

# Indice generale

<b>Prefazione</b> .....	<b>xv</b>
<b>Premessa</b> .....	<b>xvii</b>
<b>Introduzione</b> .....	<b>xix</b>
Come è organizzato il libro: una roadmap .....	xix
Tipi di capitoli.....	xix
Percorsi consigliati .....	xix
Appendici.....	xx
Requisiti e obiettivi di apprendimento .....	xx
Altre risorse .....	xx
Scriveremo del codice.....	xxi
Il codice nel testo.....	xxi
L'autore .....	xxii
Ringraziamenti.....	xxii
<b>Capitolo 1</b>	<b>Il machine learning: buon senso applicato</b>
	<b>da un computer..... 1</b>
Sono felice di unirmi a voi in questo viaggio.....	2
Il machine learning è ovunque.....	2
Occorre essere esperti di matematica e programmazione	
per studiare il machine learning? .....	2
Formule e codice sono divertenti, se li considerate	
un linguaggio .....	3
Cos'è quindi il machine learning? .....	4
Che cos'è l'intelligenza artificiale?.....	4
Che cos'è il machine learning?.....	4
Che cos'è il deep learning? .....	6
Convincere le macchine a prendere decisioni con i dati:	
la sequenza ricordare-formulare-prevedere .....	7

Come pensano gli esseri umani? .....	7
Termini tipici del machine learning: modelli e algoritmi.....	7
Esempi di modelli utilizzati dagli esseri umani.....	8
Esempi di modelli utilizzati dalle macchine .....	13
Riepilogo .....	14

## **Capitolo 2 I vari tipi di machine learning ..... 15**

Differenza tra dati etichettati e non etichettati .....	17
Che cosa sono i dati? .....	17
Che cosa sono le caratteristiche? .....	17
Etichette.....	17
Previsioni .....	17
Dati etichettati e non etichettati.....	18
Apprendimento con supervisione: il machine learning su dati etichettati.....	19
I modelli di regressione prevedono numeri .....	21
I modelli di classificazione prevedono uno stato .....	21
Apprendimento senza supervisione: il machine learning su dati non etichettati.....	22
Gli algoritmi di clustering dividono un dataset in gruppi simili .....	24
Ridurre la dimensionalità: semplificare i dati senza sacrificare troppe informazioni .....	26
Altri modi per semplificare i dati: fattorizzazione della matrice e decomposizione ai valori singolari .....	29
Machine learning generativo.....	30
Che cos'è l'apprendimento per rinforzo?.....	31
Riepilogo .....	33
Esercizi .....	34
Esercizio 2.1 .....	34
Esercizio 2.2 .....	34
Esercizio 2.3 .....	34

## **Capitolo 3 Trovare una linea più vicina ai punti con la regressione lineare ..... 35**

Il problema: prevedere il prezzo di una casa .....	37
La soluzione: costruire un modello di regressione per i prezzi delle case .....	37
“Ricordare”: considerare i prezzi delle altre case.....	39
“Formulare”: individuare una regola che stimi il prezzo della casa .....	40
“Prevedere”: che cosa facciamo quando sul mercato arriva una nuova casa?.....	41
La regressione lineare multivariata .....	42
Alcune domande e risposte rapide.....	43
Tracciare questa linea con l'algoritmo di regressione lineare.....	43

