

Indice generale

Autore e revisore	ix
L'autore	ix
Il revisore.....	ix
Ringraziamenti	xi
Introduzione	xiii
A chi si rivolge questo libro	xiii
Argomenti trattati	xiii
Prerequisiti	xiv
Download dei file di codice degli esempi	xiv
Convenzioni utilizzate nel libro.....	xv
Capitolo 1 Introduzione al machine learning	1
Le origini del machine learning	2
Usi e abusi del machine learning.....	4
I successi del machine learning.....	4
I limiti del machine learning.....	6
L'etica del machine learning.....	7
Come apprendono le macchine	10
Archivio dati.....	11
Astrazione	12
Generalizzazione.....	14
Valutazione.....	16
Il machine learning in pratica.....	17
Tipi di dati di input	18
Tipi di algoritmi di machine learning.....	20
Tipi di dati di input e relativi algoritmi	22
Machine Learning con R.....	23
Installazione dei package R.....	24
Caricamento e download dei package R.....	25

Installazione di RStudio.....	26
Riepilogo	27
Capitolo 2 Gestione e conoscenza dei dati	29
Le strutture di dati di R	30
I vettori.....	30
I fattori.....	32
Le liste.....	33
I data frame	36
Matrici e array.....	39
Gestione dei dati con R.....	40
Salvataggio, caricamento rimozione di strutture di dati R.....	40
Importazione e salvataggio di dati da file CSV.....	41
Esplorazione e comprensione dei dati.....	43
Esplorazione della struttura dei dati.....	43
Esplorazione delle variabili numeriche	44
Esplorazione delle variabili categoriche	54
Esplorazione delle relazioni fra le variabili	57
Riepilogo	61
Capitolo 3 Apprendimento “pigro”: classificazione nearest neighbors.....	63
La classificazione nearest neighbors.....	64
L’algoritmo k-NN	64
Perché si dice che l’algoritmo k-NN è “pigro”?	72
Esempio: diagnosi del tumore al seno con l’algoritmo k-NN.....	73
Passo 1: raccolta dei dati.....	73
Passo 2: esplorazione e preparazione dei dati.....	74
Passo 3: addestramento di un modello sui dati	78
Passo 4: valutazione delle prestazioni del modello	79
Passo 5: miglioramento delle prestazioni del modello.....	81
Riepilogo	83
Capitolo 4 Apprendimento probabilistico: Naive Bayes.....	85
Che cos’è la classificazione Naive Bayes.....	86
Concetti di base dei metodi bayesiani.....	86
L’algoritmo Naive Bayes	92
Esempio: filtraggio dello spam SMS su smartphone con l’algoritmo Naive Bayes.....	98
Passo 1: raccolta dei dati.....	98
Passo 2: esplorazione e preparazione dei dati.....	99
Passo 3: addestramento di un modello sui dati	113
Passo 4: valutazione delle prestazioni del modello	114
Passo 5: miglioramento delle prestazioni del modello.....	116
Riepilogo	117

Capitolo 5	Classificazione tramite alberi decisionali e regole	119
	Che cosa sono gli alberi decisionali	120
	Divide et impera.....	121
	L'algoritmo ad albero decisionale C5.0.....	124
	Esempio: identificazione dei prestiti bancari a rischio con alberi decisionali C5.0.....	129
	Passo 1: raccolta dei dati.....	130
	Passo 2: esplorazione e preparazione dei dati.....	130
	Passo 3: addestramento di un modello sui dati	133
	Passo 4: valutazione delle prestazioni del modello	136
	Passo 5: miglioramento delle prestazioni del modello.....	137
	Che cosa sono le regole di classificazione	141
	Separate and conquer: un divide et impera con separazione	142
	L'algoritmo 1R.....	144
	L'algoritmo RIPPER.....	146
	Regole dagli alberi decisionali	148
	Che cosa rende gli alberi e le regole così "avidì"?.....	149
	Esempio: identificazione dei funghi velenosi tramite sistemi di apprendimento basato su regole.....	151
	Passo 1: raccolta dei dati.....	151
	Passo 2: esplorazione e preparazione dei dati.....	152
	Passo 3: addestramento di un modello sui dati	152
	Passo 4: valutazione delle prestazioni del modello	154
	Passo 5: miglioramento delle prestazioni del modello.....	155
	Riepilogo	157
Capitolo 6	Previsione di dati numerici: i metodi a regressione	159
	Che cos'è la regressione	160
	La regressione lineare semplice.....	162
	Stima con il metodo dei minimi quadrati	165
	Le correlazioni.....	167
	La regressione lineare multipla	168
	Esempio: predizione delle spese mediche con regressione lineare	173
	Passo 1: raccolta dei dati.....	173
	Passo 2: esplorazione e preparazione dei dati.....	174
	Passo 3: addestramento di un modello sui dati	179
	Passo 4: valutazione delle prestazioni del modello	181
	Passo 5: miglioramento delle prestazioni del modello.....	183
	Che cosa sono gli alberi di regressione e gli alberi di modelli.....	189
	Aggiunta della regressione agli alberi.....	189
	Esempio: stimare la qualità dei vini con alberi di regressione e alberi di modelli.....	192

