

Indice generale

Introduzionexi
Argomenti trattati in questo libro xi
Dotazione software necessaria xii
A chi è rivolto questo libro xii
Convenzioni utilizzate xiii
Scarica i file degli esempi xiii
Capitolo 1	Essere uno scienziato dei dati 1
Che cosa si intende per scienza dei dati? 3
Terminologia di base 3
Perché scienza dei dati? 4
Il diagramma di Venn e la scienza dei dati 5
Matematica 7
Programmazione 8
Perché Python? 9
Conoscenza del dominio13
Ancora un po' di terminologia13
Scienza dei dati: casi di studio15
Caso di studio: automatizzare il flusso dei documenti15
Caso di studio: spese di marketing16
Caso di studio: i contenuti di un'offerta di lavoro18
Riepilogo20
Capitolo 2	Tipi di dati23
L'aspetto dei dati23
Perché tutte queste distinzioni?24
Dati strutturati e non strutturati24
Un esempio di pre-elaborazione dei dati25
Conteggio di parole e frasi26
Presenza di specifici caratteri speciali26
Lunghezza relativa del testo26

	Individuazione degli argomenti.....	27
	Dati quantitativi e qualitativi.....	27
	Esempio: caratteristiche delle caffetterie.....	28
	Esempio: dati sul consumo di alcol a livello mondiale.....	29
	Approfondimenti.....	31
	Ricapitolando.....	31
	I quattro livelli dei dati.....	31
	Il livello nominale.....	32
	Il livello ordinale.....	33
	Il livello degli intervalli.....	35
	Il livello dei rapporti.....	38
	Dati e presupposti.....	39
	Riepilogo.....	40
Capitolo 3	I cinque passi della scienza dei dati.....	41
	Introduzione alla scienza dei dati.....	41
	Panoramica sui cinque passi.....	42
	Porre una domanda interessante.....	42
	Ottenere i dati.....	42
	Esplorare i dati.....	42
	Creare un modello per i dati.....	43
	Comunicare e presentare i risultati.....	43
	Esplorare i dati.....	43
	Domande di base per l'esplorazione dei dati.....	43
	Dataset 1: Yelp.....	44
	Dataset 2: Titanic.....	52
	Riepilogo.....	55
Capitolo 4	Basi matematiche.....	57
	La matematica come disciplina.....	58
	Simboli e terminologia di base.....	58
	Vettori e matrici.....	58
	Simboli aritmetici.....	61
	Grafici.....	64
	Logaritmi/esponenti.....	65
	Teoria degli insiemi.....	67
	Algebra lineare.....	71
	Prodotti di matrici.....	71
	Riepilogo.....	75
Capitolo 5	Impossibile o improbabile: introduzione al calcolo delle probabilità.....	77
	Definizioni di base.....	77
	Probabilità.....	78
	Approccio bayesiano vs. approccio a frequenza.....	79
	Approccio a frequenza.....	80

Eventi composti.....	82
Probabilità condizionale.....	85
Le regole del calcolo delle probabilità.....	86
La regola della somma.....	86
Reciproca esclusività.....	87
La regola del prodotto.....	87
Indipendenza.....	88
Eventi complementari.....	88
Approfondimento.....	89
Riepilogo.....	91
Capitolo 6	Approfondimenti sul calcolo delle probabilità.....93
Eventi collettivamente esaustivi.....	94
Ancora sulle idee di Bayes.....	94
Il teorema di Bayes.....	94
Altre applicazioni del teorema di Bayes.....	97
Variabili casuali.....	100
Variabili casuali discrete.....	101
Riepilogo.....	114
Capitolo 7	Basi di statistica.....115
Che cos'è la statistica?.....	115
Come si ottengono e campionano i dati?.....	116
Ottenere i dati.....	117
Campionamento dei dati.....	119
Campionamento a probabilità.....	119
Campionamento casuale.....	119
Campionamento di probabilità diseguali.....	120
Misurazioni statistiche.....	121
Misurazioni del centro.....	121
Misurazioni della variabilità.....	122
Misurazioni della posizione relativa.....	126
La regola empirica.....	134
Riepilogo.....	135
Capitolo 8	Approfondimenti di statistica.....137
Stime dei punti.....	137
Distribuzioni di campionamento.....	142
Intervalli di confidenza.....	144
Verifica delle ipotesi.....	147
Condurre un verifica delle ipotesi.....	148
Test t per un campione.....	148
Errori di Tipo I e di Tipo II.....	152
Verifica delle ipotesi per variabili categoriche.....	152
Riepilogo.....	156

