

Auf einen Blick

Über die Autoren	13
Einführung	25
Teil I: Einführung in das maschinelle Lernen	29
Kapitel 1: Künstliche Intelligenz in Fiktion und Realität	31
Kapitel 2: Lernen im Zeitalter von Big Data	43
Kapitel 3: Ein Ausblick auf die Zukunft	53
Teil II: Einrichtung Ihrer Programmierumgebung	63
Kapitel 4: Installation einer R-Distribution	65
Kapitel 5: Programmierung mit R und RStudio	83
Kapitel 6: Installation einer Python-Distribution	107
Kapitel 7: Programmierung mit Python und Anaconda	127
Kapitel 8: Weitere Softwareprogramme für maschinelles Lernen	151
Teil III: Mathematische Grundlagen	159
Kapitel 9: Mathematische Grundlagen des maschinellen Lernens	161
Kapitel 10: Fehlerfunktionen und ihre Minimierung	179
Kapitel 11: Validierung von maschinellem Lernen	191
Kapitel 12: Einfache Lerner	209
Teil IV: Aufbereitung und Verwendung von Daten zum Lernen	225
Kapitel 13: Vorverarbeitung von Daten	227
Kapitel 14: Ausnutzung von Ähnlichkeiten in Daten	245
Kapitel 15: Einfache Anwendung von linearen Modellen	265
Kapitel 16: Komplexere Lernverfahren und neuronale Netze	287
Kapitel 17: Support Vector Machines und Kernel-Funktionen	303
Kapitel 18: Kombination von Lernalgorithmen in Ensembles	321
Teil V: Praktische Anwendung von maschinellem Lernen	337
Kapitel 19: Klassifikation von Bildern	339
Kapitel 20: Bewertung von Meinungen und Stimmungslagen	353
Kapitel 21: Produkt- und Filmempfehlungen	373
Teil VI: Der Top-Ten-Teil	387
Kapitel 22: Zehn wichtige Pakete für maschinelles Lernen	389
Kapitel 23: Zehn Methoden zur Verbesserung Ihrer maschinellen Lernmodelle	395
Stichwortverzeichnis	403

Inhaltsverzeichnis

Über die Autoren	13
Einführung	25
Über dieses Buch.....	25
Grundvoraussetzungen.....	26
Symbole, die in diesem Buch verwendet werden.....	27
Weitere Ressourcen.....	27
Und nun?.....	27
TEIL I	
EINFÜHRUNG IN DAS MASCHINELLE LERNEN	29
Kapitel 1	
Künstliche Intelligenz in Fiktion und Realität	31
Eine realistische Betrachtung von KI.....	32
Träume von elektrischen Schafen.....	33
Die Entstehungsgeschichte von KI und maschinellem Lernen.....	33
Der Beitrag von maschinellem Lernen zur KI.....	34
Die Ziele des maschinellen Lernens.....	35
Einschränkungen beim maschinellen Lernen durch Hardware.....	35
Die Grenzen zwischen Fiktion und Realität.....	36
Visionäre Ideen für KI und maschinelles Lernen.....	37
Realistische Anwendungsfälle für KI und maschinelles Lernen.....	37
Banal und trotzdem nützlich.....	38
Die Beziehung zwischen KI und maschinellem Lernen.....	39
Die technischen Spezifikationen von KI und maschinellem Lernen.....	40
Technische und kreative Vorgehensweisen.....	41
Kapitel 2	
Lernen im Zeitalter von Big Data	43
Definition von Big Data.....	44
Mögliche Quellen für Big Data.....	45
Erzeugung einer neuen Datenquelle.....	45
Nutzung vorhandener Datenquellen.....	47
Quellen für Testdaten finden.....	47
Die Statistik und das maschinelle Lernen.....	48
Die Rolle von Algorithmen.....	49
Funktionsweise von Algorithmen.....	49
Fünf wesentliche Techniken.....	49
Das Training von Algorithmen.....	51