Corso accelerato su pendenza e intercettazione- $y$ .....	46
Un semplice trucco per avvicinare una linea a un insieme di punti, un punto alla volta .....	48
Il trucco del quadrato: un modo molto più intelligente per avvicinare la linea a uno dei punti .....	51
Il trucco assoluto: un altro trucco utile per avvicinare la linea ai punti.....	53
L'algoritmo di regressione lineare: ripetere il trucco assoluto o del quadrato per spostare la linea più vicino ai punti.....	54
Caricare e tracciare i dati .....	55
Utilizzare l'algoritmo di regressione lineare nel dataset .....	56
Utilizzare il modello per fare previsioni.....	57
L'algoritmo generale di regressione lineare (opzionale) .....	58
Valutare i risultati. La funzione d'errore .....	58
L'errore assoluto: una metrica che dice quanto è efficace il modello sommando le distanze .....	59
L'errore quadratico: una metrica che dice quanto è efficace il modello sommando i quadrati delle distanze .....	60
L'errore assoluto medio e (la radice de) l'errore quadratico medio sono molto più comuni.....	60
Discesa del gradiente: minimizzare una funzione d'errore scendendo lentamente da una montagna .....	61
Tracciare la funzione d'errore per sapere quando interrompere l'esecuzione dell'algoritmo.....	64
Addestrare considerando un punto o più punti alla volta con la discesa stocastica del gradiente e batch .....	66
Applicazione reale: utilizzo di Turi Create per prevedere i prezzi delle case in India .....	66
La regressione polinomiale .....	68
Un tipo speciale di funzioni per curve: i polinomi.....	68
Approssimiamo dati non lineari con una curva polinomiale.....	69
Parametri e iperparametri .....	70
Applicazioni della regressione.....	71
Sistemi di raccomandazione .....	71
Consigli su video e musica.....	71
Consigli sui prodotti .....	71
Assistenza sanitaria .....	72
Riepilogo .....	72
Esercizi .....	72
Esercizio 3.1 .....	72
Esercizio 3.2.....	73
Esercizio 3.3 .....	73
Esercizio 3.4.....	73

## Capitolo 4 **Ottimizzare l'addestramento: underfitting, overfitting, test e regolarizzazione .....75**

Un esempio di underfitting e overfitting utilizzando

la regressione polinomiale .....	78
Un test per far sì che il computer scelga il modello giusto .....	79
Scegliere il set di test e le sue dimensioni .....	81
Non possiamo utilizzare i dati di test per addestrare il modello. ....	82
Dove abbiamo infranto la regola d'oro e come possiamo rimediare? Il set di convalida .....	82
Un modo numerico per decidere quanto dovrebbe essere complesso il modello: il grafico della complessità del modello .....	83
Un'alternativa per evitare l'overfitting: la regolarizzazione .....	84
Un altro esempio di overfitting: consigli sui film .....	86
Misurare la complessità di un modello: norma L1 e norma L2 .....	87
Modificare la funzione d'errore per risolvere il problema: regressione Lasso e regressione a cresta .....	89
Regolare le quantità delle prestazioni e della complessità nel modello: il parametro di regolarizzazione .....	89
Effetti della regolarizzazione L1 e L2 nei coefficienti del modello .....	90
Un modo intuitivo per immaginare la regolarizzazione .....	91
Regressione polinomiale, collaudo e regolarizzazione con Turi Create .....	94
Riepilogo .....	98
Esercizi .....	99
Esercizio 4.1 .....	99
Esercizio 4.2 .....	99

## **Capitolo 5 Trovare linee per suddividere i punti con l'algoritmo perceptron .....101**

Il problema: siamo su un pianeta alieno e non conosciamo la lingua usata .....	104
Un pianeta leggermente più complicato .....	107
Non necessariamente il classificatore deve essere sempre corretto .....	111
Un classificatore più generale e un modo leggermente diverso di definire le linee .....	111
La funzione step e le funzioni di attivazione: un modo condensato per ottenere previsioni .....	113
Che cosa succede se ho più di due parole? Definizione generale del classificatore perceptron .....	114
Il bias, l'intercettazione- $y$ e l'umore innato di un alieno tranquillo .....	117
Funzione d'errore: come determinare se un classificatore è efficace .....	118
Come confrontare i classificatori .....	118
Trovare un buon classificatore. L'algoritmo perceptron .....	126

Migliorare leggermente il perceptron: il trucco perceptron .....	126
Ripetere più volte il trucco perceptron: l'algoritmo perceptron .....	131
Discesa del gradiente .....	133
Discesa stocastica del gradiente e batch .....	134
Programmare l'algoritmo perceptron.....	134
Programmare il trucco perceptron.....	136
Programmare l'algoritmo perceptron.....	137
Programmare l'algoritmo perceptron con Turi Create.....	139
Applicazioni dell'algoritmo perceptron.....	140
Filtri e-mail anti-spam .....	140
Sistemi di raccomandazione .....	140
Assistenza sanitaria .....	141
Visione computerizzata.....	141
Riepilogo .....	141
Esercizi .....	142
Esercizio 5.1 .....	142
Esercizio 5.2 .....	142
Esercizio 5.3 .....	143