Passo 1: raccolta dei dati.....	192
Passo 2: esplorazione e preparazione dei dati.....	193
Passo 3: addestramento di un modello sui dati	195
Passo 4: valutazione delle prestazioni del modello	198
Passo 5: miglioramento delle prestazioni del modello.....	200
Riepilogo	203

Capitolo 7 Metodi black-box: reti neurali e macchine a vettori di supporto.....205

Che cosa sono le reti neurali	206
Dai neuroni biologici ai neuroni artificiali.....	207
Le funzioni di attivazione.....	209
La topologia della rete.....	211
Addestramento di reti neurali con retropropagazione.....	215
Esempio: modellazione della forza del calciatore con le reti neurali.....	217
Passo 1: raccolta dei dati.....	217
Passo 2: esplorazione e preparazione dei dati.....	218
Passo 3: addestramento di un modello sui dati	219
Passo 4: valutazione delle prestazioni di un modello.....	222
Passo 5: miglioramento delle prestazioni del modello.....	223
Che cosa sono le macchine a vettori di supporto	227
Classificazione tramite iperpiani	228
Impiego di kernel trick per gli spazi non lineari	233
Esempio: OCR con macchine a vettori di supporto	235
Passo 1: raccolta dei dati.....	236
Passo 2: esplorazione e preparazione dei dati.....	237
Passo 3: addestramento di un modello sui dati	238
Passo 4: valutazione delle prestazioni del modello	240
Passo 5: miglioramento delle prestazioni del modello.....	241
Riepilogo	244

Capitolo 8 Ricerca di pattern: analisi del carrello della spesa tramite regole associative.....245

Che cosa sono le regole associative.....	246
L'algoritmo Apriori per l'apprendimento di regole associative.....	247
Misurazione della significatività delle regole: supporto e confidenza.....	248
Costruzione di un insieme di regole con il principio Apriori..	250
Esempio: identificazione degli articoli acquistati di frequente tramite regole associative	251
Passo 1: raccolta dei dati.....	251

	Passo 2: esplorazione e preparazione dei dati.....	252
	Passo 3: addestramento di un modello sui dati	259
	Passo 4: valutazione delle prestazioni del modello	261
	Passo 5: miglioramento delle prestazioni del modello.....	264
	Riepilogo	266
Capitolo 9	Ricerca di gruppi di dati: clustering con k-means ...	269
	Che cos'è il clustering.....	270
	Il clustering come un compito di machine learning.....	270
	L'algoritmo di clustering k-means	272
	Ricerca dei segmenti di mercato dei teenager tramite il clustering k-means	279
	Passo 1: raccolta dei dati	280
	Passo 2: esplorazione e preparazione dei dati.....	280
	Passo 3: addestramento di un modello sui dati	284
	Passo 4: valutazione delle prestazioni di un modello.....	287
	Passo 5: miglioramento delle prestazioni del modello.....	290
	Riepilogo	291
Capitolo 10	Valutazione delle prestazioni di un modello	293
	Misurazione delle prestazioni in compiti di classificazione.....	294
	Comprendere le predizioni di un classificatore	294
	Un'occhiata più da vicino alle matrici di confusione	298
	Impiego delle matrici di confusione per misurare le prestazioni.....	300
	Oltre l'accuratezza: altre misure prestazionali	302
	Visualizzazione dei compromessi prestazionali con le curve ROC	310
	Stima delle prestazioni future	315
	Il metodo holdout	315
	Riepilogo	322
Capitolo 11	Miglioramento delle prestazioni del modello	323
	Ottimizzazione dei modelli base per migliorarne le prestazioni....	324
	Impiego di caret per l'ottimizzazione automatizzata dei parametri	325
	Miglioramento delle prestazioni del modello con il meta-learning.....	334
	Che cosa sono gli ensemble	334
	Il bagging	337
	Il boosting	338
	Le foreste casuali.....	341
	Riepilogo	347

Capitolo 12 Argomenti specialistici di machine learning349

Gestione e preparazione dei dati tratti dal mondo reale.....	350
Preparazione dei dati con i package del tidyverse.....	350
Lettura e scrittura su file	353
Query di dati su database SQL.....	355
Utilizzo di dati e servizi online.....	359
Download del testo di interesse pagine web	360
Parsing dei dati presenti nelle pagine web.....	362
Impiego di dati specifici di un dominio	368
Analisi di dati bioinformatici.....	368
Analisi e visualizzazione dei dati di rete	369
Miglioramento delle prestazioni di R	372
Gestione di grandi dataset.....	373
Apprendimento rapido con il calcolo parallelo	376
Algoritmi di apprendimento ottimizzati	384
Impiego della GPU	387
Riepilogo	389

Indice analitico.....391