Capitolo 9	Comunicare i dati	157
	Perché è importante la comunicazione?.....	157
	Identificare i metodi di presentazione efficaci e inefficaci.....	158
	Grafici a dispersione	158
	Grafici a linee	160
	Diagrammi a barre.....	161
	Istogrammi.....	163
	Grafici box-plot (a scatola e baffi).....	164
	Quando i grafici e le statistiche mentono	167
	Correlazione oppure causalità?.....	167
	Il paradosso di Simpson.....	169
	Se la correlazione non implica causalità, allora qual è il suo significato?	171
	Comunicazione verbale.....	171
	Si tratta di raccontare una storia	171
	Un approccio più formale.....	171
	La strategia di presentazione: perché, come e cosa	172
	Riepilogo	173
Capitolo 10	Quando le macchine apprendono: il machine learning	175
	Che cosa si intende con machine learning?	176
	Esempio: riconoscimento facciale.....	176
	Il machine learning non è perfetto	177
	Come funziona il machine learning?.....	178
	Tipi di machine learning.....	178
	Apprendimento con supervisione.....	179
	Apprendimento senza supervisione	184
	In tutto questo, che cosa c'entra la modellazione statistica?.....	189
	La regressione lineare	189
	Aggiunta di altri predittori.....	194
	Metriche per la regressione	196
	Regressione logistica.....	201
	Probabilità, odds e log-odds.....	203
	La matematica della regressione logistica	205
	Variabili fittizie	208
	Riepilogo	212
Capitolo 11	Le previsioni non crescono sugli alberi... 0 forse sì?	215
	Classificazione bayesiana naïve.....	215
	Alberi decisionali	222
	In quale modo un computer costruisce un albero di regressione?	224
	In quale modo un computer adatta un albero di classificazione?	224

Apprendimento senza supervisione.....	229
Quando usare l'apprendimento senza supervisione	229
Clustering K-means.....	229
Esempio: punti dei dati	231
Esempio: birre!	237
Scelta di un valore ottimale per K e convalida dei cluster.....	238
Il coefficiente di silhouette.....	239
Estrazione delle caratteristiche e analisi del componente principale.....	241
Riepilogo	250
Capitolo 12	
Oltre le basi della scienza dei dati.....	253
Il compromesso tra bias e varianza.....	253
Errori dovuti al bias	254
Errori dovuti alla varianza	254
Due casi estremi di compromesso tra bias e varianza.....	261
Il ruolo del bias e della varianza nelle funzioni d'errore	261
Convalida incrociata k-fold	263
Ricerca a griglia.....	266
Confrontare graficamente l'errore di addestramento e di convalida incrociata.....	269
Tecniche d'insieme	271
Foreste casuali.....	272
Confronto fra foreste casuali e alberi decisionali	277
Reti neurali	277
Struttura di base.....	277
Riepilogo	283
Capitolo 13	
Casi di studio.....	285
Caso di studio 1: predire le quotazioni di mercato sulla base dei social media	285
Analisi del sentiment nei testi.....	285
Analisi esplorativa dei dati	286
Ulteriori sviluppi.....	298
Caso di studio 2: perché alcune persone mentono sul loro matrimonio?.....	299
Caso di studio 3: uso di tensorflow	305
Tensorflow e le reti neurali.....	308
Riepilogo	314
Indice analitico	315