Kapitel 3	
Ein Ausblick auf die Zukunft	53
Nützliche Technologien für die Zukunft	54
Maschinelles Lernen und Roboter	54
Maschinelles Lernen im Gesundheitswesen	55
Intelligente Systeme für unterschiedlichste Anforderungen	55
Maschinelles Lernen in industriellen Anwendungsbereichen	56
Die Wichtigkeit aktueller Prozessoren und Hardware	57
Neue Arbeitsfelder durch maschinelles Lernen	57
Eine Maschine als Chef	57
Maschinelle Systeme im Alltag	58
Reparatur von Maschinen	58
Erzeugung neuer Aufgaben für maschinelles Lernen	59
Gestaltung neuer maschineller Lernumgebungen	59
Potenzielle Tücken zukünftiger Technologien	60
TEIL II	
EINRICHTUNG IHRER PROGRAMMIERUMGEBUNG	63
Kapitel 4	
Installation einer R-Distribution	65
Auswahl einer R-Distribution für maschinelles Lernen	66
Installation von R unter Windows	67
Installation von R unter Linux	74
Installation von R unter Mac OS X	76
Herunterladen der Quelltexte und Datensätze	77
Verwendete Datensätze in diesem Buch	78
Zentraler Speicherort für den Programmcode	79
Kapitel 5	
Programmierung mit R und RStudio	83
Wichtige Datentypen	83
Verwendung von Vektoren	86
Datenorganisation mit Listen	86
Verwendung von Matrizen	87
Erzeugung einer einfachen Matrix	88
Änderung der Vektoranordnung	89
Zugriff auf individuelle Elemente	89
Namen für Zeilen und Spalten	90
Nutzung mehrerer Dimensionen mit Arrays	91
Erzeugung eines einfachen Arrays	91
Namen für Zeilen und Spalten	92
Nutzung von Data-Frames	93
Funktionsweise von Faktoren	93
Erzeugung von einfachen Data-Frames	95
Interaktion mit Data-Frames	96
Erweiterung eines Data-Frames	97

Durchführung einfacher statistischer Aufgaben.....	99
Entscheidungsfindung.....	99
Nutzung von Schleifen.....	101
Ausführung schleifenartiger Aufgaben ohne Schleifen.....	102
Verwendung von Funktionen.....	103
Arithmetisches Mittel und Median.....	103
Diagrammdarstellung Ihrer Daten.....	105

Kapitel 6

Installation einer Python-Distribution 107

Auswahl einer Python-Distribution für maschinelles Lernen.....	107
Anaconda von Continuum Analytics.....	109
Canopy Express von Enthought.....	109
Python(x,y).....	110
WinPython.....	110
Installation von Python unter Linux.....	111
Installation von Python unter Mac OSX.....	112
Installation von Python unter Windows.....	113
Herunterladen der Quelltexte und Datensätze.....	117
Verwendung von Jupyter Notebook.....	117
Zentraler Speicherort für den Programmcode.....	118
Verwendete Datensätze in diesem Buch.....	124

Kapitel 7

Programmierung mit Python und Anaconda 127

Zahlen und logische Ausdrücke in Python.....	128
Variablenzuweisung.....	129
Arithmetische Operatoren.....	130
Vergleich von Daten mit booleschen Ausdrücken.....	131
Erzeugung und Verwendung von Zeichenketten.....	133
Interaktion mit Datums- und Zeitangaben.....	134
Erzeugung und Verwendung von Funktionen.....	135
Erzeugung wiederverwendbarer Funktionen.....	135
Funktionsaufruf.....	137
Globale und lokale Variablen.....	139
Bedingungen und Schleifen.....	139
Entscheidungsfindung mit der »if«-Anweisung.....	139
Auswahl zwischen mehreren Optionen durch Verschachtelung.....	141
Wiederholung von Aufgaben mit der »for«-Schleife.....	141
Verwendung der »while«-Anweisung.....	142
Datenspeicherung mit Mengen, Listen und Tupeln.....	143
Erzeugung von Mengen.....	143
Mengenoperationen.....	144
Erzeugung von Listen.....	145
Erzeugung und Verwendung von Tupeln.....	146
Definition nützlicher Iteratoren.....	147
Datenindizierung mit Wörterbüchern.....	148
Codespeicherung in Modulen.....	149