## **Capitolo 6**    **Approccio continuo alla suddivisione dei punti con i classificatori logistici .....145**

Classificatori logistici: una versione continua dei classificatori perceptron .....	147
Approccio probabilistico alla classificazione:	
la funzione sigmoidea .....	148
Dataset e previsioni.....	150
Funzioni d'errore: assoluto, quadratico e logaritmico .....	151
Confrontare i classificatori utilizzando la log loss .....	156
L'algoritmo di regressione logistica .....	157
Migliorare leggermente il perceptron continuo:	
il trucco logistico.....	158
Ripetere il trucco logistico: l'algoritmo di regressione logistica .....	161
Discesa stocastica del gradiente, mini-batch e batch .....	163
Programmare l'algoritmo di regressione logistica .....	163
Programmare manualmente l'algoritmo di regressione logistica .....	164
Applicazione nella vita reale: classificazione delle recensioni su IMDB con Turi Create .....	167
Classificazione in più classi: la funzione softmax.....	170
Riepilogo .....	171
Esercizi.....	172
Esercizio 6.1 .....	172
Esercizio 6.2 .....	172
Esercizio 6.3 .....	172

<b>Capitolo 7</b>	<b>Valutare i modelli di classificazione con l'accuratezza e altre metriche .....</b>	<b>173</b>
	Accuratezza: quanto è corretto il mio modello? .....	174
	Due esempi di modelli: Coronavirus ed e-mail di spam .....	174
	Un modello super efficace ma del tutto inutile .....	175
	Risolvere il problema dell'accuratezza definendo e misurando i diversi tipi di errori .....	175
	Falsi positivi e falsi negativi: che cosa è peggio? .....	176
	Memorizzare i punti classificati correttamente ed erroneamente in una tabella: la matrice di confusione .....	178
	Recall: tra gli esempi positivi, quanti ne abbiamo classificati correttamente? .....	179
	Precisione: tra gli esempi classificati positivi, quanti lo sono? ...	180
	Combinare insieme recall e precisione come un modo per ottimizzare entrambi: F-score .....	182
	Calcolo di $F_{\beta}$ .....	183
	Recall, precisione o F-score: quale usare? .....	184
	Uno strumento utile per valutare il modello: la curva ROC .....	185
	Sensibilità e specificità: due nuovi modi per valutare il modello .....	185
	La curva ROC per ottimizzare la sensibilità e la specificità in un modello .....	187
	Una metrica che dice quanto è efficace il modello: l'AUC .....	189
	Come prendere decisioni utilizzando la curva ROC .....	190
	Il recall equivale alla sensibilità, ma la precisione e la specificità sono cose diverse .....	193
	Riepilogo .....	195
	Esercizi .....	197
	Esercizio 7.1 .....	197
	Esercizio 7.2 .....	197
	Esercizio 7.3 .....	198
	Esercizio 7.4 .....	198
<b>Capitolo 8</b>	<b>Applicare le probabilità con il modello naive Bayes .....</b>	<b>199</b>
	Malato o sano? Il teorema di Bayes come un eroe .....	200
	A priori, evento e a posteriori .....	204
	Caso d'uso: il modello di rilevamento dello spam .....	205
	Trovare l'a priori: probabilità generale che un'e-mail sia di spam .....	205
	Trovare l'a posteriori: probabilità che un'e-mail contenente una certa parola sia di spam .....	206
	La matematica dietro tutto questo: trasformare i rapporti in probabilità .....	208
	E con due parole? L'algoritmo naive Bayes .....	212
	E con più di due parole? .....	217

Costruire un modello di rilevamento dello spam con dati reali ....	218
Pre-elaborazione dei dati .....	219
Trovare le probabilità a posteriori con il teorema di Bayes .....	220
Implementare l'algoritmo naive Bayes .....	221
Ulteriore lavoro .....	222
Riepilogo .....	222
Esercizi .....	222
Esercizio 8.1 .....	222
Esercizio 8.2 .....	223
Esercizio 8.3 .....	223

