenta di passare un ruolo ad Amazon ML.

Soluzione

Allega una policy di autorizzazioni al tuo utente IAM che ti consenta di trasferire ruoli ad Amazon ML. È possibile associare una policy di autorizzazioni all'utente IAM nella [console IAM](#).

Non è consentito trasferire un ruolo IAM tra account diversi. Il ruolo IAM deve appartenere a questo account.

Causa

Non è possibile trasferire un ruolo appartenente a un altro account IAM.

Soluzione

Accedere all'account AWS utilizzato per creare il ruolo. È possibile visualizzare i ruoli IAM nella [console IAM](#).

Il ruolo specificato non dispone delle autorizzazioni per eseguire l'operazione. Fornisci un ruolo con una politica che fornisca ad Amazon ML le autorizzazioni richieste.

Causa

Il ruolo IAM non dispone delle autorizzazioni per eseguire l'operazione richiesta.

Soluzione

Modificare la policy delle autorizzazioni associata al ruolo nella [console IAM](#) per fornire le autorizzazioni richieste.

Amazon ML non può configurare un gruppo di sicurezza su quel cluster Amazon Redshift con il ruolo IAM specificato.

Causa

Il tuo ruolo IAM non dispone delle autorizzazioni necessarie per configurare un cluster di sicurezza Amazon Redshift.

Soluzione

Modificare la policy delle autorizzazioni associata al ruolo nella [console IAM](#) per fornire le autorizzazioni richieste.

Si è verificato un errore quando Amazon ML ha tentato di configurare un gruppo di sicurezza sul cluster. Riprova più tardi.

Causa

Quando Amazon ML ha provato a connettersi al cluster Amazon Redshift, si è verificato un problema.

Soluzione

Verificare che il ruolo IAM fornito nella procedura guidata Crea origine dati disponga di tutte le autorizzazioni necessarie.

Il formato dell'ID del cluster non è valido. L'ID del cluster deve iniziare con una lettera e deve contenere solo caratteri alfanumerici e trattini. Non può contenere due trattini consecutivi o terminare con un trattino.

Causa

Il formato dell'ID del cluster Amazon Redshift non è corretto.

Soluzione

Nella procedura guidata Crea origine dati, correggere l'ID del cluster in modo che contenga solo caratteri alfanumerici e trattini e non contenga due trattini consecutivi o termini con un trattino.

Non esiste un <Amazon Redshift cluster name>cluster o il cluster non si trova nella stessa regione del servizio Amazon ML. Specificare un cluster nella stessa Regione di Amazon ML.

Causa

Amazon ML non riesce a trovare il tuo cluster Amazon Redshift perché non si trova nella regione in cui stai creando un'origine dati Amazon ML.

Soluzione

Verifica che il cluster esista sulla console Amazon Redshift [Cluster](#) pagina, che stai creando un'origine dati nella stessa regione in cui si trova il cluster Amazon Redshift e che l'ID del cluster specificato nella procedura guidata Create Datasource è corretto.

Amazon ML non è in grado di leggere i dati nel cluster Amazon Redshift. Fornisci l'ID del cluster Amazon Redshift corretto.

Causa

Amazon ML non è in grado di leggere i dati nel cluster Amazon Redshift specificato.

Soluzione

Nella procedura guidata Create Datasource, specifica l'ID del cluster Amazon Redshift corretto, verifica di creare un'origine dati nella stessa regione in cui si trova il cluster Amazon Redshift e che il cluster sia elencato su Amazon Redshift [Cluster](#) pagina.

Il <Amazon Redshift cluster name>cluster non è accessibile pubblicamente.

Causa

Amazon ML non può accedere al tuo cluster perché il cluster non è accessibile al pubblico e non dispone di un indirizzo IP pubblico.

Soluzione

Rendere il cluster accessibile pubblicamente e assegnargli un indirizzo IP pubblico. Per informazioni su come rendere i cluster accessibili al pubblico, vedere [Modifica di un cluster](#) nel Guida alla gestione di Amazon Redshift.

<Redshift>Lo stato del cluster non è disponibile per Amazon ML. Usa la console Amazon Redshift per visualizzare e risolvere questo problema di stato del cluster. Lo stato del cluster deve essere "disponibile".

Causa

Amazon ML non riesce a vedere lo stato del cluster.

Soluzione

Assicurarsi che il cluster sia disponibile. Per informazioni sulla verifica dello stato del cluster, consulta [Ottenere una panoramica dello stato del cluster](#) nel Guida alla gestione di Amazon Redshift. Per informazioni sul riavvio del cluster in modo che sia disponibile, vedere [Riavvio di un cluster](#) nel Guida alla gestione di Amazon Redshift.

Non vi è alcun database <database name> in questo cluster. Verificare che il nome del database sia corretto oppure specificare un altro cluster e database.

Causa

Amazon ML non è in grado di individuare il database specificato nel cluster specificato.

Soluzione

Verificare che il nome del database inserito nella procedura guidata Crea origine dati sia corretto, oppure specificare i nomi corretti del cluster e del database.

Amazon ML non è riuscito ad accedere al tuo database. Fornire una password valida per l'utente di database <user name>.

Causa

La password che hai fornito nella procedura guidata Create Datasource per consentire ad Amazon ML di accedere al tuo database Amazon Redshift non è corretta.

Soluzione

Fornisci la password corretta per l'utente del tuo database Amazon Redshift.

Si è verificato un errore durante il tentativo di convalida della query da parte di Amazon ML.

Causa

Esiste un problema con la query SQL.

Soluzione

Verificare che la query SQL sia valida.

Si è verificato un errore durante l'esecuzione della query SQL. Verificare il nome del database e la query fornita. Causa principale: {serverMessage}.

Causa

Amazon Redshift non è stato in grado di eseguire la tua query.

Soluzione

Verificare che il nome del database specificato nella procedura guidata Crea origine dati sia corretto e che la query SQL sia valida.

Si è verificato un errore durante l'esecuzione della query SQL. Causa principale: {serverMessage}.

Causa

Amazon Redshift non è riuscito a trovare la tabella specificata.

Soluzione

Verifica che la tabella specificata nella procedura guidata Create Datasource sia presente nel database del cluster Amazon Redshift e di aver inserito l'ID cluster, il nome del database e la query SQL corretti.

Contattare AWS Support

Se si dispone di AWS Premium Support, è possibile creare una richiesta di supporto tecnico presso il [Centro AWS Support](#).

Utilizzo dei dati da un database Amazon RDS per creare un'origine dati Amazon ML

Amazon ML consente di creare un oggetto origine dati dai dati archiviati in un database MySQL in Amazon Relational Database Service (Amazon RDS). Quando si esegue questa operazione, Amazon ML crea un oggetto AWS Data Pipeline che esegue la query SQL specificata e posiziona l'output in un bucket S3 scelto dall'utente. Amazon ML usa tali dati per creare l'origine dati.

 Note

Amazon ML supporta solo i database MySQL in VPC.

Prima che Amazon ML possa leggere i dati di input, è necessario esportare tali dati in Amazon Simple Storage Service (Amazon S3). È possibile configurare Amazon ML affinché esegua l'esportazione utilizzando l'API. (RDS è limitato all'API e non è disponibile dalla console).

Affinché Amazon ML si connetta al database MySQL in Amazon RDS e legga i dati per conto dell'utente, è necessario fornire quanto segue:

- L'identificatore istanza database RDS
- Il nome del database MySQL
- LaAWS Identity and Access Managementruolo (IAM) utilizzato per creare, attivare ed eseguire la data pipeline
- Le credenziali utente di database:
 - Nome utente
 - Password
- Le informazioni di protezione di AWS Data Pipeline:
 - Il ruolo delle risorse IAM
 - Il ruolo del servizio IAM
- Le informazioni relative alla sicurezza di Amazon RDS:
 - L'ID sottorete
 - Gli ID gruppo di sicurezza
- La query SQL che specifichi i dati che si desidera utilizzare per creare l'origine dati
- Il percorso di output S3 (bucket) utilizzato per memorizzare i risultati della query
- (Facoltativo) La posizione del file dello schema dati

Inoltre, è necessario accertarsi che gli utenti o i ruoli IAM che creano origini dati Amazon RDS utilizzando l'[CreateDataSourceFromRDS](#)operazione ha iliam:PassRoleautorizzazione. Per ulteriori informazioni, consultare [Controllo dell'accesso alle risorse Amazon ML con IAM](#).

Argomenti

- [Identificatore di istanza di database RDS](#)
- [Nome database MySQL](#)

- [Credenziali utente di database](#)
- [Informazioni di protezione AWS Data Pipeline](#)
- [Informazioni sulla sicurezza di Amazon RDS](#)
- [Query SQL MySQL](#)
- [Percorso di output S3](#)

Identificatore di istanza di database RDS

L'identificatore di istanze DB RDS è un nome univoco fornito dall'utente che identifica l'istanza di database che Amazon ML deve utilizzare durante l'interazione con Amazon RDS. Si può trovare l'identificatore di istanza di database di RDS nella console di Amazon RDS.

Nome database MySQL

Il nome database MySQL specifica il nome del database MySQL nell'istanza di database RDS.

Credenziali utente di database

Per connettersi all'istanza database RDS, è necessario specificare il nome utente e la password dell'utente di database che dispone di autorizzazioni sufficienti per eseguire la query SQL fornita.

Informazioni di protezione AWS Data Pipeline

Per abilitare l'accesso sicuro ad AWS Data Pipeline, è necessario fornire i nomi del ruolo delle risorse IAM e del ruolo di servizio IAM.

Un'istanza EC2 assume il ruolo di risorsa per copiare i dati da Amazon RDS ad Amazon S3. Il modo più semplice per creare questo ruolo di risorsa è con il modello `DataPipelineDefaultResourceRole`, inserendo **machinelearning.aws.com** come servizio attendibile. Per ulteriori informazioni sul modello, consultare la pagina relativa all'[impostazione di ruoli IAM](#) nella AWS Data Pipeline Developer Guide.

Se crei il tuo ruolo, deve avere i seguenti contenuti:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
```

```
    "Principal": {
      "Service": "machinelearning.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "StringEquals": { "aws:SourceAccount": "123456789012" },
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:datasource/*" }
    }
  }]
}
```

AWS Data Pipeline assume il ruolo di servizio per monitorare l'avanzamento della copiatura dei dati da Amazon RDS ad Amazon S3. Il modo più semplice per creare questo ruolo di risorsa è con il modello `DataPipelineDefaultRole`, inserendo `machinelearning.aws.com` come servizio attendibile. Per ulteriori informazioni sul modello, consultare la pagina relativa all'[impostazione di ruoli IAM](#) nella AWS Data Pipeline Developer Guide.

Informazioni sulla sicurezza di Amazon RDS

Per abilitare l'accesso sicuro ad Amazon RDS, è necessario fornire l'`VPC Subnet ID` e `RDS Security Group IDs`. È inoltre necessario configurare regole di ingresso appropriate per la sottorete VPC a cui punta il parametro `Subnet ID` e fornire l'ID del gruppo di sicurezza che dispone di questa autorizzazione.

Query SQL MySQL

Il parametro `MySQL SQL Query` specifica la query SQL `SELECT` che si desidera eseguire sul database MySQL. I risultati della query vengono copiati nel percorso di output S3 (bucket) specificato dall'utente.

Note

La tecnologia di Machine Learning funziona meglio quando i record di input si presentano in ordine casuale (mischiat). È possibile mischiare i risultati della query SQL MySQL utilizzando la funzione `rand()`. Ad esempio, supponiamo che questa sia la query originale:

```
"SELECT col1, col2,... FROM training_table"
```

È possibile aggiungere un mescolamento casuale aggiornando la query in questo modo:

```
"SELECT col1, col2, ... FROM training_table ORDER BY rand()"
```

Percorso di output S3

La `S3 Output Location` specifica il nome del percorso «di gestione temporanea» Amazon S3 in cui conduce i risultati della query SQL MySQL.

Note

È necessario accertarsi che Amazon ML disponga delle autorizzazioni per leggere i dati da questa posizione una volta che i dati saranno esportati da Amazon RDS. Per informazioni su come impostare le autorizzazioni, consultare la pagina relativa a come concedere ad Amazon ML le autorizzazioni per leggere i dati su Amazon S3.

Addestramento dei modelli ML

Per il processo di addestramento di un modello ML occorre fornire a un algoritmo ML (ovvero l'algoritmo di apprendimento) i dati di addestramento da cui possa apprendere. Il termine modello ML si riferisce all'artefatto del modello creato dal processo di addestramento.

I dati di addestramento devono contenere la risposta corretta, che è nota come target o attributo di destinazione. L'algoritmo di apprendimento trova nei dati di addestramento i pattern che mappano gli attributi dei dati di input al target (la risposta che si desidera prevedere) e genera un modello ML che acquisisce questi pattern.

È possibile utilizzare il modello ML per ottenere previsioni su nuovi dati di cui non si conosce il target. Ad esempio, supponiamo di voler addestrare un modello ML per prevedere se un'e-mail è spam o non spam. È possibile fornire ad Amazon ML dati di addestramento contenenti e-mail di cui è noto il target (ovvero un'etichetta che indica se un'e-mail è spam o non spam). Amazon ML addestrerà un modello ML utilizzando questi dati, generando un modello che tenta di prevedere se le nuove e-mail saranno spam o non spam.

Per informazioni generali sui modelli ML e gli algoritmi ML, consultare [Concetti di Machine Learning](#).

Argomenti

- [Tipi di modelli ML](#)
- [Processo di addestramento](#)
- [Parametri di addestramento](#)
- [Creazione di un modello ML](#)

Tipi di modelli ML

Amazon ML supporta tre tipi di modelli ML: di classificazione binaria, di classificazione multiclasse e di regressione. Il tipo di modello che è opportuno scegliere dipende dal tipo di destinazione che si desidera prevedere.

Modello di classificazione binario

I modelli ML per i problemi di classificazione binaria prevedono un esito binario (una tra due possibili classi). Per addestrare i modelli di classificazione binaria, Amazon ML utilizza l'algoritmo di apprendimento standard del settore noto come regressione logistica.

Esempi di problemi di classificazione binaria

- "Questa e-mail è spam o non spam?"
- "Il cliente acquisterà questo prodotto?"
- "Questo prodotto è un libro o un animale da fattoria?"
- "Questa recensione è scritta da un cliente o da un robot?"

Modello di classificazione multiclasse

I modelli ML per problemi di classificazione multiclasse consentono di generare previsioni per più classi (per prevedere uno tra più di due esiti). Per addestrare i modelli di classificazione multiclasse, Amazon ML utilizza l'algoritmo di apprendimento standard del settore noto come regressione logistica multinomiale.

Esempi di problemi multiclasse

- "Questo prodotto è un libro, un film o un abito?"
- "Questo film è una commedia romantica, un documentario o un thriller?"
- "Quale categoria di prodotti è più interessante per questo cliente?"

Modello di regressione

I modelli ML per i problemi di regressione prevedono un valore numerico. Per addestrare i modelli di regressione, Amazon ML utilizza l'algoritmo di apprendimento standard del settore noto come regressione lineare.

Esempi di problemi di regressione

- "Che temperatura ci sarà domani a Seattle?"
- "Per questo prodotto, quante unità si venderanno?"
- "A che prezzo sarà venduta questa casa?"

Processo di addestramento

Per addestrare un modello ML, è necessario specificare quanto segue:

- Origine dati di input per l'addestramento

- Nome dell'attributo dati che contiene la destinazione che deve essere prevista
- Istruzioni per le trasformazioni di dati richieste
- Parametri di addestramento per controllare l'algoritmo di apprendimento

Durante il processo di addestramento, Amazon ML seleziona automaticamente il corretto algoritmo di apprendimento per l'utente, in base al tipo di destinazione specificata nell'origine dati di addestramento.

Parametri di addestramento

Di solito, gli algoritmi di machine learning accettano parametri che possono essere utilizzati per controllare determinate proprietà del processo di addestramento e del modello ML risultante. In Amazon Machine Learning, questi sono chiamati parametri di addestr. È possibile impostare questi parametri utilizzando la console di Amazon ML, l'API o l'interfaccia a riga di comando (CLI). Se non si impostano i parametri, Amazon ML utilizzerà valori di default adatti a un'ampia gamma di attività di machine learning.

È possibile specificare i valori per i seguenti parametri di addestramento:

- Massima dimensione del modello
- Numero massimo di passate sui dati di addestramento
- Tipo di mescolamento
- Tipo di regolarizzazione
- Quantità di regolarizzazione

Nella console di Amazon ML, i parametri di addestramento sono impostati di default. Le impostazioni di default sono adatte alla maggior parte dei problemi di ML, ma è possibile scegliere altri valori per ottimizzare le prestazioni. Altri parametri di addestramento, come ad esempio la velocità di apprendimento, sono configurati in base ai dati dell'utente.

Nelle seguenti sezioni vengono fornite ulteriori informazioni sui parametri di addestramento.

Massima dimensione del modello

La massima dimensione del modello è la dimensione totale, in unità di byte, dei modelli che Amazon ML crea durante l'addestramento di un modello ML.

Per impostazione predefinita, Amazon ML crea un modello di 100 MB. È possibile istruire Amazon ML affinché crei un modello più piccolo o più grande specificando una dimensione diversa. Per la gamma di dimensioni disponibili, consultare [Tipi di modelli ML](#).

Se Amazon ML non è in grado di trovare un numero sufficiente di pattern per riempire le dimensioni del modello, crea un modello più piccolo. Ad esempio, se si specifica un modello di dimensioni massime di 100 MB, ma Amazon ML trova pattern solo per un totale di 50 MB, il modello risultante sarà di 50 MB. Se Amazon ML individua più pattern adatti alla dimensione specificata, applica un limite massimo tagliando i pattern che influenzano di meno la qualità del modello appreso.

La scelta della dimensione del modello consente di trovare un compromesso tra la qualità predittiva di un modello e il costo di utilizzo. Modelli più piccoli possono far sì che Amazon ML rimuova molti pattern per rientrare nel limite massimo della dimensione, influenzando la qualità delle previsioni. Modelli più grandi, invece, hanno costi di esecuzione delle query più elevati per le previsioni in tempo reale.

Note

Se si utilizza un modello ML per generare previsioni in tempo reale, verrà addebitato un piccolo costo di prenotazione di capacità determinato dalla dimensione del modello. Per ulteriori informazioni, consultare [Prezzi di Amazon ML](#).

Set di dati di input più grandi non comportano necessariamente modelli più grandi, perché i modelli archiviano i pattern, non i dati di input; se i pattern sono pochi e semplici, il modello risultante sarà piccolo. Dati di input con un numero elevato di attributi grezzi (colonne di input) o di caratteristiche derivate (output delle trasformazioni dei dati Amazon ML) avranno probabilmente più pattern trovati e memorizzati durante il processo di addestramento. La scelta della corretta dimensione dei modelli per i propri dati è un problema che è meglio affrontare con alcuni esperimenti. Il log di addestramento del modello Amazon ML (che è possibile scaricare dalla console o tramite l'API) contiene messaggi relativi a quanto il modello deve essere (eventualmente) tagliato durante il processo di addestramento, consentendo di calcolare la potenziale qualità di previsione.

Numero massimo di passate sui dati

Per ottimizzare i risultati, Amazon ML potrebbe dover fare più passate sui dati per rilevare i pattern. Per impostazione predefinita, Amazon ML fa 10 passate, ma è possibile modificare l'impostazione predefinita impostando un numero fino a 100. Amazon ML tiene traccia della qualità dei pattern

(convergenza del modello) mentre prosegue e interrompe automaticamente l'addestramento quando non vi sono più punti di dati o pattern da rilevare. Ad esempio, se si imposta il numero di passate su 20, ma Amazon ML scopre che non è possibile rilevare nuovi pattern alla fine di 15 passate, interrompe l'addestramento a 15 passate.

In generale, set di dati con poche osservazioni richiedono di norma più passate sui dati per ottenere un modello di qualità più elevata. Set di dati più grandi contengono spesso molti punti di dati simili, eliminando la necessità di un numero elevato di passate. L'impatto della scelta di più passate sui dati è duplice: l'addestramento del modello richiede più tempo e i costi sono più elevati.

Tipo di mescolamento dei dati di addestramento

In Amazon ML, è necessario mischiare i dati di addestramento. Il mescolamento consente di mischiare l'ordine dei dati in modo che l'algoritmo SGD non incontri un solo tipo di dati per troppe osservazioni in successione. Ad esempio, se si sta addestrando un modello ML per prevedere un tipo di prodotto e i dati di addestramento includono tipi di prodotti film, giocattoli e videogiochi, se i dati sono stati ordinati in base alla colonna tipo di prodotto prima di caricarli, l'algoritmo vede i dati alfabeticamente in base al tipo di prodotto. L'algoritmo vede per primi tutti i dati dei film e il modello ML inizia ad apprendere i pattern relativi ai film. Quindi, quando il modello rileva i dati sui giocattoli, ogni aggiornamento che l'algoritmo effettua adatta il modello al tipo di prodotto giocattoli, anche se gli aggiornamenti degradano i pattern che si adattano ai film. Questo improvviso passaggio dal tipo film al tipo giocattoli può produrre un modello che non apprende come prevedere in modo accurato i tipi di prodotto.

È necessario mischiare i dati di addestramento anche se si sceglie l'opzione di divisione casuale quando si divide l'origine dati di input in parti di addestramento e di valutazione. La strategia di divisione casuale sceglie un sottoinsieme di dati casuale per ogni origine dati, ma non modifica l'ordine delle righe nell'origine dati. Per ulteriori informazioni sulla divisione dei dati, consultare [Divisione dei dati](#).

Quando si crea un modello ML utilizzando la console, per impostazione predefinita Amazon ML meschia i dati con una tecnica pseudocasuale. Indipendentemente dal numero di passate richieste, Amazon ML meschia i dati solo una volta prima di addestrare il modello ML. Se si mischiano i dati prima di fornirli ad Amazon ML e non si vuole che Amazon ML mischi di nuovo i dati, è possibile impostare il tipo di mescolamento a none. Ad esempio, se hai mischiato casualmente i record nel file con estensione csv prima di caricarlo su Amazon S3, utilizza il `rand()` funzione nella query MySQL SQL durante la creazione dell'origine dati da Amazon RDS o si utilizza il `irandom()` funzione nella query SQL di Amazon Redshift durante la creazione dell'origine dati da Amazon Redshift,

impostazione Tipo di mescolamento `ononenon` influirà sulla precisione predittiva del tuo modello ML. Se si mischiano i dati solo una volta si riduce il runtime e il costo di creazione di un modello ML.

Important

Quando si crea un modello ML utilizzando l'API di Amazon ML, Amazon ML non mischia i dati per impostazione predefinita. Se si utilizza l'API al posto della console per creare il modello ML, è consigliabile mischiare i dati impostando il parametro `sgd.shuffleType` su `auto`.

Tipo e quantità di regolarizzazione

Le prestazioni predittive dei modelli ML complessi (quelli con molti attributi di input) ne risentono quando i dati contengono troppi pattern. Con l'aumento del numero di pattern cresce anche la probabilità che il modello apprenda artefatti di dati involontari invece di pattern di dati reali. In tal caso, il modello ha ottime prestazioni sui dati di addestramento, ma non è in grado di effettuare una generalizzazione corretta sui nuovi dati. Questo fenomeno è noto come *overfitting* dei dati di addestramento.

La regolarizzazione aiuta a evitare l'*overfitting* nei modelli lineari su esempi di dati di addestramento penalizzando i valori di ponderazione estremi. La regolarizzazione L1 riduce il numero di caratteristiche utilizzate nel modello spingendo a zero la ponderazione delle caratteristiche che in caso contrario avrebbero ponderazioni molto piccole. La regolarizzazione L1 produce modelli di tipo sparse e riduce la quantità di disturbo nel modello. La regolarizzazione L2 comporta valori di ponderazione globale più piccoli, il che stabilizza le ponderazioni quando vi è un'elevata correlazione tra le caratteristiche. È possibile controllare la quantità di regolarizzazione L1 o L2 utilizzando il parametro `Regularization amount`. Se si specifica un valore estremamente elevato di `Regularization amount` tutte le caratteristiche possono avere una ponderazione zero.

La selezione e il tuning del valore di regolarizzazione ottimale è un argomento attivamente discusso nell'ambito della ricerca sul machine learning. È probabilmente vantaggioso selezionare una quantità moderata di regolarizzazione L2, che è l'impostazione predefinita nella console di Amazon ML. Gli utenti esperti possono scegliere tra tre tipi di regolarizzazione (none, L1 o L2) e quantità. Per ulteriori informazioni sulla regolarizzazione, consultare la pagina [Regolarizzazione \(matematica\)](#).

parametri di addestramento: Tipi e valori predefiniti

La tabella seguente elenca i parametri di addestramento di Amazon ML, insieme ai valori di default e all'intervallo ammesso per ciascuno.

Parametro di addestramento	Type (Tipo)	Valore predefinito	Descrizione
maxMLMode ISizeInBytes	Numero intero	100.000.000 byte (100 MiB)	<p>Range consentito: 100.000 (100 KiB) a 2.147.483.648 (2 GiB)</p> <p>A seconda dei dati di input, la dimensione del modello potrebbe influenzare le prestazioni.</p>
sgd.maxPasses	Numero intero	10	Range consentito: 1-100
sgd.shuffleType	Stringa	auto	Valori accettabili: auto o none
sgd.l1RegularizationAmount	Doppio	0 (per impostazione predefinita, L1 non viene utilizzato)	<p>Range consentito: 0 per MAX_DOUBLE</p> <p>È emerso che valori L1 compresi tra 1E-4 e 1E-8 producono buoni risultati. Valori più elevati possono produrre modelli non molto utili.</p> <p>Non è possibile impostare sia L1 sia L2. È necessario scegliere l'uno o l'altro.</p>
sgd.l2RegularizationAmount	Doppio	1E-6 (Per impostazione predefinita, L2 viene utilizzato con questa quantità di regolarizzazione)	<p>Range consentito: 0 per MAX_DOUBLE</p> <p>È emerso che valori L2 compresi tra 1E-2 e 1E-6 producono buoni risultati. Valori più elevati possono produrre modelli non molto utili.</p> <p>Non è possibile impostare sia L1 sia L2. È necessario scegliere l'uno o l'altro.</p>

Creazione di un modello ML

Dopo aver creato un'origine dati, è possibile creare un modello ML. Se si utilizza la console Amazon Machine Learning per creare un modello, è possibile scegliere di utilizzare le impostazioni predefinite o personalizzare il modello applicando le opzioni personalizzate.

Le opzioni personalizzate includono:

- **Impostazioni di valutazione:** È possibile scegliere che Amazon ML riservi una parte dei dati di input per valutare la qualità predittiva del modello ML. Per ulteriori informazioni sulle valutazioni, consultare la pagina relativa alla [valutazione dei modelli ML](#).
- **Composizione:** Una composizione dice ad Amazon ML quali attributi e trasformazioni di attributi sono disponibili per il modello di addestramento. Per informazioni sulle ricette Amazon ML, consulta [Trasformazioni delle caratteristiche con le composizioni dati](#).
- **Parametro di addestramento** I parametri controllano alcune proprietà del processo di addestramento e del modello ML risultante. Per ulteriori informazioni sui parametri di addestramento, consultare [Parametri di addestramento](#).

Per selezionare o specificare i valori per queste impostazioni, scegliere l'opzione Personalizza quando si utilizza la procedura guidata Crea un modello ML. Se si desidera che Amazon ML applichi le impostazioni predefinite, scegliere Default.

Quando si crea un modello ML, Amazon ML seleziona il tipo di algoritmo di apprendimento che utilizzerà in base al tipo di attributo di destinazione. L'attributo di destinazione è l'attributo che contiene le risposte "corrette". Se l'attributo di destinazione è Binary, Amazon ML crea un modello di classificazione binaria che utilizza l'algoritmo di regressione logistica. Se l'attributo di destinazione è Categorical, Amazon ML crea un modello multiclasse che utilizza un algoritmo di regressione logistica multinomiale. Se l'attributo di destinazione è Numeric, Amazon ML crea un modello di regressione che utilizza un algoritmo di regressione lineare.

Argomenti

- [Prerequisiti](#)
- [Creazione di un modello ML con opzioni predefinite](#)
- [Creazione di un modello ML con opzioni personalizzate](#)

Prerequisiti

Prima di utilizzare la console Amazon ML per creare un modello ML, è necessario creare due origini dati, una per addestrare il modello e una per valutare il modello. Se non sono ancora state create due origini dati, consultare [Fase 2: Creare un'origine dati di addestramento](#) nel tutorial.

Creazione di un modello ML con opzioni predefinite

Seleziona le opzioni predefinite se si desidera che Amazon ML:

- Divida i dati di input per utilizzare il primo 70% per l'addestramento e il restante 30% per la valutazione
- Suggerisca una composizione basata sulle statistiche relative alle origini dati di addestramento, che costituiscono il 70% dell'origine dati di input
- Scelga i parametri di addestramento predefiniti

Per scegliere le opzioni predefinite

1. Nella console Amazon ML scegliere **Amazon Machine Learning** quindi scegliere **Modelli ML**.
2. Nella pagina di riepilogo **ML models (Modelli ML)**, scegliere **Create a new ML model (Crea un nuovo modello ML)**.
3. Nella pagina **Input data (Dati di input)** accertarsi che sia stata selezionata l'opzione **I already created a datasource pointing to my S3 data (Ho già creato un'origine dati che punta ai miei dati S3)**.
4. Nella tabella, scegliere la propria origine dati, quindi **Continua**.
5. Nella pagina **ML model settings (Impostazioni modello ML)**, inserire un nome per **ML model name (Nome modello ML)**.
6. Per **Training and evaluation settings (Impostazioni di addestramento e valutazione)**, accertarsi che sia stato selezionato **Default**.
7. Per **Denomina questa valutazione**, digitare un nome per la valutazione, quindi scegliere **Review (Revisione)**. Amazon ML ignora il resto della procedura guidata e passa al **Review (Revisione)** (Certificato creato).
8. Rivedere i dati, eliminare gli eventuali tag copiati dall'origine dati che non si vuole applicare al modello e alle valutazioni, quindi scegliere **Finish (Fine)**.

Creazione di un modello ML con opzioni personalizzate

La personalizzazione del modello ML consente di:

- Fornire la propria composizione. Per informazioni su come fornire la tua composizione, consultare [Riferimenti relativi al formato delle composizioni](#).
- Scegliere i parametri di addestramento. Per ulteriori informazioni sui parametri di addestramento, consultare [Parametri di addestramento](#).
- Scegliere un rapporto di divisione per l'addestramento/la valutazione diverso da quello predefinito di 70/30 o fornire un'altra origine dati già preparata per la valutazione. Per ulteriori informazioni sulle strategie di divisione, consultare [Divisione dei dati](#).

È anche possibile scegliere i valori predefiniti per qualsiasi di queste impostazioni.

Se è già stato creato un modello utilizzando le opzioni predefinite e si desidera migliorare le prestazioni predittive del modello, utilizzare l'opzione Custom (Personalizzato) per creare un nuovo modello con alcune impostazioni personalizzate. Ad esempio, è possibile aggiungere ulteriori trasformazioni delle caratteristiche alla composizione o aumentare il numero di passate nel parametro di addestramento.

Per creare un modello con opzioni personalizzate

1. Nella console Amazon ML scegliere Amazon Machine Learning e quindi scegliere Modelli ML.
2. Nella pagina di riepilogo ML models (Modelli ML), scegliere Create a new ML model (Crea un nuovo modello ML).
3. Se è già stata creata un'origine dati, nella pagina Input data (Dati di input) scegliere I already created a datasource pointing to my S3 data (Ho già creato un'origine dati che punta ai miei dati S3). Nella tabella, scegliere la propria origine dati, quindi Continua.

Se occorre creare un'origine dati, scegliere My data is in S3, and I need to create a datasource (I miei dati sono in S3 e devo creare un'origine dati), poi Continua. Si viene reindirizzati alla procedura guidata Create a Datasource (Crea un'origine dati). Specificare se i dati sono in S3 o in Redshift, quindi scegliere Verify (Verifica). Completare la procedura di creazione di un'origine dati.

Dopo aver creato un'origine dati, si viene reindirizzati alla fase successiva della procedura guidata Create ML Model (Crea un modello ML).

4. Nella pagina ML model settings (Impostazioni modello ML), inserire un nome per ML model name (Nome modello ML).
5. In Select training and evaluation settings (Seleziona le impostazioni di addestramento e di valutazione), scegliere Custom (Personalizzato) e Continue (Continua).
6. Nella pagina Recipe (Composizione) è possibile [customize a recipe](#). Se non si desidera personalizzare una composizione, Amazon ML ne suggerisce una. Scegliere Continue (Continua).
7. Nella pagina Advanced settings (Impostazioni avanzate) specificare Maximum ML model Size (Dimensione massima del modello ML), Maximum number of data passes (Numero massimo di passate), Shuffle type for training data (Tipo di mescolamento dei dati di addestramento), Regularization type (Tipo di regolarizzazione) e Regularization amount (Quantità di regolarizzazione). Se non si specificano queste impostazioni, Amazon ML utilizza i parametri di addestramento predefiniti.

Per ulteriori informazioni su questi parametri e le loro impostazioni predefinite, consultare [Parametri di addestramento](#).

Scegliere Continue (Continua).

8. Nella pagina Evaluation (Valutazione), specificare se si desidera valutare il modello ML immediatamente. Se non si desidera valutare il modello ML immediatamente, scegliere Review (Rivedi).

Se si desidera valutare il modello ML immediatamente:

- a. In Name this evaluation (Denomina questa valutazione), digitare un nome per la valutazione.
 - b. Per Seleziona dati di valutazione scegliere se si desidera che Amazon ML riservi una parte dei dati di input per la valutazione e, in tal caso, il modo in cui si desidera frazionare l'origine dati oppure scegliere di fornire un'altra origine dati per la valutazione.
 - c. Scegliere Review (Rivedi).
9. Nella pagina Review (Rivedi) modificare le scelte, eliminare gli eventuali tag copiati dall'origine dati che non si vuole applicare al modello e alle valutazioni, quindi scegliere Finish (Fine).

Dopo aver creato il modello, consultare [Fase 4: Esaminare le prestazioni predittive del modello ML e impostare un punteggio soglia](#).

Trasformazioni dei dati per il Machine Learning

I modelli di Machine learning sono utili solo se la qualità dei dati utilizzati per addestrarli è elevata. Una caratteristica chiave di dati di addestramento di qualità è che sono forniti secondo una modalità ottimizzata per l'apprendimento e la generalizzazione. Il processo di raccolta dei dati in questo formato ottimale è noto come trasformazione di caratteristiche.

Argomenti

- [L'importanza della trasformazione delle caratteristiche](#)
- [Trasformazioni delle caratteristiche con le composizioni dati](#)
- [Riferimenti relativi al formato delle composizioni](#)
- [Composizioni suggerite](#)
- [Riferimento per le trasformazioni di dati](#)
- [Riordino dei dati](#)

L'importanza della trasformazione delle caratteristiche

Si prenda, ad esempio, un modello di machine learning il cui compito è decidere se una transazione di carta di credito è fraudolenta o meno. In base alle conoscenze generali delle applicazioni e all'analisi dei dati, è possibile decidere quali campi dati (o caratteristiche) siano importanti ai fini dell'inclusione nei dati di input. Ad esempio, l'importo della transazione, il nome del venditore, l'indirizzo e l'indirizzo del proprietario della carta di credito sono importanti per il processo di apprendimento. Viceversa, un ID transazione casuale non apporta alcuna informazione (se sappiamo che è veramente casuale) e non è utile.

Una volta deciso i campi da includere, si trasformano queste caratteristiche per semplificare il processo di apprendimento. Le trasformazioni aggiungono l'esperienza precedente ai dati di input, consentendo al modello di machine learning di trarre vantaggio da tale esperienza. Ad esempio, il seguente indirizzo del venditore è rappresentato da una stringa:

```
"123 Main Street, Seattle, WA 98101"
```

In sé, l'indirizzo ha un potere espressivo limitato: è utile solo per i pattern di apprendimento associati a quell'indirizzo esatto. La sua suddivisione in base a elementi costitutivi, tuttavia, è in grado di creare caratteristiche aggiuntive come "Indirizzo" (123 Main Street), "Città" (Seattle), "Stato" (WA) e "CAP" (98101). L'algoritmo di apprendimento può raggruppare transazioni più disparate e individuare

pattern più ampi, e forse anche un'esperienza relativa ad alcuni codici postali di venditori con più attività fraudolenta di altri.

Per ulteriori informazioni sull'approccio e il processo relativi alla trasformazione delle caratteristiche, consultare [Concetti di Machine Learning](#).

Trasformazioni delle caratteristiche con le composizioni dati

Esistono almeno due modi per trasformare le caratteristiche prima di creare modelli ML con Amazon ML: è possibile trasformare direttamente i dati di input prima di proporli ad Amazon ML oppure si possono utilizzare le trasformazioni di dati integrate di Amazon ML. È possibile utilizzare le composizioni Amazon ML, ovvero istruzioni preformattate per comuni trasformazioni. Con le composizioni si può procedere come indicato di seguito:

- Scegli da un elenco di comuni trasformazioni di machine learning integrate e applicale alle singole variabili o a gruppi di variabili
- Seleziona le variabili di input e trasformazioni da rendere disponibili per il processo di machine learning

L'utilizzo delle composizioni Amazon ML offre diversi vantaggi. Amazon ML esegue le trasformazioni di dati per l'utente, perciò non è necessario implementarle manualmente. Inoltre, è una soluzione veloce perché Amazon ML applica le trasformazioni durante la lettura dei dati di input e fornisce i risultati per il processo di apprendimento senza il passaggio intermedio del salvataggio dei risultati su disco.

Riferimenti relativi al formato delle composizioni

Le composizioni (recipe) Amazon ML contengono le istruzioni per trasformare i dati nel quadro del processo di machine learning. Le composizioni vengono definite utilizzando una sintassi simile a JSON, ma hanno restrizioni aggiuntive rispetto a quelle normalmente presenti in JSON. Le composizioni hanno le seguenti sezioni, che devono essere visualizzate nell'ordine mostrato qui:

- **Groups (Gruppi)** abilita il raggruppamento di più variabili, per facilitare l'applicazione di trasformazioni. Ad esempio, è possibile creare un gruppo di tutte le variabili che hanno a che fare con le parti di testo libero di una pagina Web (titolo, corpo) e quindi eseguire una trasformazione su tutte le parti contemporaneamente.
- **Assignments (Assegnazioni)** abilita la creazione di variabili intermedie denominate che possono essere riutilizzate in fase di elaborazione.

- **Outputs (Risultati)** definisce quali variabili saranno utilizzate nel processo di apprendimento e quali trasformazioni (se presenti) si applicano a tali variabili.

Gruppi

È possibile definire gruppi di variabili per trasformare collettivamente tutte le variabili all'interno dei gruppi o utilizzare queste variabili per il machine learning senza trasformarle. Per impostazione predefinita, Amazon ML crea i seguenti gruppi per l'utente:

ALL_TEXT, ALL_NUMERIC, ALL_CATEGORICAL, ALL_BINARY: gruppi specifici per tipo, basati su variabili definite nello schema dell'origine dati.

Note

Non è possibile creare un gruppo con ALL_INPUTS.

Queste variabili possono essere utilizzate nella sezione degli output della composizione senza essere definite. È inoltre possibile creare gruppi personalizzati aggiungendo o togliendo variabili da gruppi esistenti, oppure direttamente da una raccolta di variabili. In questo esempio dimostriamo tutti e tre gli approcci e la sintassi per l'assegnazione dei gruppi:

```
"groups": {  
  
  "Custom_Group": "group(var1, var2)",  
  "All_Categorical_plus_one_other": "group(ALL_CATEGORICAL, var2)"  
  
}
```

I nomi dei gruppi devono iniziare con un carattere alfabetico e possono essere costituiti da 1 a 64 caratteri. Se il nome del gruppo non inizia con un carattere alfabetico o se contiene caratteri speciali (, ' "\t\r\n () \), il nome deve essere racchiuso tra virgolette per essere incluso nella composizione.

Assegnazioni

È possibile assegnare una o più trasformazioni a una variabile intermedia, per comodità e leggibilità. Ad esempio, se si dispone di una variabile di testo denominata `email_subject` e si applica la trasformazione minuscola, è possibile denominare la variabile risultante `email_subject_lowercase`,

consentendo di tenerne traccia facilmente altrove nella composizione. Le assegnazioni possono anche essere concatenate, per consentire di applicare più trasformazioni in un ordine specificato. L'esempio seguente mostra assegnazioni singole e concatenate nella sintassi della composizione:

```
"assignments": {  
  
  "email_subject_lowercase": "lowercase(email_subject)",  
  
  "email_subject_lowercase_ngram": "ngram(lowercase(email_subject), 2)"  
  
}
```

I nomi delle variabili intermedie devono iniziare con un carattere alfabetico e possono essere costituiti da 1 a 64 caratteri. Se il nome non inizia con un carattere alfabetico o se contiene caratteri speciali (, ' "\t\r\n () \), il nome deve essere racchiuso tra virgolette per essere incluso nella composizione.

Output

La sezione risultati (output) controlla quali variabili di input saranno usate per il processo di apprendimento, e quali trasformazioni sono applicate. Una sezione di output vuota o inesistente è un errore, perché significa che nessuno dato sarà passato al processo di apprendimento.

La sezione di output più semplice include solo il gruppo predefinito ALL_INPUTS, indicando ad Amazon ML di utilizzare tutte le variabili definite nell'origine dati per l'apprendimento:

```
"outputs": [  
  
  "ALL_INPUTS"  
  
]
```

La sezione di output può anche fare riferimento agli altri gruppi predefiniti, indicando ad Amazon ML di utilizzare tutte le variabili di questi gruppi:

```
"outputs": [  
  
  "ALL_NUMERIC",
```

```
"ALL_CATEGORICAL"  
]
```

La sezione di output può anche fare riferimento a gruppi personalizzati. In questo esempio, solo uno dei gruppi personalizzati definiti nella sezione assegnazioni dei raggruppamenti nell'esempio precedente verrà utilizzato per il machine learning. Tutte le altre variabili saranno rimosse:

```
"outputs": [  
  "All_Categorical_plus_one_other"  
]
```

La sezione output può anche fare riferimento ad assegnazioni di variabili definite nella sezione assegnazioni:

```
"outputs": [  
  "email_subject_lowercase"  
]
```

Inoltre le variabili di input o le trasformazioni possono essere definite direttamente nella sezione di output:

```
"outputs": [  
  "var1",  
  "lowercase(var2)"  
]
```

L'output deve specificare in modo esplicito tutte le variabili e le variabili trasformate che si prevede saranno disponibili per il processo di apprendimento. Supponiamo, ad esempio, di includere nell'output un prodotto cartesiano di var1 e var2. Se si desidera includere anche entrambe le variabili non elaborate var1 e var2, è necessario aggiungere le variabili non elaborate alla sezione di output:

```
"outputs": [  
  "cartesian(var1,var2)",  
  "var1",  
  "var2"  
]
```

Gli output possono includere commenti per migliorare la leggibilità, con l'aggiunta del testo del commento insieme alla variabile:

```
"outputs": [  
  "quantile_bin(age, 10) //quantile bin age",  
  "age // explicitly include the original numeric variable along with the  
  binned version"  
]
```

È possibile combinare tutti questi approcci all'interno della sezione di output.

Note

I commenti non sono consentiti nella console di Amazon ML quando si aggiunge una composizione.

Esempio completo di composizione

L'esempio seguente si riferisce a diversi processori di dati integrati che sono stati introdotti negli esempi precedenti:

```
{  
  
  "groups": {
```

```
"LONGTEXT": "group_remove(ALL_TEXT, title, subject)",  
  
"SPECIALTEXT": "group(title, subject)",  
  
"BINCAT": "group(ALL_CATEGORICAL, ALL_BINARY)"  
  
},  
  
"assignments": {  
  
"binned_age" : "quantile_bin(age,30)",  
  
"country_gender_interaction" : "cartesian(country, gender)"  
  
},  
  
"outputs": [  
  
"lowercase(no_punct(LONGTEXT))",  
  
"ngram(lowercase(no_punct(SPECIALTEXT)),3)",  
  
"quantile_bin(hours-per-week, 10)",  
  
"hours-per-week // explicitly include the original numeric variable  
along with the binned version",  
  
"cartesian(binned_age, quantile_bin(hours-per-week,10)) // this one is  
critical",  
  
"country_gender_interaction",  
  
"BINCAT"  
  
]  
}
```

Composizioni suggerite

Quando si crea una nuova origine dati in Amazon ML e le statistiche sono calcolate per quell'origine dati, Amazon ML crea anche una composizione (recipe) suggerita che può essere utilizzata per

creare un nuovo modello ML dall'origine dati. L'origine dati suggerita si basa sui dati e sull'attributo di destinazione presenti nei dati, e fornisce un punto di partenza utile per la creazione e il perfezionamento dei modelli ML.

Per usare la composizione suggerita nella console di Amazon ML, scegliere Datasource (Origine dati) o Datasource and ML model (Origine dati e modello ML) dall'elenco a discesa Create new (Crea nuovo). Per le impostazioni del modello ML, si dispone di una gamma di impostazioni di default o personalizzate per l'addestramento e la valutazione nella fase Impostazioni del modello ML della procedura guidata Crea modello ML. Se si sceglie l'opzione Default, Amazon ML utilizzerà automaticamente la composizione suggerita. Se si sceglie l'opzione Custom, l'editor della composizione mostrerà nella fase successiva la composizione suggerita, e sarà possibile verificarla o modificarla in base alle esigenze.

Note

Amazon ML consente di creare un'origine dati e di usarla immediatamente per creare un modello ML, prima che il calcolo delle statistiche sia completato. In questo caso, non si potrà vedere la composizione suggerita nell'opzione Custom, ma sarà comunque possibile saltare quella fase e far sì che Amazon ML utilizzi la composizione di default per l'addestramento del modello.

Per usare la composizione suggerita con l'API Amazon ML, è possibile passare una stringa vuota nei parametri sia Recipe sia RecipeUri. Non è possibile recuperare la composizione suggerita utilizzando l'API di Amazon ML.

Riferimento per le trasformazioni di dati

Argomenti

- [Trasformazione N-gramma](#)
- [Trasformazione OSB \(Orthogonal Sparse Bigram\)](#)
- [Trasformazione in minuscolo](#)
- [Trasformazione con rimozione della punteggiatura](#)
- [Trasformazione binning quantile](#)
- [Trasformazione di normalizzazione](#)

- [Trasformazione del prodotto cartesiano](#)

Trasformazione N-gramma

La trasformazione n-gramma utilizza una variabile di testo come input e produce stringhe corrispondenti allo scorrimento di una finestra di (configurabile dall'utente) n parole, generando gli output durante il processo. A titolo illustrativo, si prenda in considerazione la stringa di testo "Ho veramente apprezzato la lettura di questo libro".

Se si specifica la trasformazione n-gramma con dimensione finestra = 1 si hanno semplicemente tutte le singole parole di quella stringa:

```
{"I", "really", "enjoyed", "reading", "this", "book"}
```

Se si specifica la trasformazione n-gramma con dimensione finestra = 2 si hanno tutte le combinazioni di due parole oltre alle combinazioni con una singola parola:

```
{"I really", "really enjoyed", "enjoyed reading", "reading this", "this book", "I", "really", "enjoyed", "reading", "this", "book"}
```

Se si specifica la trasformazione n-gramma con dimensione finestra = 3 si aggiungono le combinazioni di tre parole a questo elenco, ottenendo quanto segue:

```
{"I really enjoyed", "really enjoyed reading", "enjoyed reading this", "reading this book", "I really", "really enjoyed", "enjoyed reading", "reading this", "this book", "I", "really", "enjoyed", "reading", "this", "book"}
```

È possibile richiedere n-grammi con dimensioni comprese tra 2 e 10 parole. Gli n-grammi con dimensione 1 vengono generati implicitamente per tutti gli input il cui tipo è contrassegnato come testo nello schema dei dati, pertanto non occorre richiederli. Infine, è necessario ricordare che gli n-grammi vengono generati tramite l'interruzione dei dati di input sui caratteri di spazio. Ciò significa che, ad esempio, i caratteri di punteggiatura saranno considerati una parte dei token di parole: la generazione di n-grammi con una finestra di 2 per la stringa "rosso, verde, blu" consentirà di ottenere {"rosso", "verde", "blu", "rosso, verde", "verde, blu"}. È possibile utilizzare il processore per

la rimozione di punteggiatura (descritto più avanti in questo documento) per rimuovere, se si vuole, i simboli di punteggiatura.

Calcolo di n-grammi con dimensioni della finestra 3 per variabili var1:

```
"ngram(var1, 3)"
```

Trasformazione OSB (Orthogonal Sparse Bigram)

La trasformazione OSB è concepita per facilitare l'analisi delle stringhe di testo ed è un'alternativa alla trasformazione bi-grammi (n-grammi con dimensione finestra 2). Gli OSB vengono generati facendo scorrere la finestra di dimensioni n sul testo e utilizzando come output ogni coppia di parole che include la prima parola nella finestra.

Per creare ogni OSB, le parole che lo costituiscono sono unite dalla "_" (sottolineatura) e ogni token ignorato è indicato aggiungendo un'altra sottolineatura all'OSB. Pertanto, la codifica OSB non solo i token visto all'interno di una finestra, ma anche un'indicazione del numero di token ignorato all'interno della stessa finestra.

A titolo esemplificativo si consideri la stringa "The quick brown fox jumps over the lazy dog" e OSB di dimensione 4. Le finestre da quattro parole e le ultime due finestre più brevi dalla fine della stringa sono riportate nell'esempio seguente, oltre agli OSB generati da ciascuna:

Finestra, {OSBs generate}