Kapitel 8
Weitere Softwareprogramme für maschinelles Lernen 151

- Die Vorgänger: SAS, Stata und SPSS 152
- Lernen im akademischen Sektor mit Weka 154
- Einfacher Zugriff auf komplexe Algorithmen mit LIBSVM 155
- Höchstgeschwindigkeit mit Vowpal Wabbit. 155
- Visualisierung mit Knime und RapidMiner. 156
- Verwaltung riesiger Datenmengen mit Spark 157

TEIL III
MATHEMATISCHE GRUNDLAGEN 159

Kapitel 9
Mathematische Grundlagen des maschinellen Lernens 161

- Die Arbeit mit Daten 162
 - Erzeugung einer Matrix 163
 - Grundlegende Operationen 165
 - Matrixmultiplikation 166
 - Ein Blick auf fortgeschrittene Matrixoperationen 168
 - Effektive Nutzung von Vektorisierung. 169
- Die Welt der Wahrscheinlichkeiten 171
 - Operationen mit Wahrscheinlichkeiten 172
 - Bedingte Wahrscheinlichkeiten und Satz von Bayes 173
- Nutzung der Statistik für maschinelles Lernen 176

Kapitel 10
Fehlerfunktionen und ihre Minimierung 179

- Der Lernprozess als Optimierung 180
 - Überwachtes Lernen 180
 - Unüberwachtes Lernen. 180
 - Verstärkendes Lernen 181
 - Der Lernprozess. 181
- Kostenfunktionen 184
- Minimierung der Fehlerfunktion. 186
- Aktualisierung per Mini-Batch- und Online-Lernen 188

Kapitel 11
Validierung von maschinellem Lernen 191

- Fehler durch inkorrekte Stichprobenerhebung 192
 - Suche nach Generalisierungen 193
- Der Einfluss von Bias. 194
- Beachtung der Komplexität des Modells 196
- Ausgeglichene Lösungen 197
 - Darstellung von Lernkurven. 198
- Training, Validierung und Test. 200
- Kreuzvalidierung 201

Alternativen bei der Validierung	202
Optimierung von Kreuzvalidierungsverfahren	203
Erkundung des Hyperparameterraums	204
Vermeidung von Datenlecks und Bias in Stichproben	206
Probleme durch Snooping	207
Kapitel 12	
Einfache Lerner	209
Das faszinierende Perzeptron	210
Eine clevere Formel	210
Die Grenzen der Trennbarkeit	212
Klassifikationsbäume und der Greedy-Ansatz	214
Vorhersage von Ergebnissen durch Datenzerlegung	214
Stützen von großen Bäumen	217
Wahrscheinlichkeitsbasierte Algorithmen	219
Funktionsweise des naiven Bayes-Klassifikators	219
Schätzung mit dem naiven Bayes-Klassifikator	222
TEIL IV	
AUFBEREITUNG UND VERWENDUNG VON DATEN	
ZUM LERNEN	225
Kapitel 13	
Vorverarbeitung von Daten	227
Erfassung und Bereinigung von Daten	228
Korrektur von fehlenden Daten	229
Identifizierung von fehlenden Daten	229
Auswahl einer geeigneten Ersetzungsstrategie	230
Transformation von Verteilungen	233
Erzeugung Ihrer eigenen Merkmale	235
Die Notwendigkeit neuer Merkmale	235
Automatische Erzeugung von Merkmalen	235
Komprimierung von Daten	237
Abgrenzung anomaler Daten	239
Kapitel 14	
Ausnutzung von Ähnlichkeiten in Daten	245
Messung der Ähnlichkeit zwischen Vektoren	246
Definition von »Ähnlichkeit«	246
Berechnung von Abständen beim maschinellen Lernen	247
Suche nach Clustern durch Berechnung von Abständen	248
Überprüfung von Annahmen und Erwartungen	249
Funktionsweise des k-Means-Algorithmus	250
Feinanpassung des k-Means-Algorithmus	252
Experimente zur Zuverlässigkeit von k-Means	253
Experimente zur Konvergenz von Zentroiden	255

