

Indice generale

Ringraziamenti	xiii
Introduzione	xv
Informazioni su questo libro	xv
A chi si rivolge questo libro	xvi
Questioni di software	xvi
Note sugli apici	xvii
Il forum del libro	xvii
Esercitazioni	xviii
Capitolo 1	Che cosa si intende per scienza dei dati?..... 1
Unità 1	La tipica sequenza di analisi dei dati..... 3
Unità 2	Sequenza di acquisizione dei dati..... 5
Unità 3	Struttura dei report..... 7
Esercitazioni 8
Hello, World! (☆) 8
Capitolo 2	Elementi di base di Python per la scienza dei dati 9
Unità 4	Le funzioni che operano sulle stringhe..... 11
Unità 5	La struttura dati appropriata..... 14
Unità 6	Definizione e manipolazione delle liste..... 16
Unità 7	I contatori..... 17
Unità 8	Manipolazione dei file..... 18
Unità 9	Andare sul Web..... 19
Unità 10	Ricerche a pattern tramite espressioni regolari 21
Il linguaggio delle espressioni regolari 21
Ricerca, suddivisione e sostituzione con il modulo re 23
Unità 11	Globbering di nomi di file e altre stringhe 26
Unità 12	Pickling dei dati 27

	Esercitazioni	28
	Contatore della frequenza delle parole (☆)	28
	Indicizzatore di file (☆☆).....	28
	Estrattore di numeri telefonici (☆☆☆)	28
Capitolo 3	Elaborare i dati testuali	29
	Unità 13 Elaborazione di file HTML	31
	Unità 14 Manipolazione dei file CSV.....	35
	Unità 15 Leggere i file JSON.....	37
	Unità 16 Elaborazione di testi in linguaggio naturale	39
	Corpus NLTK.....	39
	Normalizzazione	41
	Altre procedure di elaborazione del testo.....	43
	Esercitazioni	45
	Rilevatore di collegamenti ipertestuali non funzionanti (☆).....	45
	Wikipedia Miner (☆☆).....	45
	Classificatore del genere musicale (☆☆☆)	45
Capitolo 4	Utilizzare i database	47
	Unità 17 Configurazione di un database MySQL	49
	Unità 18 Uso di un database MySQL: riga di comando	52
	Inserimento	52
	Cancellazione	53
	Modifica.....	53
	Selezione.....	53
	Unione.....	55
	Unità 19 Uso di un database MySQL: pymysql.....	56
	Unità 20 Gestire gli archivi di documenti: MongoDB.....	58
	Esercitazioni	62
	Indicizzatore di file MySQL (☆)	62
	Convertitore MySQL – MongoDB (☆☆)	62
Capitolo 5	Usare i dati numerici tabulari	63
	Unità 21 Creazione di array	65
	Unità 22 Trasposizione e alterazione.....	68
	Unità 23 Indicizzazione e slicing.....	70
	Unità 24 Operazioni vettoriali	72
	Unità 25 Le funzioni universali	74
	Unità 26 Le funzioni condizionali.....	76
	Unità 27 Aggregazione e ordinamento degli array	77
	Unità 28 Trattare gli array come se fossero insieme.....	79
	Unità 29 Salvataggio e lettura di array	80
	Unità 30 Generare un'onda sinusoidale di sintesi	81

Esercitazioni	83
Sottrazioni fra array (☆)	83
HEI Locator (☆☆)	83
Calcolo delle analogie negli Stati USA (☆☆☆)	83
Capitolo 6	
 Manipolare serie di dati e frame	85
Unità 31 Le strutture di dati in Pandas	87
Le serie	87
I frame	91
Unità 32 Cambiare aspetto ai dati	93
Indicizzazione	93
Reindicizzazione	94
Indicizzazione gerarchica	95
Stack e pivot	96
Unità 33 Gestione dei dati mancanti	99
Cancellazione dei dati mancanti	99
Ricostruzione dei dati mancanti	100
Sostituzione dei valori	101
Unità 34 Combinare i dati	102
Unione	102
Concatenamento	104
Cancellazione dei duplicati	105
Unità 35 Ordinamento e descrizione dei dati	106
Ordinamento e classificazione	106
Statistica descrittiva	107
Unicità, conteggio, appartenenza	108
Unità 36 Trasformazione dei dati	110
Operazioni aritmetiche	110
Aggregazione dei dati	111
Discretizzazione	113
Mappaggio	114
Tabulazione incrociata	115
Unità 37 I/O su file in Pandas	116
Lettura di file CSV e tabulari	116
Chunking	117
Lettura di altri tipi di file	118
Esercitazioni	119
Trappole per linci (☆)	119
PIL vs. Consumo di alcol (☆☆)	119
Meteo vs. Consumo di alcol (☆☆☆)	119
Capitolo 7	
 Utilizzo dei dati delle reti	121
Unità 38 Parliamo di grafi	123
Elementi, tipi e densità dei grafi	123
Struttura di un grafo	123
Centralità	125

Unità 39 Sequenza di analisi di una rete	126
Unità 40 Parliamo di networkx	127
Costruzione e manipolazione di una rete	127
Esplorazione e analisi di una rete.....	129
Gestione degli attributi	132
Cricche e struttura comunitaria	132
Input e output	133
Esercitazioni	134
Centralità e correlazioni (☆)	134
Opere di Shakespeare (☆☆)	134
Reti e confini (☆☆)	134

Capitolo 8 Rappresentazione grafica135

Unità 41 Tracciamento di grafici con PyPlot	137
Unità 42 Altri tipi di grafici.....	140
Unità 43 Decorazione dei grafici.....	141
Unità 44 Grafici e Pandas.....	144
Esercitazioni	147
American Pie (☆)	147
Popolazione della California (☆☆).....	147

Capitolo 9 Probabilità e statistica.....149

Unità 45 Le distribuzioni delle probabilità.....	151
Distribuzione uniforme	152
Distribuzione normale	152
Distribuzione binomiale	152
Unità 46 Raccolta delle misurazioni statistiche	153
Unità 47 Statistica in Python.....	155
Generazione di numeri casuali	155
Calcolo delle misurazioni statistiche	156
Esercitazioni	159
S&P 500 nel Ventunesimo secolo (☆).....	159
Nutrienti in Rete (☆☆☆).....	159

Capitolo 10 Machine Learning.....161

Unità 48 Progettazione di un esperimento predittivo.....	163
Unità 49 La regressione lineare.....	165
Regressione OLS (metodo dei minimi quadrati)	165
Regressione ridge (Regolarizzazione di Tichonov).....	167
Regressione logistica (Logit)	168
Unità 50 Raggruppamento dei dati con il clustering K-Means ...	170
Unità 51 Sopravvivere nelle foreste decisionali casuali.....	173
Esercitazioni	175
MOSN Clusters (☆)	175
S&P 500 (☆☆)	175
Predittore di metrò (☆☆☆).....	175

Appendice A	Ulteriori letture.....	177
Appendice B	Soluzioni dei progetti di tipo (☆).....	179
	Capitolo 1 Hello, World!	179
	Capitolo 2 Contatore della frequenza delle parole	179
	Capitolo 3 Rilevatore di collegamenti ipertestuali non funzionanti	180
	Capitolo 4 Indicizzatore di file MySQL	181
	Capitolo 5 Sottrazioni fra array	182
	Capitolo 6 Trappole per linci.....	183
	Capitolo 7 Centralità e correlazioni	184
	Capitolo 8 American Pie.....	185
	Capitolo 9 S&P 500 nel Ventunesimo secolo	186
	Capitolo 10 Cluster MOSN	187
Bibliografia	189
Indice analitico	191