```
"The quick brown fox", {The_quick, The__brown, The___fox}
"quick brown fox jumps", {quick_brown, quick__fox, quick___jumps}
"brown fox jumps over", {brown_fox, brown__jumps, brown___over}
"fox jumps over the", {fox_jumps, fox__over, fox___the}
"jumps over the lazy", {jumps_over, jumps__the, jumps___lazy}
"over the lazy dog", {over_the, over__lazy, over___dog}
"the lazy dog", {the_lazy, the__dog}
```

```
"lazy dog", {lazy_dog}
```

Gli OBS sono un'alternativa rispetto ai n-grammi che potrebbero essere più adatti in alcune situazioni. Se i dati hanno campi di testo di grandi dimensioni (10 o più parole), si può fare qualche esperimento per vedere quale funziona meglio. Si noti la definizione di un campo di testo di grandi dimensioni può variare a seconda della situazione. Tuttavia, con i campi di testo di dimensioni maggiori, è stato dimostrato empiricamente che gli OSB riproducono il testo in modo univoco a causa del simbolo skip (salta) (sottolineatura).

È possibile richiedere una dimensione finestra da 2 a 10 per trasformazioni OSB su variabili di testo di input.

Per il calcolo di OBS con dimensioni della finestra 5 per la variabile var1:

```
"osb (var1, 5)"
```

Trasformazione in minuscolo

Il processore per la trasformazione in minuscolo converte gli input di testo in minuscolo. Ad esempio, dato l'input "The Quick Brown Fox Jumps Over the Lazy Dog", il risultato del processore sarà "the quick brown fox jumps over the lazy dog".

Per applicare la trasformazione in minuscolo alla variabile var1:

```
"lowercase(var1)"
```

Trasformazione con rimozione della punteggiatura

Amazon ML divide implicitamente gli input contrassegnati come testo nello schema di dati sullo spazio. La punteggiatura nella stringa termina con token di parole adiacenti oppure come token completamente separati, a seconda dello spazio che la circonda. Se questo è indesiderabile, la trasformazione con rimozione della punteggiatura può essere usata per rimuovere i simboli di punteggiatura dalle caratteristiche generate. Ad esempio, data la stringa "Welcome to AML - please fasten your seat-belts!", viene implicitamente generato il seguente set di token:

```
{"Welcome", "to", "Amazon", "ML", "-", "please", "fasten", "your", "seat-belts!"}
```

L'applicazione del processore per la rimozione di punteggiatura a questa stringa consente di ottenere questo set:

```
{"Welcome", "to", "Amazon", "ML", "please", "fasten", "your", "seat-belts"}
```

Si noti che solo i segni di punteggiatura del prefisso e del suffisso vengono rimossi. I segni di punteggiatura che appaiono a metà di un token, ad esempio il trattino in "seat-belts" (cinture di sicurezza), non vengono rimossi.

Per applicare la rimozione della punteggiatura alla variabile var1:

```
"no_punct (var1)"
```

Trasformazione binning quantile

Il processore binning quantile richiede due input, una variabile numerica e un parametro denominato bin number (numero bin) e fornisce come risultato una variabile categorica. Lo scopo è scoprire la non linearità nella distribuzione della variabile raggruppando insieme i valori osservati.

In molti casi, il rapporto tra una variabile numerica e la destinazione è non lineare (il valore della variabile numerica non aumenta né diminuisce monotonicamente con la destinazione). In questi casi, può essere utile effettuare il binning della caratteristica numerica in una caratteristica categorica che rappresenta diversi intervalli della caratteristica numerica. Ogni valore della caratteristica categorica (bin) può quindi essere modellato come se avesse la propria relazione lineare con la destinazione. Ad esempio, supponiamo di informarti che la caratteristica numerica continua `account_age` non sia correlata linearmente alla probabilità di acquistare un libro. È possibile effettuare il binning dell'età in caratteristiche categoriche che potrebbero essere in grado di acquisire in modo più preciso il rapporto con la destinazione.

Il processore binning quantile può essere utilizzato per istruire Amazon ML a definire n contenitori di pari dimensioni in base alla distribuzione di tutti i valori di input della variabile età e quindi a sostituire ogni numero con un token di testo contenente il contenitore. Il numero ottimale di bin per una variabile numerica dipende dalle caratteristiche della variabile e dal suo rapporto con la destinazione, e ciò si determina meglio attraverso la sperimentazione. Amazon ML suggerisce il numero ottimale di contenitori per una caratteristica numerica basata su dati statistici nella [composizione suggerita](#).

È possibile richiedere tra 5 e 1.000 contenitori quantili da calcolare per qualsiasi variabile di input numerici.

L'esempio seguente spiega come calcolare e utilizzare 50 contenitori al posto della variabile numerica var1:

```
"quantile_bin(var1, 50)"
```

Trasformazione di normalizzazione

Il trasformatore di normalizzazione normalizza le variabili numeriche per avere una media di zero e una varianza di uno. La normalizzazione delle variabili numeriche può aiutare il processo di apprendimento se vi sono differenze di gamma molto grandi tra le variabili numeriche, poiché variabili con la massima grandezza potrebbe dominare il modello ML, indipendentemente dal fatto che la funzione sia informativa o meno riguardo alla destinazione.

Per applicare questa trasformazione alla variabile numerica `var1`, aggiungere questo alla composizione:

```
normalize(var1)
```

Questo trasformatore può anche utilizzare come input un gruppo di variabili numeriche definito dall'utente o il gruppo predefinito per tutte le variabili numeriche (`ALL_NUMERIC`):

```
normalize (ALL_NUMERIC)
```

Nota

Non è obbligatorio utilizzare il processore di normalizzazione per le variabili numeriche.

Trasformazione del prodotto cartesiano

La trasformazione cartesiana genera permutazioni di due o più variabili di testo o di input categorico. Questa trasformazione viene utilizzata quando si sospetta un'interazione tra variabili. A titolo illustrativo, si prenda in considerazione il set di dati di marketing utilizzato nel tutorial relativo all'analisi: Utilizzare Amazon ML per prevedere le risposte a un'offerta di marketing. Con l'utilizzo di questo set di dati, vorremmo prevedere se una persona risponderebbe positivamente a una promozione, in base alle informazioni economiche e demografiche. Ci si potrebbe aspettare che il tipo di lavoro di una persona sia molto importante (forse c'è una correlazione tra lavorare in determinati campi e avere denaro a disposizione) e che il massimo livello di istruzione raggiunto sia anch'esso importante. È inoltre possibile avere un'intuizione più profonda circa il fatto che vi sia una forte interazione tra queste due variabili, per esempio che la promozione è particolarmente indicata per i clienti che sono imprenditori e che hanno conseguita una laurea.

La trasformazione del prodotto cartesiano richiede variabili categoriche o di testo come input e produce nuove caratteristiche in grado di acquisire l'interazione tra queste variabili di input. Nello specifico, per ogni esempio di addestramento, creerà una combinazione di caratteristiche

e le aggiungerà come caratteristica indipendente. Supponiamo ad esempio che le righe di input semplificate abbiano questo aspetto:

target, education, job

0, university.degree, technician

0, high.school, services

1, university.degree, admin

Se specifichiamo che la trasformazione cartesiana deve essere applicata ai cambi delle variabili categoriche istruzione (education) e lavoro (job), la caratteristica risultante `education_job_interaction` avrà questo aspetto:

target, education_job_interaction

0, university.degree_technician

0, high.school_services

1, university.degree_admin

La trasformazione cartesiana è ancora più potente quando si tratta di lavorare su sequenze di token, come avviene quando uno dei suoi argomenti è una variabile di testo implicitamente o esplicitamente divisa in token. A titolo illustrativo, si prenda in considerazione l'attività di classificazione di un libro come manuale o meno. Intuitivamente si potrebbe pensare che qualcosa nel titolo possa dirci se si tratta di un manuale (determinate parole apparire più frequentemente nei titoli dei libri di testo) e che nella rilegatura del libro vi sia qualcosa di predittivo (è più probabile che i manuali siano rilegati con una copertura rigida), ma è la combinazione di alcune parole nel titolo e della rilegatura a essere più predittiva. Per un esempio reale, la tabella riportata di seguito mostra i risultati dell'applicazione del processore cartesiano alle variabili di input rilegatura e titolo:

Mani	Title	Rilegatura	Prodotto cartesiano di no_punct (titolo) e rilegatura
1	Economia: Principi, problemi, politiche	Hardcover (copertina rigida)	{"Economics_Hardcover", "Principles_Hardcover", "Problems_Hardcover", "Policies_Hardcover"}

Man	Title	Rilegatura	Prodotto cartesiano di no_punct (titolo) e rilegatura
0	The Invisible Heart: Una storia d'amore nell'economia	Softcover (copertina morbida)	{"The_Softcover", "Invisible_Softcover", "Heart_Softcover", "An_Softcover", "Economics_Softcover", "Romance_Softcover"}
0	Fun With Problems (Divertirsi con i problemi)	Softcover (copertina morbida)	{"Fun_Softcover", "With_Softcover", "Problems_Softcover"}

L'esempio seguente mostra come applicare il trasformatore cartesiano a `var1` e `var2`:

```
cartesian(var1, var2)
```

Riordino dei dati

La funzionalità di riordino dei dati consente di creare un'origine dati che si basa solo su una parte dei dati di input a cui punta. Ad esempio, quando crei un modello ML utilizzando il Crea un modello ML nella console Amazon ML e sceglie l'opzione di valutazione predefinita, Amazon ML riserva automaticamente il 30% dei dati per la valutazione del modello ML e utilizza il restante 70% per l'addestramento. Questa funzionalità viene attivata tramite la funzione Data Rearrangement (Riordino dei dati) di Amazon ML.

Se si sta utilizzando l'API Amazon ML per creare origini dati, è possibile specificare su quale parte dei dati di input sarà basata la nuova origine dati. A questo scopo, si trasferiscono le istruzioni nel parametro `DataRearrangement` alle API `CreateDataSourceFromS3`, `CreateDataSourceFromRedshift` o `CreateDataSourceFromRDS`. Il contenuto della stringa `DataRearrangement` è una stringa JSON contenente le ubicazioni di inizio e fine dei dati, espresse come percentuali, un flag complementare e una strategia di divisione. Ad esempio, la seguente stringa `DataRearrangement` specifica che il primo 70% dei dati verrà utilizzato per creare l'origine dati:

```
{
  "splitting": {
```

```
"percentBegin": 0,  
"percentEnd": 70,  
"complement": false,  
"strategy": "sequential"  
}  
}
```

Parametri DataRearrangement

Per modificare il modo in cui Amazon ML crea un'origine dati, utilizzare i seguenti parametri.

PercentBegin (facoltativo)

Utilizzare `percentBegin` per indicare dove iniziano i dati per l'origine dati. Se non includi `percentBegin` e `percentEnd`, Amazon ML include tutti i dati quando crea l'origine dati.

I valori validi vanno da 0 a 100, inclusi.

PercentEnd (facoltativo)

Utilizzare `percentEnd` per indicare dove finiscono i dati per l'origine dati. Se non includi `percentBegin` e `percentEnd`, Amazon ML include tutti i dati quando crea l'origine dati.

I valori validi vanno da 0 a 100, inclusi.

Complement (facoltativo)

La `complement` dice ad Amazon ML di utilizzare i dati che non sono inclusi nell'intervallo di `percentBegin` e `percentEnd` per creare un'origine dati. Il parametro `complement` è utile se occorre creare origini dati complementari per l'addestramento e la valutazione. Per creare un'origine dati complementari, utilizzare gli stessi valori per `percentBegin` e `percentEnd`, insieme al parametro `complement`.

Ad esempio, le due origini dati seguenti non condividono dati e possono essere utilizzate per addestrare e valutare un modello. La prima origine dati ha il 25% dei dati, mentre la seconda ha il 75% dei dati.

Origine dati per la valutazione:

```
{  
  "splitting":{  
    "percentBegin":0,
```



```

    "percentEnd":25
  }
}

```

Origine dati per l'addestramento:

```

{
  "splitting":{
    "percentBegin":0,
    "percentEnd":25,
    "complement":"true"
  }
}

```

I valori validi sono `true` e `false`.

Strategy (facoltativo)

Per modificare il modo in cui Amazon ML divide i dati per un'origine dati, utilizzare il `strategy` Parametro .

Il valore predefinito per il `strategy` il parametro è `sequential`, il che significa che Amazon ML prende tutti i record di dati tra `percentBegin` e `percentEnd` parametri per l'origine dati, nell'ordine in cui i record appaiono nei dati di input

Le due righe seguenti `DataRearrangement` sono esempi di ordinamento sequenziale di origini dati di addestramento e valutazione:

Origine dati per la valutazione: `{"splitting":{"percentBegin":70, "percentEnd":100, "strategy":"sequential"}}`

Origine dati per l'addestramento: `{"splitting":{"percentBegin":70, "percentEnd":100, "strategy":"sequential", "complement":"true"}}`

Per creare un'origine dati da una selezione casuale di dati, impostare il parametro `strategy` su `random` e fornire una stringa che viene utilizzata come valore di origine per la suddivisione casuale dei dati (ad esempio, è possibile utilizzare il percorso di S3 per i dati come stringa di origine casuale). Se si sceglie la strategia di divisione casuale, Amazon ML assegna a ogni riga di dati un numero pseudocasuale, quindi seleziona le righe che hanno un numero assegnato compreso tra `percentBegin` e `percentEnd`. I numeri pseudocasuali sono assegnati utilizzando l'offset di byte come `seed`; perciò, se si modificano i risultati dei dati, si ottiene una divisione

diversa. Qualsiasi ordine esistente viene mantenuto. La strategia di divisione casuale garantisce che le variabili dei dati di addestramento e valutazione siano distribuite in modo analogo. Si tratta di una funzione utile, ad esempio, nel caso in cui i dati di input possano avere un ordinamento implicito; altrimenti, ciò porterebbe a origini dati di addestramento e valutazione contenenti record di dati non simili.

Le due righe seguenti `DataRearrangement` sono esempi di ordinamento non sequenziale di origini dati di addestramento e valutazione:

Origine dati per la valutazione:

```
{
  "splitting":{
    "percentBegin":70,
    "percentEnd":100,
    "strategy":"random",
    "strategyParams": {
      "randomSeed":"RANDOMSEED"
    }
  }
}
```

Origine dati per l'addestramento:

```
{
  "splitting":{
    "percentBegin":70,
    "percentEnd":100,
    "strategy":"random",
    "strategyParams": {
      "randomSeed":"RANDOMSEED"
    }
    "complement":"true"
  }
}
```

I valori validi sono `sequential` e `random`.

Strategy:RandomSeed (facoltativo)

Amazon ML utilizza ilSeed casuale per dividere i dati. Il seed di default per l'API è una stringa vuota. Per specificare un seed per la strategia di divisione casuale, effettuare una passata su una

stringa. Per ulteriori informazioni sulle origini casuali, consulta [Divisione casuale dei dati](#) nella Guida per sviluppatori Amazon Machine Learning.

Per il codice di esempio che dimostra come utilizzare la convalida incrociata con Amazon ML, visitare [la Esempi di Machine Learning Github](#).

Valutazione dei modelli ML

È sempre consigliabile valutare un modello per determinare se riuscirà a predire correttamente la destinazione per i dati nuovi e futuri. Poiché le istanze future hanno valori di destinazione ignoti, è necessario verificare il parametro di accuratezza del modello ML su dati dei quali si conosce già la risposta target e utilizzare questa valutazione come proxy per la precisione predittiva sui dati futuri.

Per la corretta valutazione di un modello, si tiene un campione di dati dell'origine dati per l'addestramento che è stato etichettato con la destinazione (dati acquisiti sul campo). Non è utile valutare la precisione predittiva di un modello ML con gli stessi dati utilizzati per l'addestramento, perché in questo modo si premiano modelli in grado di "ricordare" i dati di addestramento, anziché utilizzarli per la generalizzazione. Una volta finito di addestrare il modello ML, è possibile inviare al modello le osservazioni utilizzate di cui si conoscono i valori di destinazione. Si confrontano, quindi, le previsioni restituite dal modello ML con il valore noto di destinazione. Infine, si calcola un parametro riepilogativo che indica quanto è elevato il grado di corrispondenza tra i valori previsti e quelli reali.

In Amazon ML, si valuta un modello ML mediante creazione di una valutazione. Per creare una valutazione per un modello di ML, è necessario disporre di un modello ML da valutare e occorrono dati etichettati che non siano stati utilizzati per l'addestramento. In primo luogo, si crea un'origine dati per la valutazione creando un'origine dati Amazon ML con i dati utilizzati. I dati utilizzati nella valutazione devono avere lo stesso schema di quelli utilizzati per l'addestramento e includere i valori effettivi della variabile di destinazione.

Se tutti i dati sono in un unico file o directory, è possibile utilizzare la console di Amazon ML per dividere i dati. Il percorso di default nella procedura guidata Crea un modello ML divide l'origine dati di input e utilizza il primo 70% per un'origine dati di addestramento e il restante 30% per un'origine dati di valutazione. È anche possibile personalizzare il rapporto di divisione utilizzando l'opzione Custom (Personalizza) nella procedura guidata Crea un modello ML, dove è possibile selezionare un campione casuale del 70% per l'addestramento e utilizzare il restante 30% per la valutazione. Per specificare ulteriormente le proporzioni di divisione personalizzate, si può utilizzare la stringa di riordino dei dati nell'API [Crea origine dati](#). Una volta che si dispone di un'origine dati di valutazione e di un modello ML, è possibile creare una valutazione ed esaminare i risultati della valutazione.

Argomenti

- [Informazioni del modello ML](#)
- [Informazioni sul modello binario](#)
- [Informazioni del modello multiclasse](#)

- [Informazioni sul modello di regressione](#)
- [Prevenzione dell'overfitting](#)
- [Convalida incrociata](#)
- [Avvisi relativi alla valutazione](#)

Informazioni del modello ML

Quando si valuta un modello ML, Amazon ML fornisce un parametro standard di settore e una serie di informazioni per verificare l'accuratezza predittiva del modello. In Amazon ML, l'esito di una valutazione contiene quanto segue:

- Un parametro di accuratezza della previsione per segnalare il successo globale del modello
- Visualizzazioni per esplorare l'accuratezza del modello oltre il parametro dell'accuratezza predittiva
- La possibilità di esaminare l'impatto dell'impostazione di un punteggio soglia (solo per la classificazione binaria)
- Avvisi sui criteri per verificare la validità della valutazione

La scelta del parametro e della visualizzazione dipende dal tipo di modello ML che si sta valutando. È importante rivedere queste visualizzazioni per decidere se il modello sta avendo prestazioni in grado di soddisfare i requisiti aziendali.

Informazioni sul modello binario

Interpretazione delle previsioni

L'output effettivo di molti algoritmi di classificazione binaria è un punteggio di previsione. Il punteggio indica la certezza del sistema che una data osservazione appartenga alla classe positiva (l'effettivo valore di destinazione è 1). I modelli di classificazione binaria in Amazon ML producono un punteggio che varia da 0 a 1. In quanto consumatore di questo punteggio, per decidere se l'osservazione debba essere classificata come 1 o 0, si interpreterà il punteggio scegliendo una soglia di classificazione o limite (cut-off) e si confronterà il punteggio con tale soglia. Le eventuali osservazioni con punteggio superiore al limite vengono considerate come target = 1 mentre quelle con punteggio inferiore come target = 0.

In Amazon ML, il punteggio limite predefinito è 0,5. È possibile scegliere di aggiornare questo limite per soddisfare le tue esigenze aziendali. È possibile utilizzare le visualizzazioni nella console per comprendere il modo in cui la scelta del limite influenzerà l'applicazione.

Misurazione dell'accuratezza del modello ML

Amazon ML fornisce un parametro di accuratezza standard del settore per i modelli di classificazione binaria denominati AUC (Area Under the Curve) ROC (Receiver Operating Characteristic). L'AUC misura la capacità del modello ML di prevedere un punteggio più elevato per gli esempi positivi rispetto agli esempi negativi. Poiché è indipendente dal punteggio limite, è possibile farsi un'idea dell'accuratezza del modello attraverso il parametro AUC, senza selezionare una soglia.

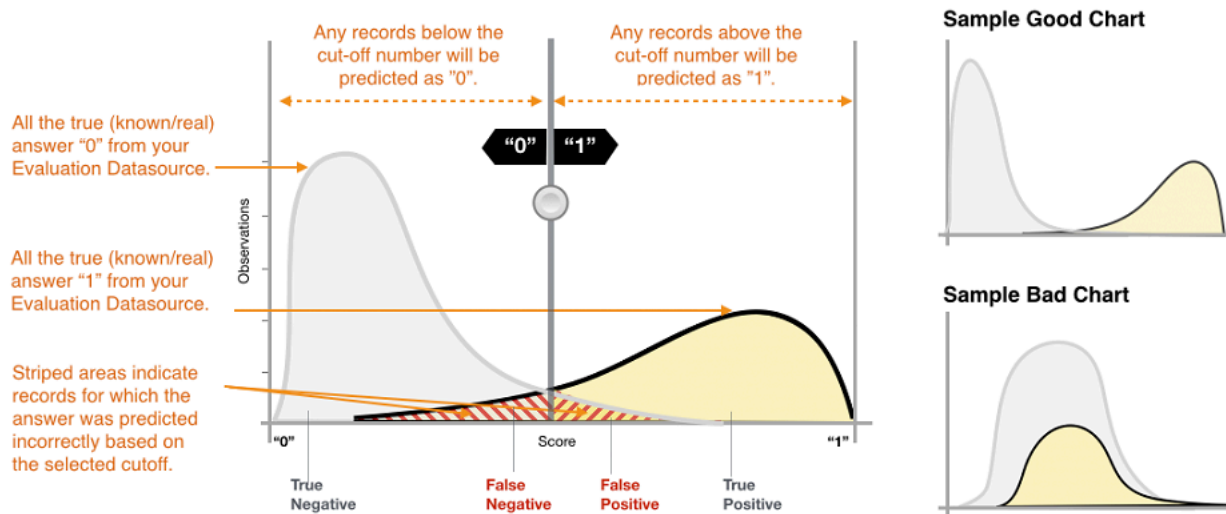
Il parametro AUC restituisce un valore decimale compreso tra 0 e 1. I valori di AUC prossimi a 1 indicano un modello di ML estremamente accurato. I valori vicino a 0,5 indicano un modello ML equivalente all'indovinare a caso. È insolito vedere valori vicino a 0 e in genere indicano un problema con i dati. Essenzialmente, un'AUC vicino a 0 dice che il modello ML ha appreso i pattern corretti, ma li sta utilizzando per fare previsioni che sono capovolte rispetto dalla realtà (gli 0 sono previsti come 1 e viceversa). Per ulteriori informazioni sull'AUC, consultare la pagina [Receiver operating characteristic](#) di Wikipedia.

La baseline del parametro AUC per un modello binario è 0,5. È il valore di un ipotetico modello ML che prevede in modo casuale una risposta 1 o 0. Il modello ML binario deve ottenere prestazioni superiori a questo valore per essere utile.

Utilizzo della Performance Visualization

Per valutare l'accuratezza del modello ML, è possibile esaminare i grafici della [Valutazione](#) pagina sulla console Amazon ML. Questa pagina mostra due istogrammi: a) un istogramma dei punteggi per i positivi effettivi (il target è 1) e b) un istogramma dei punteggi per i negativi effettivi (il target è 0) nei dati di valutazione.

Un modello ML che dispone di buona accuratezza predittiva sarà in grado di prevedere punteggi più elevati per gli 1 effettivi e punteggi inferiori per gli 0 effettivi. Un modello perfetto avrà due istogrammi a due estremità differenti dell'asse x che mostrano che tutti i punteggi positivi effettivi hanno ottenuto punteggi elevati e che tutti i negativi effettivi hanno ottenuto punteggi bassi. Tuttavia, i modelli ML commettono errori e un grafico tipico mostrerà che i due istogrammi si sovrappongono in corrispondenza di determinati punteggi. Un modello con prestazioni estremamente scarse non sarà in grado di distinguere tra le classi positive e negative ed entrambe le classi avranno per lo più istogrammi che si sovrappongono.



Utilizzando le visualizzazioni, è possibile individuare il numero di previsioni che ricadono nei due tipi di previsioni corrette e nei due tipi di previsioni errate.

Previsioni corrette

- Vero positivo (True positive, TP Amazon ML ha previsto il valore come 1 e il valore true è 1.
- Vero negativo (True negative, TN Amazon ML ha previsto il valore come 0 e il valore true è 0.

Previsioni errate

- Falso positivo (False positive, FP): Amazon ML ha previsto il valore come 1, ma il valore true è 0.
- Falso negativo (False negative, FN): Amazon ML ha previsto il valore come 0, ma il valore true è 1.

i Note

Il numero di TP, TN, FP e FN dipende dalla soglia selezionata per il punteggio e l'ottimizzazione per uno qualsiasi di questi numeri significherebbe fare un compromesso sugli altri. Un numero elevato di TP spesso comporta un numero elevato di FP e un basso numero di TN.

Regolazione del punteggio soglia

I modelli ML funzionano generando punteggi di previsione numerici, pertanto l'applicazione di un cut-off converte questi punteggi in etichette binarie 0/1. Modificando il punteggio soglia, è possibile

regolare il comportamento del modello quando fa un errore. SulValutazione nella console di Amazon ML, è possibile esaminare l'impatto di vari punteggi limite e salvare quello da utilizzare per il modello.

Quando si regola il punteggio soglia, si osserva il trade-off tra i due tipi di errori. Spostando la soglia a sinistra si acquisiscono più veri positivi, ma il trade-off consiste in un aumento del numero di errori relativi ai falsi positivi. Spostandolo a destra acquisisce meno errori relativi ai falsi positivi, ma il trade-off è che salta alcuni veri positivi. Per l'applicazione predittiva, si decide che tipo di errore è più tollerabile selezionando un punteggio soglia adeguato.

Revisione dei parametri avanzati

Amazon ML fornisce i seguenti parametri aggiuntivi per misurare l'accuratezza predittiva del modello ML: accuratezza, precisione, recall e percentuale di falsi positivi.

Accuratezza

Accuracy (Accuratezza) (ACC) misura la percentuale di previsioni corrette. L'intervallo è compreso tra 0 e 1. Un valore maggiore indica una migliore accuratezza predittiva:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precisione

Precision (Precisione) misura la percentuale di positivi effettivi tra gli esempi previsti come positivi. L'intervallo è compreso tra 0 e 1. Un valore maggiore indica una migliore accuratezza predittiva:

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall (Richiamata) misura la percentuale di positivi effettivi previsti come positivi. L'intervallo è compreso tra 0 e 1. Un valore maggiore indica una migliore accuratezza predittiva:

$$Recall = \frac{TP}{TP + FN}$$

Percentuale di falsi positivi

False positive rate (Percentuale falsi positivi) (FPR) misura il la percentuale di falsi allarmi o la percentuale di negativi effettivi previsti come positivi. L'intervallo è compreso tra 0 e 1. Un valore minore indica una migliore accuratezza predittiva:

$$FPR = \frac{FP}{FP + TN}$$

A seconda del problema aziendale, si potrebbe essere più interessati a un modello che esegua correttamente uno specifico sottoinsieme di questi parametri. Ad esempio, due applicazioni aziendali potrebbe avere requisiti molto diversi per il loro modello ML:

- Un'applicazione potrebbe essere molto sicura che le previsioni positive siano effettivamente positive (elevata precisione) e potersi permettere di classificare erroneamente alcuni esempi positivi come negativi (recall moderato).
- Un'altra applicazione potrebbe dover prevedere correttamente tutti gli esempi positivi possibile (recall elevato) e accetterà l'errata classificazione solo di alcuni esempi negativi come positivi (precisione moderata).

Amazon ML consente di scegliere un punteggio soglia che corrisponda ad un determinato valore di uno qualsiasi dei precedenti parametri avanzati. Mostra inoltre i trade-off legati all'ottimizzazione di qualsiasi parametro. Ad esempio, se si seleziona una soglia che corrisponde ad un'elevata precisione, è in genere necessario effettuare un trade-off con un minore recall.

Note

È necessario salvare il punteggio soglia al fine di renderlo effettivo per la classificazione delle future previsioni da parte del modello ML.

Informazioni del modello multiclasse

Interpretazione delle previsioni

L'output effettivo di un algoritmo di classificazione multiclasse è un insieme di punteggi di previsione. I punteggi indicano la certezza del modello che una data osservazione appartenga a ciascuna delle classi. A differenza dei problemi di classificazione binaria, non è necessario scegliere un punteggio limite per fare le previsioni. La risposta prevista è la classe (ad esempio, l'etichetta) con il miglior punteggio previsto.

Misurazione dell'accuratezza del modello ML

I parametri tipici utilizzati nella multiclasse sono gli stessi utilizzati nel caso della classificazione binaria, dopo averne calcolato la media su tutte le classi. In Amazon ML, il punteggio macro-medio F1 viene utilizzato per valutare l'accuratezza predittiva di un parametro multiclasse.

Punteggio macro medio F1

Il punteggio F1 è un parametro di classificazione binaria che considera entrambi i parametri binari precisione e recall. È la media armonica tra precisione e recall. L'intervallo è compreso tra 0 e 1. Un valore maggiore indica una migliore accuratezza predittiva:

$$F1\ score = \frac{2 * precision * recall}{precision + recall}$$

Il punteggio macro medio F1 è la media non ponderata del punteggio F1 su tutte le classi del caso multiclasse. Non tiene conto della frequenza di occorrenza delle classi nel set di dati di valutazione. Un valore maggiore indica una migliore accuratezza predittiva. L'esempio seguente mostra K classi nell'origine dati di valutazione:

$$Macro\ average\ F1\ score = \frac{1}{K} \sum_{k=1}^K F1\ score\ for\ class\ k$$

Punteggio macro medio F1 di riferimento

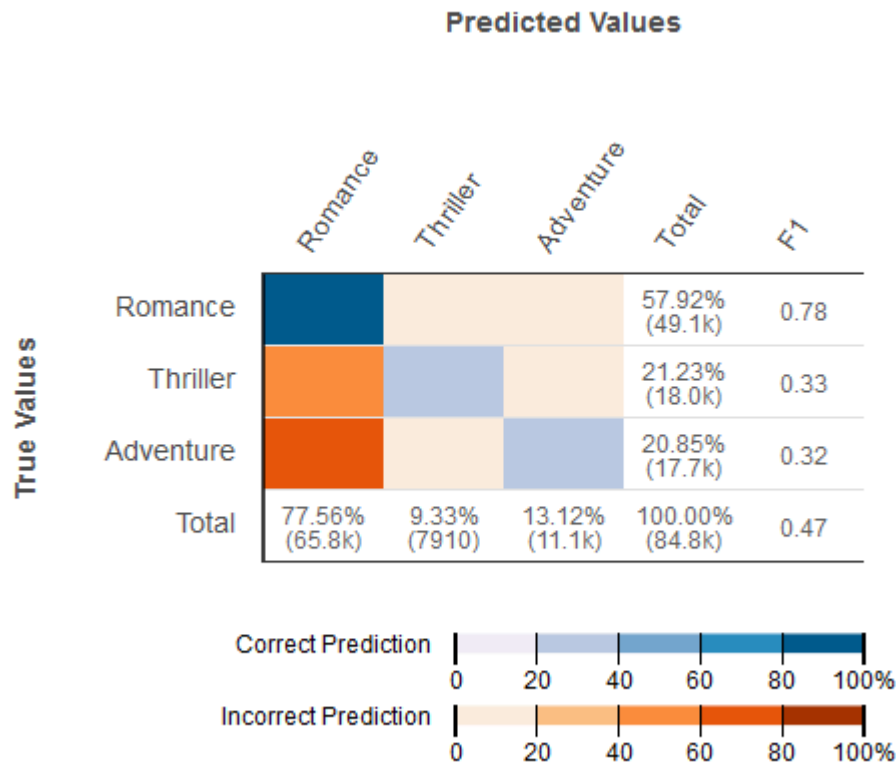
Amazon ML fornisce un parametro di riferimento per i modelli multiclasse. È il punteggio macro medio F1 di un ipotetico modello multiclasse che prevede sempre come risposta la classe più frequente. Ad esempio, se si dovesse prevedere il genere di un film e il genere più comune nei dati di addestramento è stato Romantico, il modello di riferimento prevederebbe sempre il genere come Romantico. È possibile confrontare il modello ML con questo riferimento, per convalidare se il modello ML sia meglio di un modello ML che prevede questa risposta costante.

Utilizzo della Performance Visualization

Amazon ML fornisce una Matrice di confusione come un modo per visualizzare l'accuratezza dei modelli predittivi di classificazione multiclasse. La matrice di confusione illustra in una tabella il numero o la percentuale di previsioni corrette ed errate per ogni classe, confrontando la classe prevista di un'osservazione e la sua classe "true".

Ad esempio, se si sta cercando di classificare un film in un genere, il modello di previsione potrebbero prevedere che il suo genere (classe) sia Romantico. Tuttavia, il suo genere reale potrebbe essere, di

fatto, Thriller. Quando si valuta la precisione di un modello di classificazione multiclasse ML, Amazon ML identifica queste errate classificazioni e visualizza i risultati nella matrice di confusione, come illustrato di seguito.



Le seguenti informazioni vengono visualizzate in una matrice di confusione:

- Numero di previsioni corrette e errate per ogni classe: Ogni riga della matrice di confusione corrisponde ai parametri di una delle classi «true». Ad esempio, la prima riga mostra che per i film che sono effettivamente nel genere Romance (romantico), il modello ML multiclasse fornisce le previsioni corrette per oltre l'80% dei casi. Prevede in modo errato il genere come Thriller per meno del 20% dei casi, e Adventure (avventura) per meno del 20% dei casi.
- Punteggio F1 per classe: L'ultima colonna mostra il punteggio F1 per ciascuna delle classi.
- Frequenze di classe reali nei dati di valutazione: La penultima colonna mostra che nei set di dati di valutazione, il 57,92% delle osservazioni nei dati di valutazione è Romance, il 21,23% è Thriller e il 20,85% è Adventure.
- Frequenze di classe previste per i dati di valutazione: L'ultima riga mostra la frequenza di ogni classe nelle previsioni. Il 77,56% delle osservazioni è previsto come Romance, il 9,33% come Thriller e il 13,12% come Adventure.

La console di Amazon ML fornisce una rappresentazione visiva che gestisce fino a 10 classi nella matrice di confusione, elencate nell'ordine che dalla classe più frequente a quella meno frequente nei dati di valutazione. Se i dati di valutazione hanno più di 10 classi, si possono vedere le prime 9 classi più frequenti presenti nella matrice di confusione, e tutte le altre classi sono comprese in una classe denominata "altre". Amazon ML fornisce inoltre la possibilità di scaricare l'intera matrice di confusione tramite un link sulla pagina delle visualizzazioni multiclasse.

Informazioni sul modello di regressione

Interpretazione delle previsioni

L'output di un modello ML di regressione è un valore numerico per la previsione del target da parte del modello. Ad esempio, se si stanno prevedendo i prezzi delle case, la previsione del modello potrebbe essere un valore come 254.013.

Note

La gamma di previsioni può differire dalla gamma del target nei dati di addestramento. Ad esempio, supponiamo che si stiano prevedendo i prezzi delle case e che il target nei dati di addestramento abbia valori in un intervallo da 0 a 450.000. Il target previsto non deve essere necessariamente nello stesso intervallo e potrebbe assumere qualsiasi valore positivo (superiore a 450.000) o negativo (meno di zero). È importante pianificare cosa fare in presenza di valori di previsione esterni a un determinato intervallo accettabile per l'applicazione.

Misurazione dell'accuratezza del modello ML

Per le attività di regressione, Amazon ML utilizza il parametro standard di settore dell'errore quadratico medio (RMSE). Misura la distanza tra il target numerico previsto e la risposta numerica effettiva (dati acquisiti sul campo). Minore è il valore del RMSE, migliore è l'accuratezza predittiva del modello. Un modello con previsioni perfettamente corrette avrebbe un RMSE di 0. L'esempio seguente mostra dati di valutazione che contengono N record:

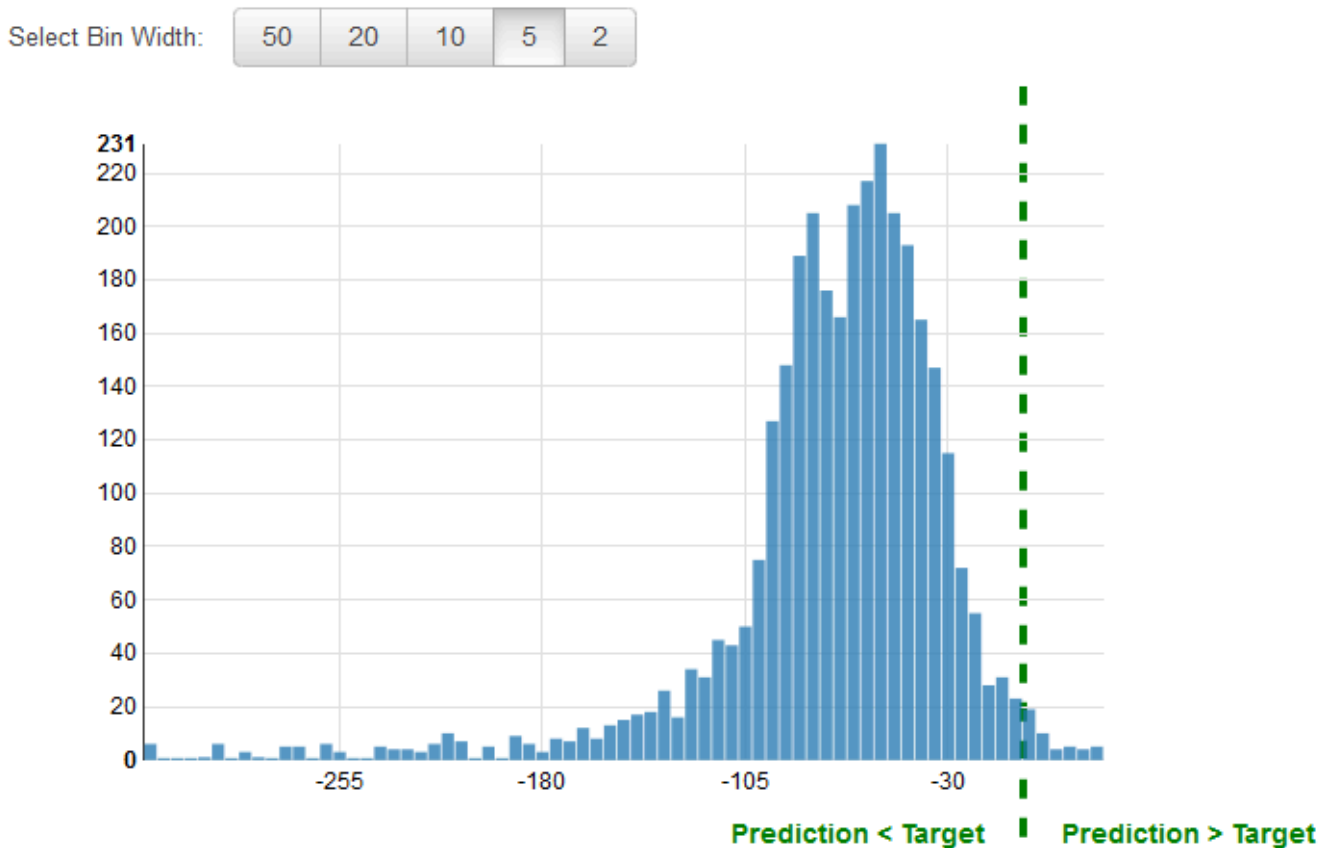
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{actual target} - \text{predicted target})^2}$$

Riferimento RMSE

Amazon ML fornisce un parametro di riferimento per i modelli di regressione. È l'RMSE per un modello di regressione ipotetico che sarebbe sempre in grado di prevedere la media del target come risposta. Ad esempio, se si dovesse prevedere l'età dell'acquirente di una casa e l'età media per le osservazioni nei dati di addestramento è 35, il modello di base sarebbe sempre in grado di prevedere la risposta come 35. È possibile confrontare il modello ML con questo riferimento, per convalidare se il modello ML sia meglio di un modello ML che prevede questa risposta costante.

Utilizzo della Performance Visualization

È prassi comune esaminare i residui per cercare problemi di regressione. Un residuo di un'osservazione nei dati di valutazione è la differenza tra il target reale e il target previsto. I residui rappresentano quella parte del target che il modello non è in grado di prevedere. Un residuo positivo indica che il modello sta sottovalutando il target (il target effettivo è più grande del target previsto). Un residuo negativo indica una sopravvalutazione (il target effettivo è più piccolo del target previsto). L'istogramma dei residui relativi ai dati di valutazione, ove distribuito con una forma a campana e centrato sullo zero, indica che il modello compie errori in modo aleatorio e non sovra-prevede o sotto-prevede sistematicamente una determinata gamma di valori target. Se i residui non assumono una forma a campana centrata sullo zero, è presente una struttura nell'errore di previsione del modello. L'aggiunta di ulteriori variabili al modello potrebbero aiutarlo ad acquisire il pattern che non viene acquisito dal modello attuale. La figura seguente mostra i residui che non sono centrati intorno allo zero.



Prevenzione dell'overfitting

Durante la creazione e l'addestramento di un modello ML, l'obiettivo è selezionare il modello che consente di ottenere le migliori previsioni, il che significa selezionare il modello con le impostazioni migliori (impostazioni del modello ML o iperparametri). In Amazon Machine Learning, esistono quattro iperparametri che è possibile impostare: numero di passate, regolarizzazione, dimensioni del modello e tipo di riproduzione casuale. Tuttavia, se si seleziona le impostazioni dei parametri del modello che producono le "migliori" prestazioni predittive riguardo alla valutazione dei dati, si potrebbe provocare un overfitting del modello. L'overfitting si verifica quando un modello ha memorizzato pattern che si verificano nelle origini dati di addestramento e valutazione, ma non è riuscito a generalizzare i pattern nei dati. Si verifica spesso quando i dati di addestramento includono tutti i dati utilizzati nella valutazione. Un modello con overfitting si comporta correttamente durante le valutazioni, ma non riesce a eseguire previsioni precise sui dati invisibili.

Per evitare di scegliere un modello con overfitting come il miglior modello, è possibile prenotare dati aggiuntivi per convalidare le prestazioni del modello ML. Ad esempio, è possibile dividere i dati in 60% per l'addestramento, 20% per la valutazione e un ulteriore 20% per la convalida. Dopo aver

selezionato i parametri del modello più adatti ai dati di valutazione, è possibile eseguire una seconda valutazione con i dati di convalida per controllare quanto è elevata la qualità delle prestazioni del modello ML con i dati di convalida. Se il modello soddisfa le aspettative riguardo ai dati di convalida, non sta effettuando l'overfitting dei dati.

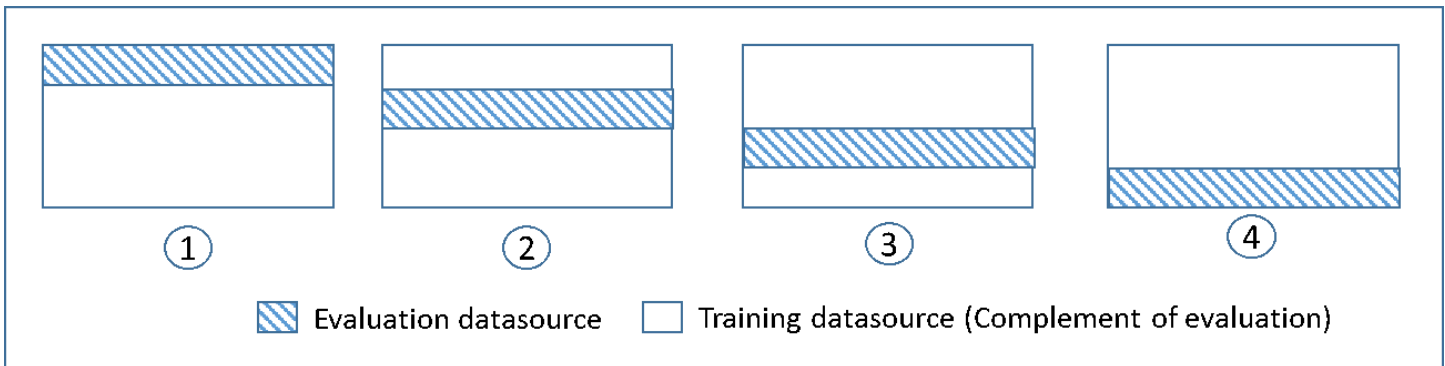
L'utilizzo di una terza serie di dati per la convalida consente di selezionare i parametri del modello ML più adatti a evitare l'overfitting. Tuttavia, se si trattengono dati del processo di addestramento per la valutazione e la convalida, vi sono meno dati disponibili per l'addestramento. Questo è un problema soprattutto con serie di dati di dimensioni ridotte, perché è sempre consigliabile utilizzare quanti più dati possibile per l'addestramento. Per risolvere questo problema, è possibile eseguire la convalida incrociata. Per ulteriori informazioni sulla convalida incrociata, consultare [Convalida incrociata](#).

Convalida incrociata

La convalida incrociata è una tecnica per valutare i modelli ML addestrando vari modelli ML su sottoinsiemi dei dati in ingresso disponibili e valutandoli sulla base di un sottoinsieme complementare dei dati. Per rilevare l'overfitting, ossia la mancata generalizzazione di un modello, usare la convalida incrociata.

In Amazon ML, è possibile utilizzare il metodo k-fold cross-validation per eseguire la convalida incrociata. Con il metodo k-fold cross-validation si dividono i dati di input in k sottoinsiemi di dati (noti anche come fold). Si addestra un modello ML su tutti i sottoinsiemi tranne uno (k-1), quindi si valuta il modello sul sottoinsieme che non è stato utilizzato per l'addestramento. Questo processo viene ripetuto k volte, ogni volta con un diverso sottoinsieme riservato per la valutazione (ed escluso dall'addestramento).

Il seguente diagramma mostra un esempio dei sottoinsiemi di addestramento e dei sottoinsiemi di valutazione complementari generati per ciascuno dei quattro modelli che vengono creati e addestrati durante una convalida incrociata per 4 volte. Il modello uno utilizza il primo 25% dei dati per la valutazione e il restante 75% per l'addestramento. Il modello due usa il secondo sottoinsieme, pari al 25% (dal 25% al 50%), per la valutazione, mentre i restanti tre sottoinsiemi dei dati sono utilizzati per l'addestramento e così via.



Ogni modello è addestrato e valutato utilizzando origini dati complementari: i dati dell'origine dati di valutazione includono e sono limitati a tutti i dati che non fanno parte dell'origine dati di addestramento. È possibile creare origini dati per ciascuno di questi sottoinsiemi con il parametro `DataRearrangement` nel `createDatasourceFromS3`, `createDatasourceFromRedShift` e nelle API `createDatasourceFromRDS`. Nel parametro `DataRearrangement`, si specifica il sottoinsieme di dati da includere in un'origine dati indicando dove iniziare e terminare ogni segmento. Per creare le origini dati complementari richieste per una 4k-fold cross-validation, si specifica il parametro `DataRearrangement` come nell'esempio seguente:

Modello uno:

Origine dati per la valutazione:

```
{"splitting":{"percentBegin":0, "percentEnd":25}}
```

Origine dati per l'addestramento:

```
{"splitting":{"percentBegin":0, "percentEnd":25, "complement":"true"}}
```

Modello due:

Origine dati per la valutazione:

```
{"splitting":{"percentBegin":25, "percentEnd":50}}
```

Origine dati per l'addestramento:

```
{"splitting":{"percentBegin":25, "percentEnd":50, "complement":"true"}}
```


Modello tre:

Origine dati per la valutazione:

```
{"splitting":{"percentBegin":50, "percentEnd":75}}
```

Origine dati per l'addestramento:

```
{"splitting":{"percentBegin":50, "percentEnd":75, "complement":"true"}}
```

Modello quattro:

Origine dati per la valutazione:

```
{"splitting":{"percentBegin":75, "percentEnd":100}}
```

Origine dati per l'addestramento:

```
{"splitting":{"percentBegin":75, "percentEnd":100, "complement":"true"}}
```

L'esecuzione di una 4-fold cross-validation genera quattro modelli, quattro origini dati per addestrare i modelli, quattro origini dati per valutare i modelli e quattro valutazioni, una per ogni modello. Amazon ML genera un parametro delle prestazioni del modello per ogni valutazione. Ad esempio, in una 4-fold cross-validation per un problema di classificazione binaria, ciascuna delle valutazioni segnala un parametro AUC (Area Under Curve). È possibile ottenere le prestazioni complessive misurate calcolando la media dei quattro parametri AUC. Per ulteriori informazioni sul parametro AUC, consultare [Misurazione dell'accuratezza del modello ML](#).

Per il codice di esempio che mostra come creare una convalida incrociata e ottenere la media dei punteggi del modello, consultare il [Codice di esempio Amazon ML](#).

Regolazione dei modelli

Dopo aver effettuato la convalida incrociata dei modelli, è possibile regolare le impostazioni per il modello successivo se le prestazioni del modello non sono all'altezza delle aspettative. Per ulteriori informazioni sull'overfitting, consultare [Fitting del modello: Underfitting vs. overfitting](#). Per ulteriori informazioni sulla regolarizzazione, consultare [Regolarizzazione](#). Per ulteriori informazioni sulla modifica delle impostazioni di regolarizzazione, consultare [Creazione di un modello ML con opzioni personalizzate](#).

Avvisi relativi alla valutazione

Amazon ML fornisce informazioni per consentire di convalidare la correttezza della valutazione del modello. Se la valutazione non soddisfa uno qualsiasi dei criteri di convalida, la console di Amazon ML avvisa l'utente visualizzando il criterio di convalida che è stato violato, come segue.

- La valutazione del modello ML è eseguita su dati utilizzati

Amazon ML avvisa se si usa la stessa origine dati per l'addestramento e per la valutazione. Se si usa Amazon ML per dividere i dati, tale criterio di validità viene soddisfatto. Se non si utilizza Amazon ML per dividere i dati, assicurarsi di valutare il modello ML con un'origine dati diversa da quella utilizzata per l'addestramento.

- Sono stati utilizzati dati sufficienti per valutare il modello predittivo

Amazon ML avvisa se il numero di osservazioni/record nella valutazione dei dati è inferiore al 10% del numero di osservazioni presenti nell'origine dati di addestramento. Per valutare correttamente il modello, è importante fornire un campione di dati sufficientemente grande. Questo criterio fornisce un controllo per informare l'utente se sta utilizzando dati troppo esigui. La quantità di dati richiesti per valutare il modello ML è soggettiva. In questo caso viene selezionato provvisoriamente il 10% in assenza di una misura migliore.

- Schema abbinato

Amazon ML avvisa se lo schema dell'origine dati per l'addestramento e quello per la valutazione non coincidono. Se si hanno alcuni attributi che non esistono nell'origine dati di valutazione o se si hanno attributi aggiuntivi, Amazon ML visualizza questo avviso.

- Tutti i record dei file di valutazione sono stati utilizzati per la valutazione delle prestazioni del modello predittivo

È importante sapere se tutti i record forniti per la valutazione siano stati effettivamente utilizzati per valutare il modello. Amazon ML avvisa se alcuni record dell'origine dati di valutazione non erano validi e non sono stati inclusi nel calcolo del parametro dell'accuratezza. Ad esempio, se manca la variabile di destinazione in alcune delle osservazioni dell'origine dati di valutazione, Amazon ML non è in grado di verificare se le previsioni del modello ML per tali osservazioni siano corrette. In questo caso, i record con valori di destinazione mancanti sono considerati non validi.

- Distribuzione della variabile di destinazione

Amazon ML mostra la distribuzione dell'attributo di destinazione delle origini dati di addestramento e di valutazione, per consentire di verificare se la destinazione sia distribuita in modo analogo

in entrambe le origini dati. Se il modello è stato addestrato su dati di addestramento con una distribuzione di destinazione diversa da quella della destinazione dei dati di valutazione, la qualità della valutazione potrebbe risentirne perché viene calcolata su dati con statistiche molto diverse. È consigliabile distribuire i dati di addestramento e quelli di valutazione in modo analogo e far sì che tali set di dati simulino quanto più possibile i dati che il modello incontrerà durante l'esecuzione delle previsioni.

Se questo avviso viene attivato, provare a usare la strategia di divisione casuale per suddividere i dati in origini dati di addestramento e di valutazione. In rari casi, questo messaggio potrebbe erroneamente avvisare l'utente riguardo a differenze nella distribuzione di destinazione anche se i dati sono stati divisi in modo casuale. Amazon ML impiega statistiche di dati approssimative per valutare le distribuzioni dei dati e talvolta attiva questo avviso per errore.

Generazione e interpretazione delle previsioni

Amazon ML offre due meccanismi per generare previsioni: asincrono (basato su batch) e sincrono (uno alla volta).

Si utilizzano le previsioni asincrone o previsioni in batch, quando si dispone di un dato numero di osservazioni e si desidera ottenere previsioni per tutte le osservazioni contemporaneamente. Il processo utilizza un'origine dati come input, mentre le previsioni di output sono raccolte in un file .csv archiviato in un bucket S3 a scelta. È necessario attendere il completamento del processo di previsione in batch prima di poter accedere ai risultati delle previsioni. La dimensione massima di un'origine dati che Amazon ML è in grado di elaborare in un file batch è di 1 TB (circa 100 milioni di record). Se l'origine dati è di dimensioni superiori a 1 TB, il processo fallirà e Amazon ML restituirà un codice di errore. Per evitare questo problema, occorre dividere i dati in più batch. Se i record sono in genere più lunghi, il limite di 1 TB verrà raggiunto prima dell'elaborazione di 100 milioni di record. In questo caso, si consiglia di contattare [AWS Support](#) per aumentare le dimensioni del processo per la previsione in batch.

Si utilizzano le previsioni sincrone o in tempo reale, per ottenere previsioni a bassa latenza. L'API di previsione in tempo reale accetta una sola osservazione di input serializzata come stringa JSON e restituisce in modo sincrono la previsione e i metadati associati come parte della risposta dell'API. È possibile richiamare l'API simultaneamente più di una volta per ottenere previsioni sincrone in parallelo. Per ulteriori informazioni sui limiti di throughput dell'API di previsione in tempo reale, consultare la sezione sui limiti di previsione in tempo reale nella [documentazione di riferimento delle API di Amazon ML](#).

Argomenti

- [Creazione di una previsione in batch](#)
- [Revisione dei parametri delle previsioni in batch](#)
- [Lettura dei file di output delle previsioni in batch](#)
- [Richiesta di previsioni in tempo reale](#)

Creazione di una previsione in batch

Per creare una previsione in batch, è necessario creare una previsione in `batch.BatchPrediction` oggetto utilizzando la console o l'API di Amazon Machine Learning

(Amazon ML). `UNBatchPredictionobject` descrive un insieme di previsioni che Amazon ML genera utilizzando il modello ML e una serie di osservazioni di input. Quando crei un `UNBatchPrediction` oggetto, Amazon ML avvia un flusso di lavoro asincrono che calcola le previsioni.

È necessario utilizzare lo stesso schema per l'origine dati utilizzata per ottenere previsioni in batch e l'origine dati che è stata utilizzata per addestrare il modello di ML soggetto alle query per le previsioni. L'unica eccezione è che non serve che l'origine dati per una previsione in batch includa l'attributo di destinazione, poiché Amazon ML prevede la destinazione. Se si fornisce l'attributo di destinazione, Amazon ML ignora il suo valore.

Creazione di una previsione in batch (console)

Per creare una previsione Batch con la console di Amazon ML, utilizzare la procedura guidata Crea previsione in batch.

Per creare una previsione in batch (console)

1. Accedi alla `AWS Management Console` e apri la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Sul pannello di controllo Amazon ML, sotto `Oggetti`, scegli `Creare nuovo...` e quindi scegli `previsione in Batch`.
3. Scegliere il modello Amazon ML che si desidera utilizzare per creare la previsione in batch.
4. Per confermare che si desidera usare quel modello, scegliere `Continua`.
5. Scegliere l'origine dati che si desidera utilizzare per creare le previsioni. L'origine dati deve avere lo stesso schema del modello, anche se non occorre includere l'attributo di destinazione.
6. Scegliere `Continue (Continua)`.
7. Per `S3 destination (destinazione S3)`, digitare il nome del bucket S3.
8. Scegliere `Review (Rivedi)`.
9. Rivedere le impostazioni e scegliere `Create batch prediction (Crea previsione in batch)`.

Creazione di una previsione in batch (API)

Creazione di un `UNBatchPrediction` utilizzando l'API Amazon ML, è necessario fornire i seguenti parametri:

Datasource ID

L'ID dell'origine dati che punta alle osservazioni di cui si vogliono ottenere le previsioni. Ad esempio, se si desiderano le previsioni per i dati in un file denominato `s3://examplebucket/input.csv`, è necessario creare un oggetto origine dati che punti al file di dati e quindi passare l'ID di tale origine dati con questo parametro.

BatchPrediction ID

L'ID da assegnare alla previsione in batch.

ML Model ID

L'ID del modello ML su cui Amazon ML dovrebbe eseguire query relative alle previsioni.

Output Uri

L'URI del bucket S3 in cui memorizzare l'output della previsione. Amazon ML deve disporre delle autorizzazioni per scrivere i dati in questo bucket.

Il parametro `OutputUri` deve fare riferimento a un percorso di S3 che termina con una barra (/), come nell'esempio seguente:

```
s3://examplebucket/examplepath/
```

Per ulteriori informazioni sulla configurazione delle autorizzazioni S3, consultare [Concessione ad Amazon ML delle autorizzazioni per generare previsioni in Amazon S3](#).

BatchPrediction Name (facoltativo)

(Facoltativo) Un nome in formato leggibile per la previsione in batch.

Revisione dei parametri delle previsioni in batch

Dopo che Amazon Machine Learning (Amazon ML) ha creato una previsione in batch, fornisce due parametri: `Records seen` e `Records failed to process`. `Records seen` indica il numero di record che Amazon ML ha esaminato quando ha eseguito la previsione in batch. `Records failed to process` indica il numero di record che Amazon ML non è stato in grado di elaborare.

Per consentire ad Amazon ML di elaborare i record non elaborati, verificare la formattazione dei record nei dati utilizzati per creare l'origine dati e accertarsi che tutti gli attributi richiesti siano presenti e che tutti i dati siano corretti. Dopo aver sistemato i dati, è possibile ricreare la previsione in batch.

oppure creare una nuova origine dati con i record non elaborati, e quindi creare una nuova previsione in batch utilizzando la nuova origine dati.

Revisione dei parametri delle previsioni in batch (console)

Per vedere le metriche nella console Amazon ML, apri il Riepilogo delle previsioni in Batch pagina e guarda nell'Informazioni elaborate sezione.

Revisione dei parametri e dei dettagli delle previsioni in batch (API)

È possibile utilizzare le API Amazon ML per recuperare i dettagli `BatchPrediction` oggetti, inclusi i parametri dei record. Amazon ML fornisce le seguenti chiamate API per le previsioni in batch:

- `CreateBatchPrediction`
- `UpdateBatchPrediction`
- `DeleteBatchPrediction`
- `GetBatchPrediction`
- `DescribeBatchPredictions`

Per ulteriori informazioni, consulta la [.Informazioni di riferimento delle API Amazon ML.](#)

Lettura dei file di output delle previsioni in batch

Eseguire la procedura seguente per recuperare i file di output delle previsioni in batch:

1. Individuare il file manifest delle previsioni in batch.
2. Leggere il file manifest per determinare le posizioni dei file di output.
3. Recuperare i file di output che contengono le previsioni.
4. Interpretare i contenuti dei file di output. I contenuti variano in base al tipo di modello ML utilizzato per generare le previsioni.

Le seguenti sezioni descrivono in modo più dettagliato le varie fasi.

Individuazione del file manifest delle previsioni in batch.

I file manifest delle previsioni in batch contengono le informazioni che mappano i file di input ai file di output delle previsioni.

Per individuare il file manifest, iniziare con il percorso di output specificato al momento della creazione dell'oggetto previsioni in batch. È possibile eseguire query su un oggetto previsioni in batch completato per recuperare la posizione S3 di questo file utilizzando l'[API di Amazon ML](https://console.aws.amazon.com/machinelearning/) o il <https://console.aws.amazon.com/machinelearning/>.

Il file manifest si trova nella posizione di output di un percorso che comprende la stringa statica /batch-prediction/ collegata al percorso di output e il nome del file manifest, ovvero l'ID della previsione in batch, con l'estensione collegata .manifest.

Ad esempio, se si crea un oggetto previsione in batch con l'ID bp-example e si specifica la posizione S3 s3://examplebucket/output/ come posizione di output, si potrà trovare il file manifest qui:

```
s3://examplebucket/output/batch-prediction/bp-example.manifest
```

Lettura del file manifest

Il contenuto del file manifest è codificato come mappa JSON, dove la chiave è una stringa del nome di un file di dati di input S3 e il valore è una stringa del file associato dei risultati delle previsioni in batch. Esiste una sola riga di mappatura per ogni coppia di file di input/output. Proseguendo con l'esempio, se l'input per la creazione dell'oggetto BatchPrediction è costituito da un singolo file denominato data.csv che si trova nella posizione s3://examplebucket/input/, potrebbe essere visualizzata una stringa di mappatura simile alla seguente:

```
{"s3://examplebucket/input/data.csv":  
s3://examplebucket/output/batch-prediction/result/bp-example-data.csv.gz"}
```

Se l'input per la creazione dell'oggetto BatchPrediction è costituito da tre file denominati data1.csv, data2.csv e data3.csv, tutti memorizzati nella posizione S3 s3://examplebucket/input/, potrebbe essere visualizzata una stringa di mappatura simile alla seguente:

```
{"s3://examplebucket/input/data1.csv": "s3://examplebucket/output/batch-prediction/  
result/bp-example-data1.csv.gz",  
  
"s3://examplebucket/input/data2.csv": "  
s3://examplebucket/output/batch-prediction/result/bp-example-data2.csv.gz",  
  
"s3://examplebucket/input/data3.csv": "  
s3://examplebucket/output/batch-prediction/result/bp-example-data3.csv.gz"}
```


Recupero dei file di output delle previsioni in batch

È possibile scaricare ciascuno dei file della previsione in batch ottenuti dalla mappatura del manifest ed elaborarli in locale. I file sono in formato CSV, compressi con l'algoritmo gzip. All'interno di tale file vi è una sola riga per osservazione input nel corrispondente file di input.

Per unire le previsioni con i file di input della previsione in batch, è possibile eseguire una semplice unione record per record dei due file. Il file di output della previsione in batch contiene sempre lo stesso numero di record del file di input della previsione, nello stesso ordine. Se l'elaborazione di un'osservazione di input ha esito negativo e non è possibile generare alcuna previsione, il file di output della previsione in batch avrà una riga vuota nella posizione corrispondente.

Interpretazione dei contenuti dei file della previsione in batch per un modello ML di classificazione binaria

Le colonne del file della previsione in batch per un modello di classificazione binaria vengono denominate `bestAnswer` (migliore risposta) e `score` (punteggio).

La colonna `bestAnswer` (migliore risposta) contiene l'etichetta di previsione ("1" o "0") ottenuta valutando il punteggio della previsione sulla base del punteggio soglia. Per ulteriori informazioni sui punteggi soglia, consultare [Regolazione del punteggio soglia](#). È possibile impostare un punteggio soglia per il modello ML utilizzando l'API Amazon ML o la funzionalità di valutazione del modello presente nella console di Amazon ML. Se non si imposta un punteggio soglia, Amazon ML usa il valore predefinito di 0,5.

La colonna `score` contiene il punteggio di previsione non elaborato assegnato dal modello ML a questa previsione. Amazon ML usa modelli di regressione logistica, perciò questo punteggio tenta di modellare la probabilità che l'osservazione corrisponda a un valore true («1»). Lo score (punteggio) è riportato in notazione scientifica, perciò nella prima riga dell'esempio seguente il valore `8.7642E-3` è pari a 0,0087642.

Ad esempio, se il punteggio soglia per il modello ML è 0,75, il contenuto del file di output della previsione in batch per un modello di classificazione binaria potrebbe avere questo aspetto:

```
bestAnswer,score
0,8.7642E-3
1,7.899012E-1
```

```
0,6.323061E-3
```

```
0,2.143189E-2
```

```
1,8.944209E-1
```

La seconda e la quinta osservazione nel file di input hanno ottenuto punteggi di previsione superiori a 0,75, perciò la colonna `bestAnswer` relativa a tali osservazioni indica il valore "1", mentre altre osservazioni hanno il valore "0".

Interpretazione dei contenuti dei file della previsione in batch per un modello ML di classificazione multiclasse

Il file della previsione in batch per un modello multiclasse contiene una colonna per ogni classe presente nei dati di addestramento. I nomi delle colonne compaiono nella riga di intestazione del file della previsione in batch.

Quando si richiedono previsioni a un modello multiclasse, Amazon ML calcola vari punteggi di previsione per ogni osservazione nel file di input, uno per ciascuna delle classi definite nel set di dati di input. È come chiedere "Quant'è la probabilità (misurata tra 0 e 1) che questa osservazione rientri in questa classe, rispetto a una qualsiasi delle altre classi?" Ogni punteggio può essere interpretato come la "probabilità che l'osservazione appartenga a questa classe." Poiché i punteggi di previsione modellano le probabilità sottostanti che l'osservazione appartenga a una classe o a un'altra, la somma di tutti i punteggi di previsione di una riga è 1. È necessario selezionare una classe come classe prevista per il modello. Nella maggior parte dei casi, si seleziona la classe che ha la massima probabilità di essere la risposta migliore.

A titolo illustrativo, si prenda in considerazione il tentativo di prevedere la valutazione che un cliente assegnerà a un prodotto, sulla base di una scala di stelle da 1 a 5. Se le classi sono denominate `1_star`, `2_stars`, `3_stars`, `4_stars` e `5_stars`, il file di output della previsione multiclasse potrebbe avere questo aspetto:

```
1_star, 2_stars, 3_stars, 4_stars, 5_stars
```

```
8.7642E-3, 2.7195E-1, 4.77781E-1, 1.75411E-1, 6.6094E-2
```

```
5.59931E-1, 3.10E-4, 2.48E-4, 1.99871E-1, 2.39640E-1
```

```
7.19022E-1, 7.366E-3, 1.95411E-1, 8.78E-4, 7.7323E-2  
1.89813E-1, 2.18956E-1, 2.48910E-1, 2.26103E-1, 1.16218E-1  
3.129E-3, 8.944209E-1, 3.902E-3, 7.2191E-2, 2.6357E-2
```

In questo esempio, la prima osservazione ha il miglior punteggio di previsione per la classe `3_stars` (punteggio previsione = $4.77781E-1$), pertanto è possibile interpretare i risultati come un'attestazione del fatto che la classe `3_stars` è la risposta migliore per questa osservazione. Si noti che i punteggi di previsione sono riportati in notazione scientifica, perciò un punteggio di previsione $4.77781E-1$ è pari a 0,477781.

È possibile che in determinate circostanze non si desideri scegliere la classe con la massima probabilità. Ad esempio, si potrebbe voler stabilire una soglia minima al di sotto della quale non si considera una classe come la risposta migliore anche se ha il punteggio di previsione più alto. Si supponga di classificare i film in generi e di volere che il punteggio di previsione sia almeno $5E-1$ prima di dichiarare che il genere è la risposta migliore. Si ottiene un punteggio di previsione $3E-1$ per Commedie, $2.5E-1$ per Drammatici, $2.5E-1$ per Documentari e $2E-1$ per Film d'azione. In questo caso, il modello ML prevede che Commedie sia la scelta più probabile, ma si decide di non sceglierla come risposta migliore. Poiché nessuno dei punteggi di previsione ha superato il punteggio di previsione di base $5E-1$, si decide che la previsione è insufficiente per determinare con sicurezza il genere e si decide di scegliere qualcos'altro. L'applicazione potrebbe quindi trattare il campo Genere per questo film come "sconosciuto".

Interpretazione dei contenuti dei file della previsione in batch per un modello ML di regressione

Il file della previsione in batch per un modello di regressione contiene una sola colonna denominata `score` (punteggio). Questa colonna contiene la previsione numerica non elaborata per ogni osservazione nei dati di input. I valori sono riportati in notazione scientifica, pertanto il valore dello `score` (punteggio) $-1.526385E1$ è pari a -15.26835 nella prima riga dell'esempio seguente.

Questo esempio illustra un file di output per una previsione in batch eseguita su un modello di regressione:

```
score  
-1.526385E1
```

```
-6.188034E0  
  
-1.271108E1  
  
-2.200578E1  
  
8.359159E0
```

Richiesta di previsioni in tempo reale

Una previsione in tempo reale è una chiamata sincrona ad Amazon Machine Learning (Amazon ML). La previsione viene effettuata quando Amazon ML riceve la richiesta e la risposta viene fornita immediatamente. Le previsioni in tempo reale vengono comunemente utilizzate per abilitare le funzionalità predittive all'interno di applicazioni interattive Web, mobili o desktop. È possibile eseguire query in un modello ML creato con Amazon ML per le previsioni in tempo reale utilizzando la bassa latenza `PredictAPI`. L'operazione `Predict` accetta una sola osservazione di input nel payload della richiesta e restituisce la previsione in modo sincrono nella risposta. Ciò la distingue dall'API di previsione in batch, che viene richiamata con l'ID di un oggetto origine dati Amazon ML che punta alla posizione delle osservazioni di input, e restituisce in modo asincrono un URI a un file che contiene le previsioni per tutte queste osservazioni. Amazon ML risponde alla maggior parte delle richieste di previsione in tempo reale entro 100 millisecondi.

È possibile provare le previsioni in tempo reale senza costi aggiuntivi nella console di Amazon ML. Se si decide di utilizzare le previsioni in tempo reale, è necessario creare innanzi tutto un endpoint per la generazione delle previsioni in tempo reale. È possibile farlo nella console di Amazon ML o utilizzando il `CreateRealtimeEndpointAPI`. Quando si dispone di un endpoint, utilizzare l'API di previsione in tempo reale per generare le previsioni in tempo reale.

Note

Dopo la creazione di un endpoint in tempo reale per il modello, verrà addebitato un costo di prenotazione di capacità che dipenderà dalle dimensioni del modello. Per ulteriori informazioni, consultare [Prezzi](#). Se si decide di creare l'endpoint in tempo reale nella console, la console visualizzerà una ripartizione dei costi stimati addebitati in modo regolare per l'endpoint. Per interrompere l'addebito dei costi quando non sarà più necessario ottenere previsioni in tempo reale da quel modello, eliminare l'endpoint in tempo reale tramite la console o l'operazione `DeleteRealtimeEndpoint`.

Ad esempio, per richieste e risposte, consulta [Stime](#) nel Riferimento API Amazon Machine Learning. Per vedere un esempio del formato esatto di risposta che utilizza il modello, consultare [Prova delle previsioni in tempo reale](#).

Argomenti

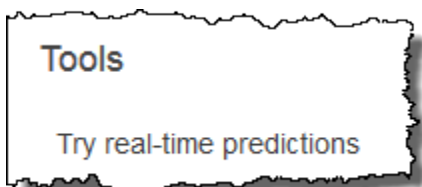
- [Prova delle previsioni in tempo reale](#)
- [Creazione di un endpoint in tempo reale](#)
- [Individuazione dell'endpoint delle previsioni in tempo reale \(console\)](#)
- [Individuazione dell'endpoint delle previsioni in tempo reale \(API\)](#)
- [Creazione di una richiesta di previsioni in tempo reale](#)
- [Eliminazione di un endpoint in tempo reale](#)

Prova delle previsioni in tempo reale

Per consentire di decidere se abilitare la previsione in tempo reale, Amazon ML permette di provare la generazione di previsioni su singoli record di dati senza incorrere nei costi aggiuntivi associati alla configurazione di un endpoint di previsione in tempo reale. Per provare la previsione in tempo reale è necessario disporre di un modello ML. Per creare previsioni in tempo reale su larga scala, utilizzare la [Stime API](#) nel Riferimento API Amazon Machine Learning.

Per provare le previsioni in tempo reale

1. Accedi alla [AWS Management Console](#) e apri la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, nell'elenco a discesa Amazon Machine Learning, scegliere ML models (Modelli ML).
3. Scegliere il modello che si desidera usare per provare le previsioni in tempo reale, come ad esempio il `Subscription propensity model` dal tutorial.
4. Nella pagina del report del modello ML, in Predictions (Previsioni), scegliere Summary (Riepilogo) e Try real-time predictions (Prova previsioni in tempo reale).



Amazon ML mostra l'elenco delle variabili nei record di dati che Amazon ML ha utilizzato per addestrare il modello.

5. È possibile procedere immettendo i dati in tutti i campi del modulo o incollando un singolo record di dati, in formato CSV, nella casella di testo.

Per utilizzare il modulo, per ogni campo Value (Valore), immettere i dati che si desidera utilizzare per testare le previsioni in tempo reale. Se il record di dati che si sta inserendo non contiene valori per uno o più attributi di dati, lasciare vuoti i campi di immissione.

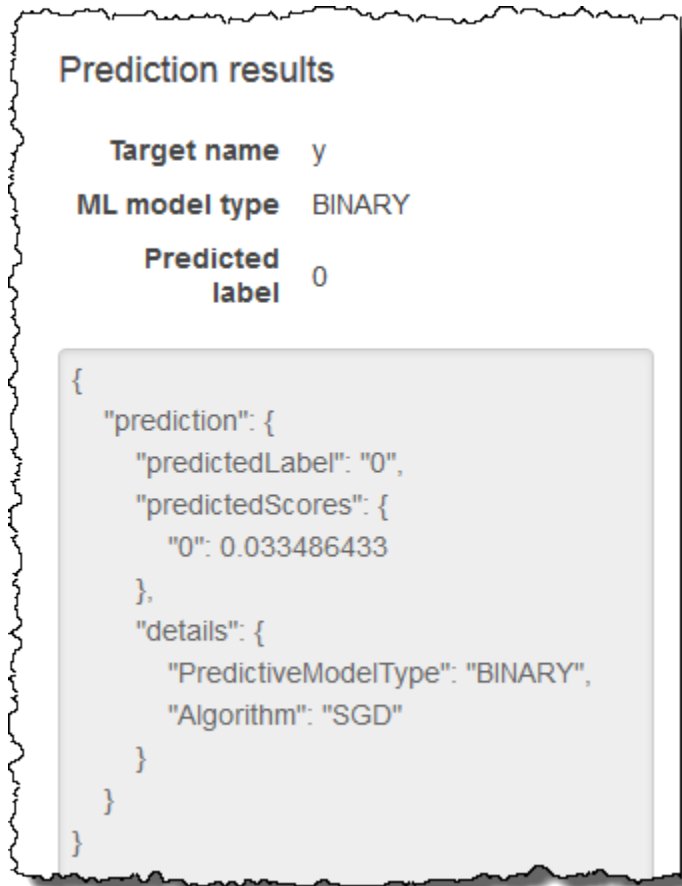
Per fornire un record di dati, scegliere Paste a record (Incolla un record). Incollare una singola riga di dati in formato CSV nel campo di testo e scegliere Invia. Amazon ML popola automaticamente la Valore campi per te.

Note

I dati del record di dati devono avere lo stesso numero di colonne dei dati di addestramento ed essere disposti nello stesso ordine. L'unica eccezione è che è necessario omettere il valore di destinazione. Se si include un valore di destinazione, Amazon ML lo ignorerà.

6. Nella parte inferiore della pagina, scegliere Create prediction (Crea previsione). Amazon ML restituisce la previsione immediatamente.

Nel riquadro Prediction results (Risultati della previsione), si vede l'oggetto previsione restituito dalla chiamata API Predict, insieme al tipo di modello ML, al nome della variabile di destinazione e alla classe o al valore previsto. Per informazioni sull'interpretazione dei risultati, consultare [Interpretazione dei contenuti dei file della previsione in batch per un modello ML di classificazione binaria](#).



Creazione di un endpoint in tempo reale

Per generare previsioni in tempo reale, è necessario creare un endpoint in tempo reale. Per creare un endpoint in tempo reale, occorre disporre già di un modello ML per il quale si desidera generare previsioni in tempo reale. È possibile creare un endpoint in tempo reale utilizzando la console di Amazon ML o chiamando il `CreateRealtimeEndpointAPI`. Per ulteriori informazioni sull'utilizzo del `CreateRealtimeEndpointAPI`, consulta https://docs.aws.amazon.com/machine-learning/latest/APIReference/API_CreateRealtimeEndpoint.html (riferimento delle API di Amazon Machine Learning).

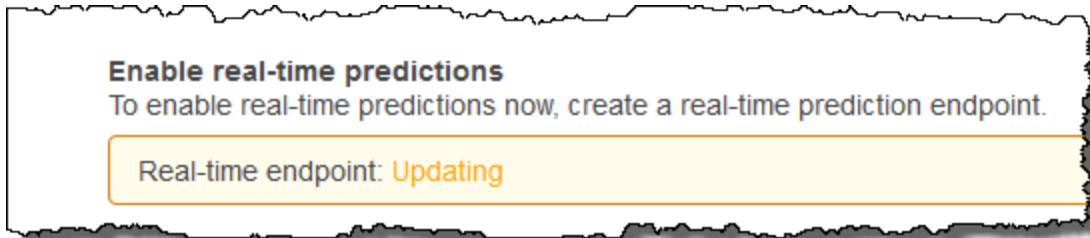
Per creare un endpoint in tempo reale

1. Accedi alla AWS Management Console e apri la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, nell'elenco a discesa Amazon Machine Learning, scegliere ML models (Modelli ML).
3. Scegliere il modello per cui si desidera generare previsioni in tempo reale.

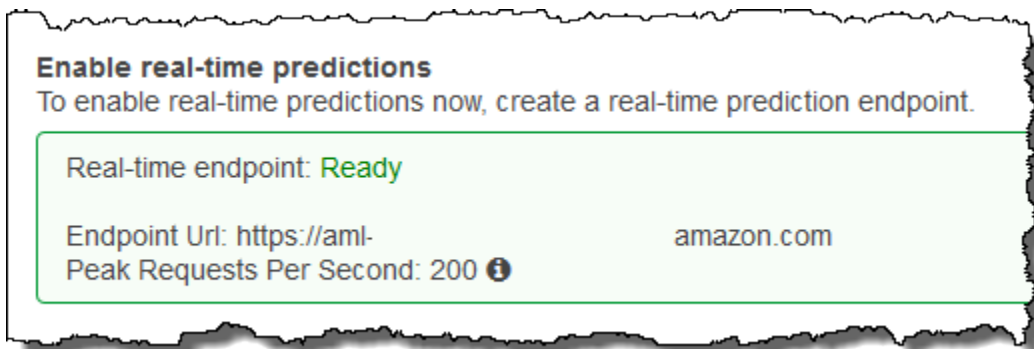
4. Nella pagina ML model summary (Riepilogo del modello ML), in Predictions (Previsioni), scegliere Create real-time endpoint (Crea endpoint in tempo reale).

Verrà visualizzata una finestra di dialogo che spiega i prezzi applicati alle previsioni in tempo reale.

5. Scegliere Create (Crea). La richiesta di endpoint in tempo reale viene inviata ad Amazon ML e inserita in una coda. Lo stato dell'endpoint in tempo reale è Updating (Aggiornamento in corso).



6. Quando l'endpoint in tempo reale è pronto, lo stato cambia in Pronto e Amazon ML visualizza l'URL dell'endpoint. Utilizzare l'URL dell'endpoint per creare richieste di previsioni in tempo reale con l'API Predict. Per ulteriori informazioni sull'utilizzo dell'API Predict, consulta https://docs.aws.amazon.com/machine-learning/latest/APIReference/API_Predict.html riferimento delle API di Amazon Machine Learning.



Individuazione dell'endpoint delle previsioni in tempo reale (console)

Per utilizzare la console di Amazon ML per trovare l'URL dell'endpoint di un modello ML, accedere alla Riepilogo del modello ML (Certificato creato).

Per individuare l'URL di un endpoint in tempo reale

1. Accedi alla AWS Management Console e apri la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.

2. Nella barra di navigazione, nell'elenco a discesa Amazon Machine Learning, scegliere ML models (Modelli ML).
3. Scegliere il modello per cui si desidera generare previsioni in tempo reale.
4. Nella pagina ML model summary (Riepilogo del modello ML), scorrere verso il basso fino a visualizzare la sezione Predictions (Previsioni).
5. L'URL dell'endpoint del modello è elencato in Real-time prediction (Previsione in tempo reale). Utilizzare l'URL come Endpoint Url (URL dell'endpoint) per le chiamate delle previsioni in tempo reale. Per informazioni su come utilizzare l'endpoint per generare previsioni, consultare https://docs.aws.amazon.com/machine-learning/latest/APIReference/API_Predict.html riferimento delle API di Amazon Machine Learning.

Individuazione dell'endpoint delle previsioni in tempo reale (API)

Quando viene creato un endpoint in tempo reale utilizzando l'operazione `CreateRealtimeEndpoint`, l'URL e lo stato dell'endpoint vengono restituiti nella risposta. Se l'endpoint in tempo reale è stato creato utilizzando la console o se si desidera recuperare l'URL e lo stato di un endpoint creato in precedenza, richiamare l'operazione `GetMLModel` con l'ID del modello su cui si desidera eseguire query per le previsioni in tempo reale. Le informazioni sull'endpoint sono contenute nella sezione `EndpointInfo` della risposta. Per un modello che dispone di un endpoint in tempo reale, l'`EndpointInfo` potrebbe avere questo aspetto:

```
"EndpointInfo":{
  "CreatedAt": 1427864874.227,
  "EndpointStatus": "READY",
  "EndpointUrl": "https://endpointUrl",
  "PeakRequestsPerSecond": 200
}
```

Un modello senza un endpoint in tempo reale restituirebbe quanto segue:

```
EndpointInfo":{
  "EndpointStatus": "NONE",
  "PeakRequestsPerSecond": 0
}
```

Creazione di una richiesta di previsioni in tempo reale

Un esempio di payload della richiesta `Predict` potrebbe avere questo aspetto:

```
{
  "MLModelId": "model-id",
  "Record":{
    "key1": "value1",
    "key2": "value2"
  },
  "PredictEndpoint": "https://endpointUrl"
}
```

La `PredictEndpoint` deve corrispondere alla `EndpointUrl` campo del `EndpointInfo` struttura. Amazon ML utilizza questo campo per instradare la richiesta ai server appropriati del parco istanze delle previsioni in tempo reale.

Il `MLModelId` è l'identificatore di un modello addestrato in precedenza con un endpoint in tempo reale.

Un `Record` è la mappatura dei nomi delle variabili ai valori delle variabili. Ogni coppia rappresenta un'osservazione. La `Record` map contiene gli input del modello Amazon ML. È analoga a una singola riga di dati del set di dati di addestramento, senza la variabile di destinazione. Indipendentemente dal tipo di valori nei dati di addestramento, `Record` contiene una mappatura stringa-a-stringa.

Note

È possibile omettere le variabili per cui non si dispone di un valore, anche se ciò potrebbe ridurre l'accuratezza della previsione. Più variabili si possono includere, più è accurato il modello.

Il formato della risposta restituita dalle richieste `Predict` dipende dal tipo di modello su cui sono eseguite le query di previsione. In tutti i casi, il campo `details` contiene informazioni sulla richiesta di previsione, in particolare il campo `PredictiveModelType` con il tipo di modello.

L'esempio seguente mostra una risposta per un modello binario:

```
{
  "Prediction":{
    "details":{
      "PredictiveModelType": "BINARY"
    },
    "predictedLabel": "0",
  }
}
```

```
    "predictedScores":{
      "0": 0.47380468249320984
    }
  }
}
```

Avvertenze `predictedLabel` che contiene l'etichetta prevista, in questo caso 0. Amazon ML calcola l'etichetta prevista confrontando il punteggio di previsione alla classificazione limite:

- Si può ottenere la classificazione limite attualmente associata a un modello ML controllando `laScoreThreshold` nella risposta del `GetMLModel` funzionamento o visualizzando le informazioni sul modello nella console Amazon ML. Se non si imposta un punteggio soglia, Amazon ML utilizza il valore predefinito di 0,5.
- È possibile ottenere l'esatto punteggio di previsione per un modello di classificazione binaria controllando la mappa `predictedScores`. In questa mappa, l'etichetta prevista viene associata all'esatto punteggio di previsione.

Per ulteriori informazioni sulle previsioni binarie, consultare [Interpretazione delle previsioni](#).

L'esempio seguente mostra una risposta per un modello di regressione. Si noti che il valore numerico previsto si trova nel campo `predictedValue`:

```
{
  "Prediction":{
    "details":{
      "PredictiveModelType": "REGRESSION"
    },
    "predictedValue": 15.508452415466309
  }
}
```

L'esempio seguente mostra una risposta per un modello multiclasse:

```
{
  "Prediction":{
    "details":{
      "PredictiveModelType": "MULTICLASS"
    },
    "predictedLabel": "red",
    "predictedScores":{
```

```
        "red": 0.12923571467399597,  
        "green": 0.08416014909744263,  
        "orange": 0.22713537514209747,  
        "blue": 0.1438363939523697,  
        "pink": 0.184102863073349,  
        "violet": 0.12816807627677917,  
        "brown": 0.10336143523454666  
    }  
}  
}
```

Analogamente ai modelli di classificazione binaria, l'etichetta/classe prevista si trova nel campo `predictedLabel`. È possibile comprendere meglio la forte correlazione esistente tra la previsione e ogni classe osservando la mappa `predictedScores`. Maggiore è il punteggio di una classe all'interno di questa mappa, più forte è la correlazione della previsione con la classe, dove il valore più alto alle fine viene selezionato come `predictedLabel`.

Per ulteriori informazioni sulle previsioni multiclasse, consultare [Informazioni del modello multiclasse](#).

Eliminazione di un endpoint in tempo reale

Dopo avere completato le previsioni in tempo reale, eliminare l'endpoint in tempo reale per evitare di incorrere in costi aggiuntivi. Gli addebiti si interrompono non appena si elimina l'endpoint.

Per eliminare un endpoint in tempo reale

1. Accedi alla AWS Management Console e apri la console Amazon Machine Learning all'indirizzo <https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, nell'elenco a discesa Amazon Machine Learning, scegliere ML models (Modelli ML).
3. Scegliere il modello che non richiede più le previsioni in tempo reale.
4. Nella pagina del report del modello ML, in Predictions (Previsioni), scegliere Summary (Riepilogo).
5. Scegliere Delete real-time endpoint (Elimina endpoint in tempo reale).
6. Nella finestra di dialogo Delete real-time endpoint (Elimina endpoint in tempo reale), scegliere Delete (Elimina).

Gestione di oggetti Amazon ML

Amazon ML offre quattro oggetti che è possibile gestire tramite la console Amazon ML o l'API Amazon ML:

- Origini dati
- Modelli ML
- Valutazioni
- Previsioni in batch

Ogni oggetto svolge una funzione diversa nel ciclo di vita della creazione di un'applicazione di machine learning e ogni oggetto ha attributi e funzionalità specifici che si applicano solo a quell'oggetto. Nonostante queste differenze, è necessario gestire gli oggetti in modo simile. Ad esempio, è possibile usare i processi quasi identici per elencare gli oggetti, recuperare le loro descrizioni e aggiornarli o eliminarli.

Le seguenti sezioni descrivono le operazioni di gestione comuni a tutti e quattro gli oggetti ed evidenziano le eventuali differenze.

Argomenti

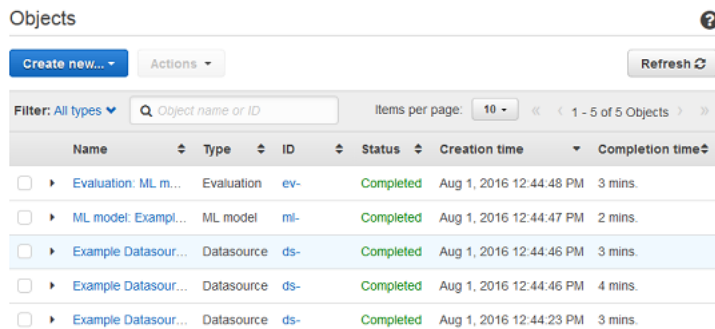
- [Elenco degli oggetti](#)
- [Recupero delle descrizioni degli oggetti](#)
- [Aggiornamento di oggetti](#)
- [Eliminazione di oggetti](#)

Elenco degli oggetti

Per avere informazioni approfondite sulle origini dati Amazon Machine Learning (Amazon ML), i modelli ML, le valutazioni e le previsioni in batch, è possibile elencarli. Per ogni oggetto verrà visualizzato il nome, il tipo, l'ID, il codice di stato e l'ora di creazione. È anche possibile visualizzare i dettagli specifici di un determinato tipo di oggetto. Ad esempio, è possibile visualizzare le informazioni sui dati di un'origine dati.

Elencazione degli oggetti (Console)

Per consultare un elenco degli ultimi 1.000 oggetti creati, nella console Amazon ML aprire il **Oggetti** pannello di controllo. Per visualizzare **Oggetti** Pannello di controllo, accedere alla console Amazon ML.



Objects

Create new... Actions Refresh

Filter: All types Items per page: 10 << < 1 - 5 of 5 Objects > >>

Name	Type	ID	Status	Creation time	Completion time
<input type="checkbox"/> ▶ Evaluation: ML m...	Evaluation	ev-	Completed	Aug 1, 2016 12:44:48 PM	3 mins.
<input type="checkbox"/> ▶ ML model: Examl...	ML model	ml-	Completed	Aug 1, 2016 12:44:47 PM	2 mins.
<input type="checkbox"/> ▶ Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:46 PM	3 mins.
<input type="checkbox"/> ▶ Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:46 PM	4 mins.
<input type="checkbox"/> ▶ Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:23 PM	3 mins.

Per visualizzare ulteriori informazioni su un oggetto, inclusi i dettagli specifici di quel tipo di oggetto, scegliere il nome o l'ID dell'oggetto. Ad esempio, per visualizzare Data insights (Informazioni dati) per un'origine dati, scegliere il nome dell'origine dati.

Le colonne del pannello di controllo Objects (Oggetti) mostrano le seguenti informazioni riguardo a ogni oggetto.

Nome

Il nome dell'oggetto.

Type (Tipo)

Il tipo di oggetto. I valori validi includono Datasource (Origine dati), ML model (Modello ML), Evaluation (Valutazione) e Batch prediction (Previsione in batch).

Note

Per vedere se un modello è configurato per supportare previsioni in tempo reale, visitare la pagina ML model summary (Riepilogo del modello ML) selezionando il nome o l'ID del modello.

ID

L'ID dell'oggetto.

Stato

Lo stato dell'oggetto. I valori includono Pending (In sospeso), In Progress (In corso), Completed (Completato) e Failed (Non riuscito). Se lo stato è Failed (Non riuscito), controllare i dati e riprovare.

Creation time

La data e l'ora in cui Amazon ML ha terminato la creazione di questo oggetto.

Completion time

Il tempo impiegato da Amazon ML per creare questo oggetto. È possibile utilizzare il tempo di completamento di un modello per stimare il tempo di addestramento di un nuovo modello.

Datasource ID

Per gli oggetti creati utilizzando un'origine dati, ad esempio modelli e valutazioni, l'ID dell'origine dati. Se si cancella l'origine dati, non è più possibile utilizzare i modelli ML creati con quell'origine dati per creare previsioni.

È possibile ordinare per qualsiasi colonna selezionando l'icona del doppio triangolo accanto all'intestazione della colonna.

Elencazione degli oggetti (API)

Nella [API Amazon ML](#) è possibile elencare gli oggetti per tipo, utilizzando le seguenti operazioni:

- `DescribeDataSources`
- `DescribeMLModels`
- `DescribeEvaluations`
- `DescribeBatchPredictions`

Ogni operazione include i parametri di filtraggio, ordinamento e paginazione per un lungo elenco di oggetti. Non esiste alcun limite al numero di oggetti a cui è possibile accedere tramite l'API. Per limitare le dimensioni dell'elenco, utilizzare il parametro `Limit`, che può accettare un valore massimo di 100.

La risposta dell'API a un comando `Describe*` include un token di paginazione (`nextPageToken`), se necessario, e brevi descrizioni di ogni oggetto. Le descrizioni degli oggetti includono le stesse

informazioni per ciascuno dei tipi di oggetto che vengono visualizzate nella console, inclusi i dettagli specifici per un tipo di oggetto.

Note

Anche se la risposta include un minor numero di oggetti rispetto al limite specificato, è possibile includere un `nextPageToken` che indica che vi sono altri risultati disponibili. Anche una risposta che contiene 0 elementi potrebbe contenere un `nextPageToken`.

Per ulteriori informazioni, consulta la [.Informazioni di riferimento delle API Amazon ML.](#)

Recupero delle descrizioni degli oggetti

È possibile visualizzare descrizioni dettagliate di qualsiasi oggetto tramite la console o tramite l'API.

Descrizioni dettagliate nella console

Per visualizzare le descrizioni nella console, accedere a un elenco per un determinato tipo di oggetto (origine dati, modello ML, valutazione o previsione in batch). Quindi, individuare la riga nella tabella che corrisponde all'oggetto, sfogliando l'elenco o ricercando il suo nome o l'ID.

Descrizioni dettagliate dall'API

Ogni tipo di oggetto dispone di un'operazione che recupera i dettagli completi di un oggetto Amazon ML:

- `GetDataSource`
- `GetMLModel`
- `GetEvaluation`
- `GetBatchPrediction`

Ogni operazione richiede esattamente due parametri: l'ID dell'oggetto e un flag booleano denominato `Verbose`. Le chiamate con `Verbose` impostato su `true` includeranno ulteriori dettagli sull'oggetto, con conseguenti latenze più elevate e risposte di maggiori dimensioni. Per scoprire quali campi sono inclusi impostando il flag `Verbose`, consultare la pagina relativa alla [documentazione di riferimento delle API di Amazon ML.](#)

Aggiornamento di oggetti

Ogni tipo di oggetto dispone di un'operazione che aggiorna i dettagli di un oggetto Amazon ML (consultare la pagina relativa alla documentazione di [Guida di riferimento all'API Amazon ML](#)):

- UpdateDataSource
- UpdateMLModel
- UpdateEvaluation
- UpdateBatchPrediction


Ogni operazione richiede l'ID dell'oggetto, per specificare quale l'oggetto viene aggiornato. È possibile aggiornare i nomi di tutti gli oggetti. Non è possibile aggiornare le altre proprietà degli oggetti per le origini dati, le valutazioni e le previsioni in batch. Per i modelli ML, è possibile aggiornare il campo ScoreThreshold, purché il modello ML non abbia un endpoint di previsione in tempo reale associato.

Eliminazione di oggetti

Quando le origini dati, i modelli ML, le valutazioni e le previsioni in batch non servono più, è possibile eliminarli. Anche se non è applicato alcun costo aggiuntivo qualora si tengano oggetti Amazon ML diversi dalle previsioni in batch dopo che sono stati utilizzati, l'eliminazione di oggetti tiene il workspace ordinato e più facile da gestire. È possibile eliminare uno o più oggetti utilizzando la console di Amazon Machine Learning (Amazon ML) o l'API.

Warning

Quando si eliminano oggetti Amazon ML, l'effetto è immediato, permanente e irreversibile.

Objects 

Create new... Actions Refresh

Filter: All types Items per page: 10 << 1 - 5 of 5 Objects >>

Name	Type	ID	Status	Creation time	Completion time
<input type="checkbox"/> Evaluation: ML m...	Evaluation	ev-	Completed	Aug 1, 2016 12:44:48 PM	3 mins.
<input type="checkbox"/> ML model: Examl...	ML model	ml-	Completed	Aug 1, 2016 12:44:47 PM	2 mins.
<input type="checkbox"/> Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:46 PM	3 mins.
<input type="checkbox"/> Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:46 PM	4 mins.
<input type="checkbox"/> Example Datasour...	Datasource	ds-	Completed	Aug 1, 2016 12:44:23 PM	3 mins.

Eliminazione di oggetti (console)

È possibile utilizzare la console Amazon ML per eliminare gli oggetti, inclusi modelli. La procedura utilizzata per eliminare un modello varia se si sta utilizzando il modello per generare previsioni in tempo reale oppure no. Per eliminare un modello utilizzato per generare previsioni in tempo reale, eliminare per primo l'endpoint in tempo reale.

Per eliminare oggetti Amazon ML (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Selezionare gli oggetti Amazon ML che si intendono eliminare. Per selezionare più di un oggetto, utilizzare il tasto MAIUSC. Per deselezionare tutti gli oggetti selezionati, utilizzare il pulsante



o



3. In Actions (Azioni), scegliere Delete (Elimina).
4. Nella finestra di dialogo, scegliere Delete (Elimina) per eliminare il modello.

Per eliminare un modello Amazon ML con un endpoint in tempo reale (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Selezionare il modello che si intende eliminare.
3. In Azioni scegliere Delete real-time endpoint (Elimina endpoint in tempo reale).
4. Scegliere Delete (Elimina) per eliminare l'endpoint.
5. Selezionare di nuovo il modello.
6. In Actions (Azioni), scegliere Delete (Elimina).
7. Scegliere Delete (Elimina) per eliminare il modello.

Eliminazione di oggetti (API)

È possibile eliminare oggetti Amazon ML utilizzando le seguenti chiamate API:

- `DeleteDataSource` - Prende il parametro `DataSourceId`.

- `DeleteMLModel` - Prende il parametro `MLModelId`.
- `DeleteEvaluation` - Prende il parametro `EvaluationId`.
- `DeleteBatchPrediction` - Prende il parametro `BatchPredictionId`.

Per ulteriori informazioni, consultare la pagina relativa alla [documentazione di riferimento delle API di Amazon Machine Learning](#).

Monitoraggio di Amazon ML con i parametri di Amazon CloudWatch

Amazon ML invia automaticamente parametri ad Amazon CloudWatch per raccogliere e analizzare le statistiche di utilizzo per i modelli ML. Ad esempio, per tenere traccia delle previsioni in batch e in tempo reale, è possibile monitorare il parametro PredictCount secondo la dimensione RequestMode. I parametri vengono automaticamente raccolti e inviati ad Amazon CloudWatch ogni cinque minuti. È possibile monitorare questi parametri utilizzando la console di Amazon CloudWatch, l'interfaccia a riga di comando AWS o SDK AWS.

Non è previsto alcun costo per i parametri di Amazon ML segnalati attraverso CloudWatch. Se si impostano allarmi sui parametri, saranno addebitate le [tariffe CloudWatch](#) standard.

Per ulteriori informazioni, consultare l'elenco dei parametri di Amazon ML nella pagina di [riferimento di parametri, dimensioni e spazi dei nomi di Amazon CloudWatch](#) della guida per sviluppatori di Amazon CloudWatch.

Registrazione delle chiamate API di Amazon ML conAWS CloudTrail

Amazon Machine Learning (Amazon ML) è integrato conAWS CloudTrail, un servizio che fornisce un record delle operazioni eseguite da un utente, un ruolo o unAWSservizio in Amazon ML. CloudTrail acquisisce tutte le chiamate API per Amazon ML come eventi. Le chiamate acquisite includono le chiamate dalla console Amazon ML e le chiamate di codice alle operazioni API Amazon ML. Se crei un trail, puoi abilitare la distribuzione continua di eventi CloudTrail in un bucket Amazon S3, inclusi gli eventi per Amazon ML. Se non si configura un trail, è comunque possibile visualizzare gli eventi più recenti nella console di CloudTrail in Event history (Cronologia eventi). Le informazioni raccolte da CloudTrail consentono di determinare la richiesta effettuata ad Amazon ML, l'indirizzo IP da cui è stata eseguita la richiesta, l'autore della richiesta, il momento in cui è stata eseguita e altri dettagli.

Per ulteriori informazioni su CloudTrail, incluso come configurarlo e abilitarlo, consulta la [Guida per l'utente di AWS CloudTrail](#).

Informazioni su Amazon ML in CloudTrail

CloudTrail è abilitato sull'account AWS al momento della sua creazione. Quando si verifica un'attività supportata in Amazon ML, questa viene registrata in un evento CloudTrail insieme ad altri.AWSeventi del servizio inCronologia eventi. È possibile visualizzare, cercare e scaricare gli eventi recenti nell'account AWS. Per ulteriori informazioni, consulta [Visualizzazione di eventi nella cronologia degli eventi di CloudTrail](#).

Per una registrazione continuativa di attività ed eventi inAWSl'account, inclusi gli eventi per Amazon ML, crea un trail. Un trail consente a CloudTrail di distribuire i file di log in un bucket Amazon S3. Per impostazione predefinita, quando si crea un trail nella console, il trail sarà valido in tutte le regioni AWS. Il trail registra gli eventi di tutte le Regioni nella partizione AWS e distribuisce i file di registro nel bucket Amazon S3 specificato. Inoltre, è possibile configurare altri servizi AWS per analizzare con maggiore dettaglio e usare i dati evento raccolti nei registri CloudTrail. Per ulteriori informazioni, consulta gli argomenti seguenti:

- [Panoramica della creazione di un percorso](#)
- [Servizi e integrazioni CloudTrail supportati](#)
- [Configurazione delle notifiche Amazon SNS per CloudTrail](#)
- [Ricezione di file di log CloudTrail da più regioni](#) e [Ricezione di file di log CloudTrail da più account](#)

Amazon ML supporta la registrazione delle operazioni seguenti come eventi nei file di log CloudTrail:

- [AddTags](#)
- [CreateBatchPrediction](#)
- [CreateDataSourceFromRDS](#)
- [CreateDataSourceFromRedshift](#)
- [CreateDataSourceFromS3](#)
- [CreateEvaluation](#)
- [CreateMLModel](#)
- [CreateRealtimeEndpoint](#)
- [DeleteBatchPrediction](#)
- [DeleteDataSource](#)
- [DeleteEvaluation](#)
- [DeleteMLModel](#)
- [DeleteRealtimeEndpoint](#)
- [DeleteTags](#)
- [DescribeTags](#)
- [UpdateBatchPrediction](#)
- [UpdateDataSource](#)
- [UpdateEvaluation](#)
- [UpdateMLModel](#)

Le seguenti operazioni Amazon ML usano i parametri della richiesta che contengono le credenziali. Prima che queste richieste vengano inviate a CloudTrail, le credenziali sono sostituite da tre asterischi («***»):

- [CreateDataSourceFromRDS](#)
- [CreateDataSourceFromRedshift](#)

Quando le seguenti operazioni Amazon ML sono eseguite con la console Amazon ML, l'attributo `ComputeStatistics` non è incluso nel `RequestParameters` componente del log CloudTrail:

- [CreateDataSourceFromRedshift](#)
- [CreateDataSourceFromS3](#)

Ogni evento o voce del registro contiene informazioni sull'utente che ha generato la richiesta. Le informazioni di identità consentono di determinare quanto segue:

- Se la richiesta è stata effettuata con credenziali utente root o AWS Identity and Access Management (IAM).
- Se la richiesta è stata effettuata con le credenziali di sicurezza temporanee per un ruolo o un utente federato.
- Se la richiesta è stata effettuata da un altro servizio AWS.

Per ulteriori informazioni, consulta [Elemento userIdentity di CloudTrail](#).

Esempio: Voci del file di log di Amazon ML

Un trail è una configurazione che consente l'implementazione di eventi come i file di log in un bucket Amazon S3 che specifichi. I file di registro di CloudTrail possono contenere una o più voci di registro. Un evento rappresenta una singola richiesta da un'origine e include informazioni sull'operazione richiesta, sulla data e sull'ora dell'operazione, sui parametri richiesti e così via. I file di log CloudTrail non sono una traccia di pila ordinata delle chiamate API pubbliche e di conseguenza non devono apparire in base a un ordine specifico.

L'esempio seguente mostra una voce di log di CloudTrail che illustra l'operazione .

