

Indice

Introduzione	1
I Gli strumenti del Data Scientist	9
1 Nuovo lavoro, nuovo linguaggio	11
1.1 Installazione	13
1.2 Le basi di Python	14
1.3 Stringhe, Tuple e Liste	17
1.3.1 Stringhe	17
1.3.2 Stringhe usate come template	19
1.3.3 Tuple	19
1.3.4 Liste	20
1.4 Dizionari e Set	21
1.4.1 Dizionari	21
1.4.2 Set e Frozenset	24
1.5 Controllo del Flusso di Esecuzione	26
1.6 Composizione e Spacchettamento	28
1.7 List, Set e Dict Comprehension	29
1.8 Funzioni e Moduli	32
1.9 Generatori	34
1.10 Funzioni annidate, Closures e Decoratori	35
1.11 Classi e oggetti	37
1.12 File e Serializzazione di oggetti Python	40
1.12.1 File di testo	40
1.12.2 Serializzazione di oggetti Python	41
1.13 Errori, e come gestirli	42
1.13.1 Gestione delle Eccezioni	42
1.13.2 Gerarchia	44
1.13.3 Sollevare Eccezioni	45
1.14 Conclusioni	46
2 Gestione dei Dati: iPython, Numpy e Pandas	47
2.1 iPython e Jupiter	47
2.2 La Shell	48
2.2.1 I Notebook	53

2.3	Numpy	57
2.3.1	Attributi degli Array	58
2.3.2	Creazione veloce di nuovi array	59
2.3.3	Operazioni comuni con le matrici	60
2.3.4	Indicizzazione e iterazione degli array	61
2.4	Pandas	64
2.4.1	Matrici di dati: date il benvenuto ai Dataframe	65
3	Rappresentazione grafica dei dati	83
3.1	Matplotlib	83
3.2	Import e considerazioni generali	84
3.3	Curve	84
3.4	Grafici a Dispersione	87
3.5	Grafici a barre	90
3.6	Time Series	94
3.7	Seaborn	96
3.8	Conclusioni	99
II	Apprendimento Automatico Supervisionato	101
4	La migliore offerta: Regressione Multivariata	103
4.1	Pesi e Misure	105
4.2	L'obiettivo è ridurre i costi	107
5	Prestiti e alberi decisionali	115
5.1	Pulizia dei dati	116
5.2	Analisi preliminare	118
5.3	Alberi e Foreste	123
5.3.1	Scikit-learn	125
5.4	Creazione dell'albero	125
5.5	Misura della qualità di un classificatore	129
5.6	Convalida Incrociata	131
5.7	Bias e Varianza	133
5.8	L'unione fa la forza	134
6	Analisi del sentimento:	
	Il teorema di Bayes	137
6.1	Il lancio di ArBot	137
6.2	Le aspettative del Business	138
6.3	Twitter Search API	139
6.4	Una valigia di parole	141
6.5	Un classificatore Bayesiano	143
6.5.1	La probabilità condizionata	143
6.5.2	Il Teorema di Bayes	144
6.5.3	Come funziona un classificatore	144
6.6	Dalla teoria alla pratica	147

7	Chatbot su Messenger con le Reti Neurali	153
7.1	L'era dei roBOT	153
7.2	Si torna a casa!	154
7.3	Come funziona un Bot	154
7.4	Costruiamo un piccolo dataset	157
7.5	Creazione del database	158
7.6	Come ricavare il significato di una frase	160
7.7	Reti Neurali Artificiali	163
7.8	Tensorflow	168
7.9	Da un testo a un numero	170
7.10	Il cervello del bot	172
7.11	Aggiunta del contesto	175
7.12	Deploy di un'applicazione web con Heroku	180
7.12.1	Heroku	180
7.12.2	Preparazione dell'ambiente e scrittura dei file di corredo	183
7.12.3	Scrittura dell'app con Flask	184
7.13	Facebook time!	189
7.14	Ultimi ritocchi e compiti per casa	194
III	Apprendimento Automatico non Supervisionato	195
8	Consigliami un libro!	
	Le regole di Associazione	197
8.1	Libri e lettori	197
8.2	thedotcompany	198
8.3	Misurare le associazioni	198
8.3.1	Supporto	199
8.3.2	Confidenza	199
8.3.3	Lift	200
8.4	Il sentiero più breve	200
8.5	È tempo di mettersi al lavoro	201
8.5.1	Implementazione dell'algoritmo apriori	202
8.6	Regole di Associazione	205
9	Raggruppare oggetti: gli algoritmi di clusterizzazione	209
9.1	K-means	209
9.2	Scelta del numero di cluster	213
9.2.1	Il metodo Elbow	213
9.2.2	Qualità e grafici a silhouettes	217
9.3	Clustering basato sulla densità: DBSCAN	221