

Introduzione

Il *machine learning*, fondamentalmente, si occupa di algoritmi che trasformano le informazioni in conoscenza fruibile. Questo fatto rende il machine learning particolarmente adatto all'attuale epoca dei big data. Senza il machine learning, sarebbe quasi impossibile riuscire ad avere ragione di tutta quella massa di informazioni.

Data la crescente importanza di R, un ambiente di programmazione statistico, interpiattaforma e a costo zero, non vi è mai stato un momento migliore per iniziare a occuparsi di machine learning. R offre un insieme di strumenti che sono, al contempo, semplici ma potenti, e che possono assistervi nell'estrarre informazioni utili dai dati.

Utilizzando in modo combinato i casi di studio con la teoria necessaria per comprendere come funzionano le cose dietro le quinte, questo libro fornisce tutte le conoscenze necessarie per iniziare a lavorare con il machine learning.

A chi si rivolge questo libro

Questo libro si rivolge a tutti coloro che sperano di utilizzare i dati per prendere decisioni informate. Magari già conoscete qualcosa di machine learning, ma non avete mai usato R; o magari conoscete un po' R, ma non sapete nulla di machine learning. In ogni caso, questo libro ha lo scopo di colmare rapidamente queste lacune. È utile avere una certa familiarità con i concetti di base della matematica e della programmazione, ma in realtà non è richiesta alcuna precedente esperienza. Tutto ciò che occorre è un po' di curiosità.

Argomenti trattati

- Il Capitolo 1 presenta la terminologia e i concetti che definiscono e caratterizzano il machine learning, così come un metodo per passare da un compito di apprendimento all'algoritmo appropriato.
- Il Capitolo 2 fornisce l'opportunità per iniziare a “mettere le mani” sui dati con R. Il capitolo tratta le strutture di dati e le procedure essenziali impiegate per il caricamento, l'esplorazione e la conoscenza dei dati.

- Il Capitolo 3 insegna a comprendere e applicare un semplice ma potente algoritmo di machine learning a un primo compito reale: identificare dei campioni di tumore maligno.
- Il Capitolo 4 rivela i concetti essenziali della probabilità, che vengono impiegati nei più avanzati sistemi di filtraggio anti-spam. Apprenderete le basi del *text mining* nel processo di creazione di un filtro anti-spam.
- Il Capitolo 5 esplora un paio di algoritmi di apprendimento le cui predizioni non solo sono accurate, ma anche facili da motivare. Applicheremo questi metodi a quei compiti in cui la trasparenza è importante.
- Il Capitolo 6 introduce gli algoritmi di machine learning impiegati per eseguire predizioni numeriche. Poiché queste tecniche affondano a piene mani nel campo della statistica, apprenderete anche le metriche essenziali necessarie per comprendere le relazioni numeriche.
- Il Capitolo 7 tratta due complessi ma potenti algoritmi di machine learning. Anche se le operazioni matematiche possono sembrare complicate, utilizzeremo degli esempi che illustrano il loro funzionamento interno in termini più semplici.
- Il Capitolo 8 svela l'algoritmo impiegato nei sistemi di "consigli per gli acquisti" utilizzati da molti rivenditori. Se vi siete mai chiesti come fanno i rivenditori a conoscere le vostre abitudini di spesa meglio di voi stessi, questo capitolo vi rivelerà i loro segreti.
- Il Capitolo 9 è dedicato a una procedura che individua dei cluster di elementi correlati. Utilizzeremo questo algoritmo per identificare dei profili all'interno di una community online.
- Il Capitolo 10 fornisce informazioni sulla valutazione del successo di un progetto di machine learning e sul modo in cui ottenere una stima affidabile delle sue prestazioni su dati futuri.
- Il Capitolo 11 rivela i metodi impiegati dai team delle principali competizioni di machine learning. Se siete tipi competitivi o semplicemente volete trarre il massimo dai vostri dati, dovete assolutamente aggiungere queste tecniche al vostro repertorio.
- Il Capitolo 12 esplora le ultime frontiere del machine learning. Dall'utilizzo dei big data alla creazione di codice R più veloce, gli argomenti trattati vi aiuteranno a giungere fino ai limiti di ciò che è possibile fare oggi con R.

Prerequisiti

Gli esempi presenti in questo libro sono stati scritti e collaudati con R versione 3.5.2 su Microsoft Windows e macOS, ma dovrebbero funzionare con ogni versione recente di R.

Download dei file di codice degli esempi

Potete accedere ai file di codice degli esempi di questo libro visitando l'URL <http://bit.ly/apo-mlr>. Il codice sorgente è inoltre disponibile su GitHub agli URL <https://github.com/PacktPublishing/Machine-Learning-with-R-Third-Edition> e <https://github.com/dataspelunking/MLwR>.

Convenzioni utilizzate nel libro

In questo libro, abbiamo impiegato vari stili di testo per distinguere i vari tipi di informazioni fornite. Ecco alcuni esempi di questi stili e una descrizione del loro significato. Il codice sorgente nel testo, i nomi di funzioni, i nomi di file, le estensioni di file, gli input dell'utente e i nomi di package R sono riportati come segue: “La funzione `knn()` contenuta nel package `class` fornisce un'implementazione standard e classica dell'algoritmo k-NN”.

L'input e l'output R sono riportati come segue:

```
> table(mushrooms$type)
```

```
edible poisonous
 4208      3916
```

I *nuovi termini* e le *parole chiave* sono riportati in corsivo. Le parole che compaiono sullo schermo, per esempio, nei menu o nelle finestre di dialogo, appaiono sempre in corsivo: “Il link *Task Views* sul lato sinistro della pagina CRAN fornisce un elenco filtrato di package”.

Nel codice sorgente in generale, è stata rispettata la convenzione impiegata da R per i separatori: la virgola (,) come separatore delle migliaia e il punto (.) come separatore dei valori decimali: “123,456.78”.

Al contrario, nel testo descrittivo si è adottata la convenzione consueta italiana: il punto (.) come separatore delle migliaia e la virgola (,) decimale: “123.456,78”.