```
{
  "Records": [
    {
      "eventVersion": "1.03",
      "userIdentity": {
        "type": "IAMUser",
        "principalId": "EX_PRINCIPAL_ID",
        "arn": "arn:aws:iam::012345678910:user/Alice",
        "accountId": "012345678910",
        "accessKeyId": "EXAMPLE_KEY_ID",
        "userName": "Alice"
      },
      "eventTime": "2015-11-12T15:04:02Z",
```

```

"eventSource": "machinelearning.amazonaws.com",
"eventName": "CreateDataSourceFromS3",
"awsRegion": "us-east-1",
"sourceIPAddress": "127.0.0.1",
"userAgent": "console.amazonaws.com",
"requestParameters": {
  "data": {
    "dataLocationS3": "s3://aml-sample-data/banking-batch.csv",
    "dataSchema": "{\"version\":\"1.0\",\"rowId\":null,\"rowWeight
\":null,
    \"targetAttributeName\":null,\"dataFormat\":\"CSV\",
    \"dataFileContainsHeader\":false,\"attributes\":[
      {\"attributeName\":\"age\",\"attributeType\":\"NUMERIC\"},
      {\"attributeName\":\"job\",\"attributeType\":\"CATEGORICAL
\"},
      {\"attributeName\":\"marital\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"education\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"default\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"housing\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"loan\",\"attributeType\":\"CATEGORICAL
\"},
      {\"attributeName\":\"contact\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"month\",\"attributeType\":\"CATEGORICAL
\"},
      {\"attributeName\":\"day_of_week\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"duration\",\"attributeType\":\"NUMERIC
\"},
      {\"attributeName\":\"campaign\",\"attributeType\":\"NUMERIC
\"},
      {\"attributeName\":\"pdays\",\"attributeType\":\"NUMERIC\"},
      {\"attributeName\":\"previous\",\"attributeType\":\"NUMERIC
\"},
      {\"attributeName\":\"poutcome\",\"attributeType\":
\"CATEGORICAL\"},
      {\"attributeName\":\"emp_var_rate\",\"attributeType\":
\"NUMERIC\"},
      {\"attributeName\":\"cons_price_idx\",\"attributeType\":
\"NUMERIC\"},

```



```

        {"attributeName": "cons_conf_idx", "attributeType":
\"NUMERIC\"},
        {"attributeName": "euribor3m", "attributeType": \"NUMERIC
\"},
        {"attributeName": "nr_employed", "attributeType":
\"NUMERIC\"}
    ], "excludedAttributeNames": []}
  },
  "dataSourceId": "exampleDataSourceId",
  "dataSourceName": "Banking sample for batch prediction"
},
"responseElements": {
  "dataSourceId": "exampleDataSourceId"
},
"requestID": "9b14bc94-894e-11e5-a84d-2d2deb28fdec",
"eventID": "f1d47f93-c708-495b-bff1-cb935a6064b2",
"eventType": "AwsApiCall",
"recipientAccountId": "012345678910"
},
{
  "eventVersion": "1.03",
  "userIdentity": {
    "type": "IAMUser",
    "principalId": "EX_PRINCIPAL_ID",
    "arn": "arn:aws:iam::012345678910:user/Alice",
    "accountId": "012345678910",
    "accessKeyId": "EXAMPLE_KEY_ID",
    "userName": "Alice"
  },
  "eventTime": "2015-11-11T15:24:05Z",
  "eventSource": "machinelearning.amazonaws.com",
  "eventName": "CreateBatchPrediction",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "127.0.0.1",
  "userAgent": "console.amazonaws.com",
  "requestParameters": {
    "batchPredictionName": "Batch prediction: ML model: Banking sample",
    "batchPredictionId": "exampleBatchPredictionId",
    "batchPredictionDataSourceId": "exampleDataSourceId",
    "outputUri": "s3://EXAMPLE_BUCKET/BatchPredictionOutput/",
    "mlModelId": "exampleModelId"
  },
  "responseElements": {
    "batchPredictionId": "exampleBatchPredictionId"
  }
}

```

```
    },  
    "requestID": "3e18f252-8888-11e5-b6ca-c9da3c0f3955",  
    "eventID": "db27a771-7a2e-4e9d-bfa0-59deee9d936d",  
    "eventType": "AwsApiCall",  
    "recipientAccountId": "012345678910"  
  }  
]  
}
```

Tagging dei tuoi oggetti Amazon ML

Organizzare e gestire gli oggetti Amazon Machine Learning (Amazon ML) assegnando loro metadata con tag. Un tag è una coppia chiave-valore definita per un oggetto.

Oltre a usare i tag per organizzare e gestire gli oggetti Amazon ML, è possibile utilizzarli per categorizzare e monitorare i costi di AWS. Quando si applicano tag agli oggetti AWS, compresi i modelli ML, il report di allocazione dei costi di AWS include l'utilizzo e i costi aggregati in base ai tag. Applicando tag che rappresentano categorie di business (come centri di costo, nomi di applicazioni o proprietari), è possibile organizzare i costi tra più servizi. Per ulteriori informazioni, consulta [Utilizzo dei tag per l'allocazione dei costi ai fini dei report di fatturazione personalizzati](#) nella AWS Billing User Guide (Guida per l'utente di Amazon API Gateway).

Indice

- [Nozioni di base sui tag](#)
- [Limitazioni applicate ai tag](#)
- [Tagging degli oggetti Amazon ML \(console\)](#)
- [Tagging degli oggetti Amazon ML \(API\)](#)

Nozioni di base sui tag

Utilizzare i tag per categorizzare gli oggetti e facilitarne la gestione. Ad esempio, è possibile categorizzare gli oggetti in base allo scopo, al proprietario o all'ambiente. Quindi, è possibile definire un set di tag che consenta di monitorare i modelli in base al proprietario e all'applicazione associata. Di seguito sono riportati vari esempi:

- Progetto: Nome progetto
- Proprietario: Nome
- Scopo: Previsioni di marketing
- Applicazione: Nome applicazione
- Ambiente: Produzione

È possibile utilizzare la console o l'API di Amazon ML per completare le seguenti attività:

- Aggiungere tag a un oggetto

- Visualizzare i tag degli oggetti
- Modificare i tag degli oggetti
- Eliminare i tag da un oggetto

Per impostazione predefinita, i tag applicati a un oggetto Amazon ML vengono copiati negli oggetti creati utilizzando quell'oggetto. Ad esempio, se un'origine dati Amazon Simple Storage Service (Amazon S3) ha un costo di marketing: «campagna di marketing mirata», un modello creato con quell'origine dati avrebbe un «Costo di marketing: Tag «campagna di marketing mirata», così come la valutazione del modello. In questo modo è possibile usare i tag per tenere traccia degli oggetti correlati, come ad esempio tutti gli oggetti utilizzati per una campagna di marketing. Se si verifica un conflitto tra le origini dei tag, ad esempio un modello con il tag «Costo di marketing: Campagna di marketing mirata» e un'origine dati con il tag «Costo di marketing: Clienti marketing target», Amazon ML applica il tag dal modello.

Limitazioni applicate ai tag

Ai tag si applicano le limitazioni seguenti.

Limitazioni di base:

- Il numero massimo di tag per ogni oggetto è 50.
- i valori e le chiavi dei tag rispettano la distinzione tra maiuscole e minuscole;
- Non è possibile cambiare o modificare i tag di un oggetto eliminato.

Limitazioni applicate alle chiavi di tag:

- Ogni chiave di tag deve essere univoca. Se si aggiunge un tag con una chiave già in uso, il nuovo tag sovrascrive la coppia chiave-valore di quell'oggetto.
- Non è possibile avviare una chiave di tag con `aws :` perché questo prefisso è riservato all'uso da parte di AWS. AWS crea tag che iniziano con questo prefisso, ma è possibile modificarli o eliminarli.
- Le chiavi di tag devono avere una lunghezza compresa tra 1 e 128 caratteri Unicode.
- I tasti tag devono essere composti dai seguenti caratteri: Lettere Unicode, cifre, spazio e i seguenti caratteri speciali: `_ . / = + - @`.

Limitazioni applicate ai valori dei tag.

- I valori dei tag devono avere una lunghezza compresa tra 0 e 255 caratteri Unicode.
- I valori dei tag possono essere vuoti. In caso contrario, devono essere costituiti dai seguenti caratteri: lettere Unicode, cifre, spazio e i seguenti caratteri speciali: _ . / = + - @.

Tagging degli oggetti Amazon ML (console)

È possibile visualizzare, aggiungere, modificare ed eliminare i tag utilizzando la console di Amazon ML.

Per visualizzare i tag di un oggetto (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, espandere il selettore delle regioni e scegliere una regione.
3. Nella pagina Objects (Oggetti), scegliere un oggetto.
4. Scorrere fino alla sezione Tags (Tag) dell'oggetto selezionato. I tag di quell'oggetto sono elencati nella parte inferiore della sezione.

Per aggiungere un tag a un oggetto (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, espandere il selettore delle regioni e scegliere una regione.
3. Nella pagina Objects (Oggetti), scegliere un oggetto.
4. Scorrere fino alla sezione Tags (Tag) dell'oggetto selezionato. I tag di quell'oggetto sono elencati nella parte inferiore della sezione.
5. Scegli Add or edit tags (Aggiungi o modifica tag).
6. In Add Tag (Aggiungi tag), specificare la chiave di tag nel campo Key (Chiave), specificare (facoltativo) un valore di tag nel campo Value (Valore), quindi scegliere Apply changes (Applica modifiche).

Se il pulsante Apply changes (Applica modifiche) non è abilitato, la chiave di tag o il valore di tag specificato non soddisfa le limitazioni per i tag. Per ulteriori informazioni, consulta la pagina [Limitazioni applicate ai tag](#) .

7. Per visualizzare il nuovo tag nell'elenco della sezione Tags (Tag), aggiornare la pagina.

Per modificare un tag (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, espandere il selettore delle regioni e selezionare una regione.
3. Nella pagina Objects (Oggetti), scegliere un oggetto.
4. Scorrere fino alla sezione Tags (Tag) dell'oggetto selezionato. I tag di quell'oggetto sono elencati nella parte inferiore della sezione.
5. Scegli Add or edit tags (Aggiungi o modifica tag).
6. In Applied tags (Tag applicati), modificare il valore di un tag nel campo Value (Valore), quindi scegliere Apply changes (Applica modifiche).

Se il pulsante Apply changes (Applica modifiche) non è abilitato, il valore di tag specificato non soddisfa le limitazioni per i tag. Per ulteriori informazioni, consulta la pagina [Limitazioni applicate ai tag](#).

7. Per visualizzare il tag aggiornato nell'elenco della sezione Tags (Tag), aggiornare la pagina.

Per eliminare un tag da un oggetto (console)

1. Accedi allaAWS Management Consolee apri la console Amazon Machine Learning all'indirizzo<https://console.aws.amazon.com/machinelearning/>.
2. Nella barra di navigazione, espandere il selettore delle regioni e scegliere una regione.
3. Nella pagina Objects (Oggetti), scegliere un oggetto.
4. Scorrere fino alla sezione Tags (Tag) dell'oggetto selezionato. I tag di quell'oggetto sono elencati nella parte inferiore della sezione.
5. Scegli Add or edit tags (Aggiungi o modifica tag).
6. In Applied tag (Tag applicato), selezionare il tag che si desidera eliminare, quindi scegliere Apply changes (Applica modifiche).

Tagging degli oggetti Amazon ML (API)

È possibile aggiungere, elencare ed eliminare i tag utilizzando l'API di Amazon ML. Per alcuni esempi, consultare la seguente documentazione:

[AddTags](#)

Consente di aggiungere o modificare i tag per l'oggetto specificato.

[DescribeTags](#)

Elenca i tag per l'oggetto specificato.

[DeleteTags](#)

Elimina i tag dall'oggetto specificato.

Riferimento Amazon Machine Learning

Argomenti

- [Concessione ad Amazon ML delle autorizzazioni per leggere i dati da Amazon S3](#)
- [Concessione ad Amazon ML delle autorizzazioni per generare previsioni in Amazon S3](#)
- [Controllo dell'accesso alle risorse Amazon ML con IAM](#)
- [Prevenzione del confused deputy tra servizi](#)
- [Gestione delle dipendenze nelle operazioni asincrone](#)
- [Verifica dello stato della richiesta](#)
- [Limiti di sistema](#)
- [Nomi e ID per tutti gli oggetti](#)
- [Durate degli oggetti](#)

Concessione ad Amazon ML delle autorizzazioni per leggere i dati da Amazon S3

Per creare un oggetto origine dati dai dati di input in Amazon S3, è necessario concedere ad Amazon ML le autorizzazioni seguenti per la posizione S3 in cui i dati di input sono memorizzati:

- `GetObject`permessi per il bucket e il prefisso S3.
- `ListBucket`permessi per il bucket S3. A differenza di altre azioni, `ListBucket`devono essere concesse autorizzazioni a livello di bucket (anziché sul prefisso). Tuttavia, è possibile ampliare la portata dell'autorizzazione a un determinato prefisso utilizzando una clausola `Condition` (Condizione).

Se si utilizza la console di Amazon ML per creare l'origine dati, queste autorizzazioni possono essere aggiunte al bucket per l'utente. Ti verrà richiesto di confermare se desideri aggiungerli man mano che completi i passaggi della procedura guidata. La seguente policy di esempio mostra come concedere ad Amazon ML l'autorizzazione a leggere i dati dalla posizione di esempio `s3://esempio di secchio/prefisso di esempio`, durante l'esplorazione del `ListBucket`permesso solo al *prefisso di esempio* percorso di input.

```
{
  "Version": "2008-10-17",
```



```
"Statement": [
  {
    "Effect": "Allow",
    "Principal": { "Service": "machinelearning.amazonaws.com" },
    "Action": "s3:GetObject",
    "Resource": "arn:aws:s3:::examplebucket/exampleprefix/*"
    "Condition": {
      "StringEquals": { "aws:SourceAccount": "123456789012" }
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
    }
  },
  {
    "Effect": "Allow",
    "Principal": {"Service": "machinelearning.amazonaws.com"},
    "Action": "s3:ListBucket",
    "Resource": "arn:aws:s3:::examplebucket",
    "Condition": {
      "StringLike": { "s3:prefix": "exampleprefix/*" }
      "StringEquals": { "aws:SourceAccount": "123456789012" }
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
    }
  }
]
```

Per applicare questa policy ai dati, è necessario modificare l'istruzione di policy associata al bucket S3 in cui i dati sono archiviati.

Per modificare la policy delle autorizzazioni per un bucket S3 (tramite la console precedente)

1. Accedi alla AWS Management Console e apri la console di Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Selezionare il nome del bucket in cui sono archiviati i dati.
3. Scegli Properties (Proprietà).
4. Scegliere Edit bucket policy (Modifica policy bucket).
5. Inserire la policy riportata sopra, personalizzarla secondo le proprie esigenze, quindi scegliere Save (Salva).
6. Seleziona Salva.

Per modificare la policy delle autorizzazioni per un bucket S3 (tramite la nuova console)

1. Accedi alla AWS Management Console e apri la console di Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Selezionare il nome del bucket, quindi Permissions (Autorizzazioni).
3. Scegli Bucket Policy (Policy del bucket).
4. Inserire la policy mostrata sopra, personalizzandola secondo le proprie esigenze.
5. Seleziona Salva.

Concessione ad Amazon ML delle autorizzazioni per generare previsioni in Amazon S3

Per generare i risultati dell'operazione di previsione in batch in Amazon S3, è necessario concedere ad Amazon ML le seguenti autorizzazioni per la posizione dei risultati, che viene fornita come input dell'operazione Crea previsione in batch:

- GetObjectpermessi per bucket e prefisso S3.
- PutObjectpermessi per bucket e prefisso S3.
- PutObjectAclsul bucket e sul prefisso S3.
 - Amazon ML necessita di questa autorizzazione per garantire che possa concedere il prodotto in scatola [ACL](#) bucket-owner-full-control autorizzazione al tuo account AWS, dopo la creazione degli oggetti.
- ListBucketpermessi per il bucket S3. A differenza di altre azioni, ListBucketdevono essere concesse autorizzazioni a livello di bucket (anziché sul prefisso). Tuttavia, è possibile ampliare la portata dell'autorizzazione a un determinato prefisso utilizzando una clausola Condition (Condizione).

Se si utilizza la console di Amazon ML per creare la richiesta di previsione in batch, queste autorizzazioni possono essere aggiunte al bucket per l'utente. All'utente verrà richiesto di confermare se desidera aggiungerli al termine della procedura guidata.

La seguente policy di esempio mostra come concedere ad Amazon ML l'autorizzazione a scrivere dati nella posizione di esempio s3://examplebucket/exampleprefix, definendo l'ambito delListBucketautorizzazione al solo percorso di input del prefisso di esempio e autorizzazione ad Amazon ML di impostare gli ACL degli oggetti put sul prefisso di output:

```
{
```

```
"Version": "2008-10-17",
"Statement": [
  {
    "Effect": "Allow",
    "Principal": { "Service": "machinelearning.amazonaws.com"},
    "Action": [
      "s3:GetObject",
      "s3:PutObject"
    ],
    "Resource": "arn:aws:s3:::examplebucket/exampleprefix/*"
    "Condition": {
      "StringEquals": { "aws:SourceAccount": "123456789012" }
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
    }
  },
  {
    "Effect": "Allow",
    "Principal": { "Service": "machinelearning.amazonaws.com"},
    "Action": "s3:PutObjectAcl",
    "Resource": "arn:aws:s3:::examplebucket/exampleprefix/*",
    "Condition": {
      "StringEquals": { "s3:x-amz-acl": "bucket-owner-full-control" }
      "StringEquals": { "aws:SourceAccount": "123456789012" }
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
    }
  },
  {
    "Effect": "Allow",
    "Principal": {"Service": "machinelearning.amazonaws.com"},
    "Action": "s3:ListBucket",
    "Resource": "arn:aws:s3:::examplebucket",
    "Condition": {
      "StringLike": { "s3:prefix": "exampleprefix/*" }
      "StringEquals": { "aws:SourceAccount": "123456789012" }
      "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
    }
  }
]
```


- Assegnare credenziali di sicurezza univoche a ciascun utente all'interno dell'account AWS
- Controllare le autorizzazioni di ciascun utente per eseguire attività utilizzando le risorse AWS
- Condividere facilmente le risorse AWS con gli utenti nell'account AWS
- Creare ruoli per l'account AWS e gestire le autorizzazioni assegnate loro per definire gli utenti o i servizi che possono utilizzarle
- È possibile creare ruoli in IAM e gestire le autorizzazioni per controllare quali operazioni possono essere eseguite dall'entità, o dal servizio AWS, che assume il ruolo. Puoi inoltre definire quale entità può assumere quel ruolo.

Se l'organizzazione dispone già di identità IAM, è possibile utilizzarle per concedere le autorizzazioni per eseguire attività utilizzando le risorse AWS.

Per ulteriori informazioni su IAM, consulta la [Guida per l'utente di IAM](#).

Sintassi delle policy

Una policy IAM è un documento JSON costituito da una o più dichiarazioni. Ogni dichiarazione ha la seguente struttura:

```
{
  "Statement": [{
    "Effect": "effect",
    "Action": "action",
    "Resource": "arn",
    "Condition": {
      "condition operator": {
        "key": "value"
      }
    }
  }]
}
```

Una dichiarazione di policy include i seguenti elementi:

- Effetto controlla l'autorizzazione a utilizzare le risorse e operazioni API che saranno specificate successivamente nella dichiarazione. I valori validi sono Allow e Deny. Per impostazione predefinita, gli utenti IAM non dispongono dell'autorizzazione per l'utilizzo di risorse e operazioni API, pertanto tutte le richieste vengono rifiutate. Un Allow esplicito sostituisce l'impostazione predefinita. Un Deny esplicito sostituisce qualsiasi Allows.

- **Operazione:** L'elemento action corrisponde all'operazione API specifica per la quale si nega l'autorizzazione.
- **Risorsa:** La risorsa che viene modificata dall'operazione. Per specificare una risorsa nella dichiarazione, si utilizza il suo ARN (Amazon Resource Name).
- **Condition (facoltativa):** Controlla quando la tua politica sarà in vigore.

Per semplificare la creazione e la gestione delle policy IAM, puoi usare AWS Policy Generator e IAM Policy Simulator.

Specificazione delle azioni politiche IAM per Amazon ML

In una dichiarazione di policy IAM, è possibile specificare un'operazione API per qualsiasi servizio che supporta IAM. Quando crei una dichiarazione di policy per le operazioni API Amazon ML, utilizziamo `machinelearning:` al nome dell'operazione API, come illustrato negli esempi seguenti:

- `machinelearning:CreateDataSourceFromS3`
- `machinelearning:DescribeDataSources`
- `machinelearning>DeleteDataSource`
- `machinelearning:GetDataSource`

Per specificare più operazioni in una sola istruzione, separarle con la virgola:

```
"Action": ["machinelearning:action1", "machinelearning:action2"]
```

Puoi anche specificare più operazioni tramite caratteri jolly. Ad esempio, è possibile specificare tutte le operazioni il cui nome inizia con una determinata parola:

```
"Action": "machinelearning:Get*"
```

Per specificare tutte le operazioni Amazon ML, utilizza il carattere jolly `*`:

```
"Action": "machinelearning:*"
```

Per l'elenco completo delle operazioni API Amazon ML, consulta la [Amazon Machine Learning delle API](#).

Specificazione degli ARN per le risorse Amazon ML nelle policy IAM

Le dichiarazioni politiche IAM si applicano a una o più risorse. È possibile specificare le risorse per le policy attraverso i loro ARN.

Per specificare gli ARN per le risorse Amazon Machine Learning

"Resource": `arn:aws:machinelearning:region:account:resource-type/identifier`

I seguenti esempi mostrano come specificare ARN comuni.

ID origine dati: `my-s3-datasource-id`

```
"Resource":  
arn:aws:machinelearning:<region>:<your-account-id>:datasource/my-s3-datasource-id
```

ID modello ML: `my-ml-model-id`

```
"Resource":  
arn:aws:machinelearning:<region>:<your-account-id>:mlmodel/my-ml-model-id
```

ID previsione batch: `my-batchprediction-id`

```
"Resource":  
arn:aws:machinelearning:<region>:<your-account-id>:batchprediction/my-batchprediction-id
```

ID valutazione: `my-evaluation-id`

```
"Resource": arn:aws:machinelearning:<region>:<your-account-id>:evaluation/my-evaluation-id
```

Esempi di policy per Amazon MLs

Esempio 1: Consenti agli utenti di leggere i metadati delle risorse di machine learning

La seguente policy consente a un utente o a un gruppo di leggere i metadati di sorgenti dati, modelli ML, previsioni in batch e valutazioni

eseguendo [DescribeDataSources](#), [DescribeMLModels](#), [DescribeBatchPredictions](#), [DescribeEvaluations](#), [GetDat](#) sulle risorse specificate. Le autorizzazioni per le operazioni Descrivimi* non possono essere limitate a una specifica risorsa.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "machinelearning:Get*"
    ],
    "Resource": [
      "arn:aws:machinelearning:<region>:<your-account-id>:datasource/S3-DS-ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:datasource/REDSHIFT-DS-
ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:mlmodel/ML-MODEL-ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:batchprediction/BP-
ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:evaluation/EV-ID1"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "machinelearning:Describe*"
    ],
    "Resource": [
      "*"
    ]
  }
]}
```

Esempio 2: Consenti agli utenti di creare risorse di apprendimento automatico

La policy seguente permette a un utente o a un gruppo di creare origini dati di machine learning, modelli ML, previsioni in batch e valutazioni eseguendo le operazioni `CreateDataSourceFromS3`, `CreateDataSourceFromRedshift`, `CreateDataSourceFromRDS`, `CreateMLModel`, `CreateBatchPrediction` e `CreateEvaluation`. Non è possibile limitare le autorizzazioni per tali operazioni a una risorsa specifica.

```
{
  "Version": "2012-10-17",
```



```

    "Statement": [{
      "Effect": "Allow",
      "Action": [
        "machinelearning:CreateDataSourceFrom*",
        "machinelearning:CreateMLModel",
        "machinelearning:CreateBatchPrediction",
        "machinelearning:CreateEvaluation"
      ],
      "Resource": [
        "*"
      ]
    }]
  }

```

Esempio 3: Consenti agli utenti di creare (ed eliminare) endpoint in tempo reale ed eseguire previsioni in tempo reale su un modello ML

La policy seguente permette a utenti o gruppi di creare e cancellare endpoint in tempo reale e di eseguire previsioni in tempo reale per un determinato modello ML eseguendo le operazioni `CreateRealtimeEndpoint` `DeleteRealtimeEndpoint` `Predict` su tale modello.

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "machinelearning:CreateRealtimeEndpoint",
      "machinelearning>DeleteRealtimeEndpoint",
      "machinelearning:Predict"
    ],
    "Resource": [
      "arn:aws:machinelearning:<region>:<your-account-id>:mlmodel/ML-MODEL"
    ]
  }]
}

```

Esempio 4 Consenti agli utenti di aggiornare ed eliminare risorse specifiche

La policy seguente permette a un utente o gruppo di aggiornare ed eliminare risorse specifiche nell'account AWS, fornendo l'autorizzazione per eseguire le operazioni `UpdateDataSource`, `UpdateMLModel`, `UpdateBatchPrediction`, `UpdateEvaluation`, `DeleteDataSource`,

DeleteMLModel, DeleteBatchPrediction e DeleteEvaluation su tali risorse nel proprio account.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "machinelearning:Update*",
      "machinelearning>DeleteDataSource",
      "machinelearning>DeleteMLModel",
      "machinelearning>DeleteBatchPrediction",
      "machinelearning>DeleteEvaluation"
    ],
    "Resource": [
      "arn:aws:machinelearning:<region>:<your-account-id>:datasource/S3-DS-ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:datasource/REDSHIFT-DS-
ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:mlmodel/ML-MODEL-ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:batchprediction/BP-
ID1",
      "arn:aws:machinelearning:<region>:<your-account-id>:evaluation/EV-ID1"
    ]
  }]
}
```

Esempio 5 Consenti qualsiasi Amazon MLAction

La policy seguente permette a un gruppo o un gruppo di utilizzare qualsiasi operazione Amazon Machine Learning. Poiché questa policy concede l'accesso completo a tutte le risorse di machine learning, è necessario limitarla ai soli amministratori.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "machinelearning:*"
    ],
    "Resource": [
      "*"
    ]
  }]
}
```

```
}
```

Prevenzione del confused deputy tra servizi

Con "confused deputy" si intende un problema di sicurezza in cui un'entità che non dispone dell'autorizzazione per eseguire una certa operazione può costringere un'entità con più privilegi a eseguire tale operazione. In AWS, la rappresentazione cross-service può comportare il problema confused deputy. La rappresentazione tra servizi può verificarsi quando un servizio (il servizio chiamante) effettua una chiamata a un altro servizio (il servizio chiamato). Il servizio chiamante può essere manipolato per utilizzare le proprie autorizzazioni e agire sulle risorse di un altro cliente, a cui normalmente non avrebbe accesso. Per evitare ciò, AWS fornisce strumenti per poterti a proteggere i tuoi dati per tutti i servizi con entità di servizio a cui è stato concesso l'accesso alle risorse del tuo account.

Si consiglia di utilizzare `aws:SourceArn` e `aws:SourceAccount` chiavi di contesto delle condizioni globali nelle policy delle risorse per limitare le autorizzazioni con cui Amazon Machine Learning. Se il valore `aws:SourceArn` non contiene l'ID account, ad esempio un ARN di un bucket Amazon S3, è necessario utilizzare entrambe le chiavi di contesto delle condizioni globali per limitare le autorizzazioni. Se si utilizzano entrambe le chiavi di contesto delle condizioni globali e il valore `aws:SourceArn` contiene l'ID account, il valore `aws:SourceAccount` e l'account nel valore `aws:SourceArn` deve utilizzare lo stesso ID account nella stessa dichiarazione di policy. Utilizzare `aws:SourceArn` se si desidera consentire l'associazione di una sola risorsa all'accesso tra servizi. Utilizzare `aws:SourceAccount` se si desidera consentire l'associazione di qualsiasi risorsa in tale account all'uso tra servizi.

Il modo più efficace per proteggersi dal problema "confused deputy" è quello di usare la chiave di contesto della condizione globale `aws:SourceArn` con l'ARN completo della risorsa. Se non si conosce l'ARN completo della risorsa o si scelgono più risorse, è necessario utilizzare la chiave di contesto della condizione globale `aws:SourceArn` con caratteri jolly (*) per le parti sconosciute dell'ARN. Ad esempio, `arn:aws:service:*:123456789012:*`.

Nell'esempio seguente viene illustrato come utilizzare `aws:SourceArn` e `aws:SourceAccount` chiavi di contesto relative alle condizioni globali in Amazon ML per evitare il confuso problema dei deputati durante la lettura di dati da un bucket Amazon S3.

```
{
  "Version": "2008-10-17",
  "Statement": [
```

```

{
  "Effect": "Allow",
  "Principal": { "Service": "machinelearning.amazonaws.com" },
  "Action": "s3:GetObject",
  "Resource": "arn:aws:s3:::examplebucket/exampleprefix/*"
  "Condition": {
    "StringEquals": { "aws:SourceAccount": "123456789012" }
    "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
  }
},
{
  "Effect": "Allow",
  "Principal": {"Service": "machinelearning.amazonaws.com"},
  "Action": "s3:ListBucket",
  "Resource": "arn:aws:s3:::examplebucket",
  "Condition": {
    "StringLike": { "s3:prefix": "exampleprefix/*" }
    "StringEquals": { "aws:SourceAccount": "123456789012" }
    "ArnLike": { "aws:SourceArn": "arn:aws:machinelearning:us-
east-1:123456789012:*" }
  }
}]
}

```

Gestione delle dipendenze nelle operazioni asincrone

Il corretto completamento delle operazioni in batch in Amazon ML dipende da altre operazioni. Per gestire queste dipendenze, Amazon ML identifica le richieste che hanno dipendenze e verifica il completamento delle operazioni. Se le operazioni non sono state completate, Amazon ML mette da parte le richieste iniziali finché le operazioni da cui dipendono non sono state completate.

Vi sono alcune dipendenze tra le operazioni in batch. Ad esempio, prima di poter creare un modello di ML, è necessario avere creato un'origine dati con cui addestrare il modello ML. Amazon ML non è in grado di addestrare un modello ML in assenza di un'origine dati disponibile.

Tuttavia, Amazon ML supporta la gestione delle dipendenze per le operazioni asincrone. Ad esempio, non è necessario attendere che le statistiche dei dati siano state calcolate prima di poter inviare una richiesta per addestrare un modello ML sull'origine dati. Al contrario, non appena l'origine dati viene creata, è possibile inviare una richiesta per addestrare un modello ML utilizzando l'origine dati. Amazon ML non avvia l'operazione di addestramento finché le statistiche sull'origine dati non sono

state calcolate. La richiesta `createMLModel` viene messa in una coda fino a quando le statistiche non sono state calcolate; una volta completata tale l'operazione, Amazon ML prova immediatamente a eseguire l'operazione `createMLModel`. Analogamente, è possibile inviare richieste di previsione e valutazione in batch per modelli ML che non hanno completato l'addestramento.

La tabella riportata di seguito mostra i requisiti necessari a procedere con diverse azioni AmazonML

Per...	È necessario avere...
Creare un modello ML (<code>createMLModel</code>)	L'origine dati con le statistiche dei dati calcolate
Crea una previsione in batch (<code>createBatchPrediction</code>)	L'origine dati modello di ML
Crea una valutazione in batch (<code>createBatchEvaluation</code>)	L'origine dati modello di ML

Verifica dello stato della richiesta

Quando invii una richiesta, puoi controllarne lo stato con l'API Amazon Machine Learning (Amazon ML). Ad esempio, se invii un `createMLModel` richiesta, è possibile verificarne lo stato utilizzando `describeMLModel` chiamare. Amazon ML risponde con uno dei seguenti stati.

Stato	Definizione
PENDING (INSOSPESO)	<p>Amazon ML sta convalidando la richiesta.</p> <p>O</p> <p>Amazon ML è in attesa di risorse di calcolo disponibili prima di eseguire la richiesta. Questo potrebbe verificarsi se l'account ha superato il numero massimo di richieste simultanee in esecuzione di operazioni in batch. In tal caso, lo stato passa a <code>InProgress</code> quando altre richieste in esecuzione sono state completate o vengono annullate.</p>

Stato	Definizione
	O Amazon ML è in attesa di un'operazione in batch da cui la richiesta dipende per essere completata.
INPROGRESS	La richiesta è ancora in esecuzione.
COMPLETED	La richiesta è terminata e l'oggetto è pronto per essere utilizzato (modelli ML e origini dati) o visualizzato (previsioni in batch e valutazioni).
Non riuscito	Si è verificato un problema con i dati forniti o l'operazione è stata annullata. Ad esempio, se si tenta di calcolare dati statistici in un'origine dati che non è stata completata, è possibile che venga visualizzato un messaggio di stato Invalid (Non valido) o Failed (Non riuscito). Il messaggio di errore spiega perché l'operazione non è stata completata.
DELETED	L'oggetto è già stato eliminato.

Amazon ML fornisce anche informazioni su un oggetto, ad esempio quando Amazon ML ha terminato di crearlo. Per ulteriori informazioni, consulta [Elenco degli oggetti](#).

Limiti di sistema

Per fornire un servizio efficace e affidabile, Amazon ML impone determinati limiti per le richieste effettuate al sistema. La maggior parte dei problemi di ML si adatta facilmente all'interno di questi vincoli. Tuttavia, se si ritiene che l'uso di Amazon ML venga limitato da questi limiti, è possibile contattare il [servizio clienti AWS](#) e richiedere l'aumento di un limite. Ad esempio, potrebbe esserci un limite di cinque per il numero di processi che è possibile eseguire simultaneamente. Se spesso si hanno processi in coda in attesa di risorse a causa di questo limite, allora probabilmente ha senso aumentare tale limite per il proprio account.

La tabella che segue mostra i limiti di default per account in Amazon ML. Non tutti questi limiti possono essere aumentati dal servizio clienti AWS.

Tipo di limite	Limite di sistema
Dimensioni di ciascuna osservazione	100 KB
Dimensioni dei dati di addestramento*	100 GB
Dimensioni dell'input di previsioni in batch	1 TB
Dimensioni dell'input di previsioni in batch (numero di record)	100 milioni
Numero di variabili in un file di dati (schema)	1.000
Complessità delle composizioni (numero di variabili di output elaborate)	10.000
TPS per ciascun endpoint di previsione in tempo reale	200
Totale TPS per tutti gli endpoint di previsione in tempo reale	10.000
Totale RAM per tutti gli endpoint di previsione in tempo reale	10 GB
Numero di operazioni simultanee	25
Runtime più lungo per qualsiasi processo	7 giorni
Numero di classi per i modelli ML multiclasse	100
Dimensione del modello ML	Minimo 1 MB, massimo 2 GB
Numero di tag per oggetto	50

- La dimensione del file di dati è limitata per assicurare che le operazioni si concludano tempestivamente. I processi che sono in esecuzione da più di 7 giorni verranno automaticamente terminati, generando uno stato FAILED.

Nomi e ID per tutti gli oggetti

Ogni oggetto di Amazon ML deve avere un identificatore, o ID. La console di Amazon ML genera i valori di ID per l'utente, ma se questi utilizza l'API deve generare i propri valori. Ogni ID deve essere

univoco tra tutti gli oggetti di Amazon ML dello stesso tipo nell'account AWS. Ciò significa che non è possibile avere due valutazioni con lo stesso ID. È invece possibile avere una valutazione e un'origine dati con lo stesso ID, anche se non è consigliato.

È consigliabile utilizzare identificatori generati casualmente per gli oggetti, con una breve stringa come prefisso per identificare il tipo. Ad esempio, quando la console Amazon ML genera un'origine dati, assegna alla sorgente dati un ID casuale e univoco come «ds-ZSCW».luWiOxF. Questo ID è sufficientemente casuale per evitare conflitti all'utente, ed è anche compatto e leggibile. Il prefisso "ds-" è adottato per praticità e chiarezza, ma non è obbligatorio. Se non si è sicuri di cosa utilizzare per le proprie stringhe di ID, consigliamo di utilizzare valori di UUID esadecimali (ad esempio 28b1e915-57e5-4e6c-a7bd-6fb4e729cb23), che sono immediatamente disponibili in qualsiasi ambiente di programmazione moderno.

Le stringhe di ID possono contenere lettere ASCII, numeri, trattini e trattini bassi, e possono avere una lunghezza massima di 64 caratteri. È possibile, e probabilmente utile, codificare i metadata in una stringa ID. Tuttavia, non è consigliabile farlo, perché una volta che l'oggetto è stato creato, il relativo ID non può essere modificato.

I nomi degli oggetti forniscono un modo semplice per associare metadata di facile utilizzo con ogni oggetto. È possibile aggiornare i nomi dopo che un oggetto è stato creato. In questo modo si può fare in modo che il nome dell'oggetto rifletta un dato aspetto del flusso di lavoro ML. Ad esempio, è possibile denominare inizialmente un modello ML "esperimento # 3", e successivamente rinominarlo "modello finale di produzione". I nomi possono essere costituiti da qualsiasi stringa si voglia, fino a 1.024 caratteri.

Durate degli oggetti

Qualsiasi oggetto origine dati, modello ML, valutazione o previsione in batch creato con Amazon ML sarà disponibile per l'uso per almeno due anni dopo la creazione. Amazon ML potrebbe rimuovere automaticamente gli oggetti che non sono stati utilizzati o a cui non è stato effettuato alcun accesso da più di due anni.

Risorse

Le seguenti risorse correlate possono rivelarsi utili durante l'utilizzo di questo servizio.

- [Informazioni sui prodotti Amazon ML](#) Acquisisce tutte le informazioni pertinenti su Amazon ML in una posizione centrale.
- [Domande frequenti su Amazon ML](#)— Copre le domande principali che gli sviluppatori hanno posto in merito a questo prodotto.
- [Codice di esempio Amazon ML](#)— Applicazioni di esempio che usano Amazon ML. È possibile utilizzare il codice di esempio come punto di partenza per creare le proprie applicazioni ML.
- [Informazioni di riferimento dell'API Amazon ML](#)— Descrive nel dettaglio tutte le operazioni delle API per Amazon ML. Fornisce, inoltre, esempi di richieste e risposte per i protocolli dei servizi Web supportati.
- [Centro risorse per gli sviluppatori di AWS](#) Fornisce un punto di partenza centrale per trovare documentazione, esempi di codice, note di rilascio e altre informazioni utili per creare applicazioni innovative con AWS.
- [Formazione e corsi AWS](#) Collegamenti a corsi basati su ruoli e di specializzazione nonché a corsi gestiti dall'utente per affinare le proprie competenze AWS e acquisire esperienza pratica.
- [Strumenti per sviluppatori AWS](#) Collegamenti a strumenti e risorse per sviluppatori che forniscono documentazione, esempi di codice, note di rilascio e altre informazioni utili per creare applicazioni innovative con AWS.
- [Centro AWS Support](#)— Il centro per creare e gestire i tuoi casi di AWS Support. Include inoltre link ad altre risorse utili, quali forum, domande frequenti di tipo tecnico, stato del servizio e AWS Trusted Advisor.
- [AWS Support](#) pagina Web principale che include le informazioni su AWS Support, un canale di assistenza rapida individuale che aiuta a creare ed eseguire applicazioni nel cloud.
- [Contattaci](#) Un punto di contatto centrale per domande relative a fatturazione, account, eventi, uso illecito e altri problemi.
- [Condizioni del sito AWS](#) Informazioni dettagliate sul copyright e i marchi, l'account, la licenza, l'accesso al sito e altri argomenti.

Cronologia dei documenti

La tabella seguente descrive le modifiche importanti apportate alla documentazione dall'ultima versione di Amazon Machine Learning (Amazon ML).

- Versione API: 09-04-2015
- Ultimo aggiornamento della documentazione: 2016-08-02

Modifica	Descrizione	Data della modifica
Aggiunti parametri	<p>Questa release di Amazon ML aggiunge nuove metriche per gli oggetti Amazon ML.</p> <p>Per ulteriori informazioni, consultare Elenco degli oggetti.</p>	2 agosto 2016
Eliminazione di più oggetti	<p>Questa release di Amazon ML aggiunge la possibilità di eliminare più oggetti Amazon ML.</p> <p>Per ulteriori informazioni, consultare Eliminazione di oggetti.</p>	20 luglio 2016
Aggiunta tagging	<p>Questa release di Amazon ML aggiunge la possibilità di applicare tag agli oggetti Amazon ML.</p> <p>Per ulteriori informazioni, consultare Tagging dei tuoi oggetti Amazon ML.</p>	23 giugno 2016
Copia delle origini dati di Amazon Redshift	<p>Questa release di Amazon ML aggiunge la possibilità di copiare le impostazioni delle origini dati di Amazon Redshift in una nuova origine dati di Amazon Redshift.</p> <p>Per ulteriori informazioni sulla copia delle impostazioni delle origini dati di Amazon Redshift, consulta Copia di un'origine dati (console).</p>	11 aprile 2016
Aggiunta possibilità di	Questa release di Amazon ML aggiunge la possibilità di mescolare (shuffle) i dati di input.	5 aprile 2016

Modifica	Descrizione	Data della modifica
mischiare i dati	Per ulteriori informazioni su come usare il parametro Shuffle type (Tipo di mescolamento), consultare Tipo di mescolamento dei dati di addestramento .	
Creazione di origini dati migliori con Amazon Redshift	Questa release di Amazon ML aggiunge la possibilità di testare le impostazioni di Amazon Redshift al momento della creazione di un'origine dati di Amazon ML nella console, per verificare che la connessione funzioni. Per ulteriori informazioni, consultare Creazione di un'origine dati con Amazon Redshift Data (console) .	21 marzo 2016
Migliore conversione degli schemi di dati di Amazon Redshift	Questa versione di Amazon ML migliora la conversione degli schemi di dati Amazon Redshift (Amazon Redshift) in schemi di dati Amazon ML. Per ulteriori informazioni sull'utilizzo di Amazon Redshift con Amazon ML, consulta Creazione di un'origine dati Amazon ML dai dati in Amazon Redshift .	9 Febbraio 2016
Aggiunta la registrazione di CloudTrail	Questa release di Amazon ML aggiunge la possibilità di registrare e richieste utilizzando AWS CloudTrail (CloudTrail). Per ulteriori informazioni sull'uso della registrazione di CloudTrail, consulta Registrazione delle chiamate API di Amazon ML con AWS CloudTrail .	10 dicembre 2015
Aggiunta di ulteriori opzioni DataRearrangement	Questa release di Amazon ML aggiunge la possibilità di dividere i dati di input in modo casuale e di creare origini dati complementari. Per ulteriori informazioni sull'uso della DataRearrangement parametro, vedi Riordino dei dati . Per informazioni su come utilizzare le nuove opzioni per la convalida incrociata, consultare la pagina Convalida incrociata .	3 dicembre 2015

Modifica	Descrizione	Data della modifica
Prova di previsioni in tempo reale	<p>Questa release di Amazon ML aggiunge la possibilità di provare previsioni in tempo reale nella console di servizio.</p> <p>Per ulteriori informazioni su come provare previsioni in tempo reale, consultare Richiesta di previsioni in tempo reale nella Guida per sviluppatori di Machine Learning.</p>	19 novembre 2015
Nuova regione	<p>Questa release di Amazon ML aggiunge il supporto per la regione UE (Irlanda).</p> <p>Per ulteriori informazioni su Amazon ML nell'area UE (Irlanda), consulta Regioni ed endpoint nella Guida per sviluppatori di Machine Learning.</p>	20 Agosto 2015
Versione iniziale	Questa è la prima versione della Guida per gli sviluppatori di ML.	9 aprile 2015