## **Capitolo 9   Suddividere i dati ponendo domande con gli alberi decisionali.....225**

Il problema: consigliare delle app agli utenti, in base a ciò che scaricheranno con maggiore probabilità .....	232
La soluzione: costruire un sistema di suggerimenti delle app .....	233
Primo passo: porre la domanda migliore.....	233
Secondo passo: iterazione .....	244
Ultimo passo: quando interrompere la costruzione dell'albero e altri iperparametri .....	246
L'algoritmo dell'albero decisionale: costruire un albero decisionale e fare previsioni.....	247
Al di là di semplici domande sì/no .....	248
Suddividere i dati utilizzando caratteristiche categoriali non binarie .....	249
Suddividere i dati utilizzando caratteristiche continue .....	249
Il confine grafico degli alberi decisionali.....	252
Utilizzare Scikit-Learn per costruire un albero decisionale.....	254
Applicazione nella vita reale: modellazione delle ammissioni degli studenti con Scikit-Learn .....	256
Impostare gli iperparametri in Scikit-Learn .....	257
Alberi decisionali per la regressione.....	259
Applicazioni .....	263
Gli alberi decisionali in ambito sanitario .....	263
Gli alberi decisionali nei sistemi di raccomandazione.....	263
Riepilogo .....	264
Esercizi .....	265
Esercizio 9.1 .....	265
Esercizio 9.2 .....	265
Esercizio 9.3 .....	266

## **Capitolo 10   Combinare più elementi con le reti neurali .....267**

Reti neurali in un esempio: un pianeta alieno più complicato.....	269
Soluzione: se una linea non basta, usiamo due linee per classificare il dataset.....	270

Classificatore di analisi del sentiment.....	270
Classificatore per l'analisi del sentiment.....	271
Perché due linee? La felicità non è lineare?.....	272
Combinare gli output dei perceptron in un altro perceptron.....	273
Una rappresentazione grafica dei perceptron.....	275
Una rappresentazione grafica delle reti neurali.....	277
Il confine di una rete neurale.....	279
L'architettura generale di una rete neurale pienamente connessa.....	281
Addestrare le reti neurali.....	282
Funzione d'errore: un modo per misurare le prestazioni della rete neurale.....	283
Backpropagation: il passaggio chiave nell'addestramento della rete neurale.....	283
Potenziali problemi: dall'overfitting alla scomparsa dei gradienti.....	285
Tecniche di addestramento delle reti neurali: regolarizzazione e dropout.....	285
Altre funzioni di attivazione: la tangente iperbolica (tanh) e l'unità lineare rettificata (ReLU).....	287
Reti neurali con più di un output: la funzione softmax.....	288
Iperparametri.....	289
Programmare le reti neurali in Keras.....	290
Un esempio grafico bidimensionale.....	290
Costruire il modello in Keras.....	292
Addestrare una rete neurale per il riconoscimento delle immagini.....	295
Reti neurali per la regressione.....	298
Altre architetture per dataset più complessi.....	299
Come vedono le reti neurali: le reti neurali convoluzionali CNN (Convolutional Neural Network).....	300
Come parlano le reti neurali: RNN, GRU e LSTM.....	300
Come le reti neurali realizzano dipinti: GAN.....	301
Riepilogo.....	301
Esercizi.....	302
Esercizio 10.1.....	302
Esercizio 10.2.....	303
Esercizio 10.3.....	304

## Capitolo 11 **Trovare i confini con le macchine a vettori di supporto e il metodo kernel .....305**

Nuova funzione d'errore per classificatori migliori.....	308
Funzione d'errore di classificazione: il tentativo di classificare correttamente i punti.....	308
Funzione d'errore di distanza: separare il più possibile le due linee.....	310



Sommare le due funzioni per ottenere la funzione d'errore .....	312
Classificatore più concentrato sulla classificazione	
o sulla distanza? Il parametro C .....	313
Programmare le macchine a vettori di supporto in Scikit-Learn ..	314
Programmare una semplice macchina a vettori	
di supporto .....	315
Il parametro C .....	316
Addestrare macchine a vettori di supporto con confini	
non lineari: il metodo kernel .....	317
Kernel, mappe delle caratteristiche e teoria degli operatori .....	318
Usare le equazioni polinomiali a nostro vantaggio:	
il kernel polinomiale .....	318
Utilizzare “monti” e “valli”: il kernel RBF .....	324
Addestrare una macchina a vettori di supporto con	
il kernel RBF .....	330
Programmare il metodo kernel .....	336
Riepilogo .....	339
Esercizi .....	339
Esercizio 11.1 .....	339
Esercizio 11.2 .....	340

## **Capitolo 12 Ensemble learning: combinare più modelli per massimizzare i risultati .....341**

Un aiuto dai nostri amici .....	342
Bagging: unione casuale di più learner deboli per creare	
un learner forte .....	344
Mandare in overfitting un albero decisionale .....	346
Approssimare manualmente una foresta casuale .....	346
Addestrare una foresta casuale in Scikit-Learn .....	348
AdaBoost: unire i learner deboli in modo intelligente	
per costruire un learner forte .....	350
L'idea alla base di AdaBoost: rafforzare i learner deboli .....	350
Combinare i learner deboli a formare un learner forte .....	352
Programmare AdaBoost in Scikit-Learn .....	357
Boosting del gradiente: utilizzare alberi decisionali	
per creare learner forti .....	359
XGBoost: un modo estremo per applicare il boosting	
del gradiente .....	364
Punteggio di similarità di XGBoost: misurare la similarità	
in un set .....	365
Costruire i learner deboli .....	366
Potare gli alberi: ridurre l'overfitting semplificando	
i learner deboli .....	368
Fare previsioni .....	370
Addestrare un modello XGBoost in Python .....	371

Applicazioni dei metodi ensemble .....	373
Riepilogo .....	374
Esercizi .....	374
Esercizio 12.1 .....	374
Esercizio 12.2 .....	374

## **Capitolo 13 Esempio pratico di ingegneria dei dati e machine learning .....375**

Il dataset Titanic.....	376
Le caratteristiche del dataset.....	377
Pandas e il caricamento del dataset.....	377
Utilizzare Pandas per studiare il dataset.....	379
Ripulire il dataset: valori mancanti e come gestirli .....	380
Eliminare le colonne con dati mancanti .....	381
Come non perdere l'intera colonna: integrazione dei dati mancanti .....	382
Ingegneria delle caratteristiche: trasformare le caratteristiche nel dataset prima di addestrare i modelli .....	383
Trasformare i dati categoriali in dati numerici: codifica one-hot .....	384
Trasformare i dati numerici in dati categoriali (e perché dovremmo farlo): raggruppamenti .....	386
Selezionare le caratteristiche: eliminare le caratteristiche non necessarie .....	387
Addestrare i modelli.....	388
Suddividere i dati in caratteristiche ed etichette, e poi addestramento e convalida.....	389
Addestrare più modelli sul dataset .....	390
Quale modello è migliore? Valutare i modelli .....	390
Sottoporre a test il modello.....	392
Ottimizzare gli iperparametri per trovare il modello migliore: ricerca su griglia.....	393
Utilizzare la cross-validation K-fold per riutilizzare i dati come addestramento e convalida .....	396
Riepilogo .....	397
Esercizi .....	397
Esercizio 13.1 .....	397

## **Appendice A Soluzioni degli esercizi.....399**

## **Appendice B I calcoli per la discesa del gradiente .....435**

Utilizzare la discesa del gradiente per minimizzare le funzioni .....	436
Utilizzare la discesa del gradiente per addestrare i modelli.....	438
Algoritmo di discesa del gradiente per l'addestramento di modelli di machine learning .....	439

---

Utilizzare la discesa del gradiente per addestrare modelli di regressione lineare.....	440
Utilizzare la discesa del gradiente per addestrare modelli di classificazione.....	442
Utilizzare la discesa del gradiente per addestrare le reti neurali .....	448
Utilizzare la discesa del gradiente per la regolarizzazione .....	450
Rimanere bloccati sui minimi locali: come accade e come risolverlo .....	451
<b>Appendice C Riferimenti.....</b>	<b>455</b>
<b>Indice analitico.....</b>	<b>463</b>