Klassifikation mit k-Nearest Neighbors	258
Auswahl des korrekten Parameters k.	259
Die Rolle des Parameters k.	259
Experimente mit einem flexiblen Algorithmus	260
Kapitel 15	
Einfache Anwendung von linearen Modellen	265
Kombination von Variablen	266
Vermischung von Variablen unterschiedlichen Typs	271
Nutzung von Wahrscheinlichkeiten	274
Spezifikation einer binären Reaktion	275
Verfahrensweise bei mehr als zwei Klassen.	277
Schätzung der richtigen Merkmale	278
Vermeidung irreführender Ergebnisse durch inkompatible Merkmale. ...	278
Merkmalsauswahl zur Vermeidung einer Überanpassung	279
Lernen aus einzelnen Beispielen.	281
Verwendung des Gradientenabstiegs.	281
Stochastische Gradientenabstiegsverfahren	282
Kapitel 16	
Komplexere Lernverfahren und neuronale Netze	287
Imitation der Natur beim Lernen	288
Vorwärtsausrichtung in Feedforward-Netzen	289
Schichten und noch mehr Schichten.	291
Fehlerkorrektur mit Rückpropagierung	294
Vermeidung von Überanpassung	296
Ursache einer Überanpassung.	297
Ein Blick hinter die Kulissen	297
Einführung in Deep Learning.	300
Kapitel 17	
Support Vector Machines und Kernel-Funktionen	303
Ein neuer Ansatz für das Problem der Separierbarkeit	304
Die Funktionsweise des Algorithmus	305
Mathematische Grundlagen der SVM.	307
Vermeidung von Problemen durch Nichtseparierbarkeit	308
Nichtlinearität.	309
Beispiel für den Kernel-Trick	311
Unterschiedliche Kernel	312
Implementierung und Hyperparameter	313
Klassifikation und Schätzung mit einer SVM	315
Kapitel 18	
Kombination von Lernalgorithmen in Ensembles	321
Kombination von Entscheidungsbäumen	322
Ein ganzer Wald aus Entscheidungsbäumen	323
Wichtigkeitsmaße	327

Verwendung beinahe zufälliger Schätzungen	330
Bagging von Prädiktoren mit Adaboost	331
Boosting von intelligenten Prädiktoren	333
Nutzung eines Gradientenabstiegsverfahrens	334
Durchschnitt verschiedener Prädiktoren	335

TEIL V

PRAKTISCHE ANWENDUNG VON MASCHINELLEM LERNEN ... 337

Kapitel 19 Klassifikation von Bildern 339

Die Arbeit mit Bildern	340
Extraktion visueller Merkmale	344
Gesichtserkennung mit Eigengesichtern	345
Klassifikation von Bildern	348

Kapitel 20

Bewertung von Meinungen und Stimmungslagen 353

Einführung in die Verarbeitung natürlicher Sprache	353
Lesende Maschinen	354
Verarbeitung und Aufbereitung von Text	356
Auslesen von Textdaten aus dem Internet	360
Probleme mit reinen Textdaten	363
Bewertung und Klassifikation von Texten	365
Durchführung von Klassifikationsaufgaben	365
Analyse von Produktrezensionen	367

Kapitel 21

Produkt- und Filmempfehlungen 373

Revolutionäre Systeme	374
Bewertungsdaten aus dem Internet	375
Der MovieLens-Datensatz	375
Ein anonymisierter Webdatensatz	377
Bewertungsdaten und ihre Grenzen	378
Nutzung der Singulärwertzerlegung	380
Ursprünge der SWZ	380
Erkenntnisse dank SWZ	381
Die SWZ in Aktion	382

TEIL VI

DER TOP-TEN-TEIL 387

Kapitel 22

Zehn wichtige Pakete für maschinelles Lernen 389

Oryx 2	390
CUDA-Convnet	390
ConvNetJS	390
e1071	391

gbm	391
Gensim	392
glmnet	392
randomForest	392
SciPy.....	393
XGBoost.....	393

Kapitel 23

Zehn Methoden zur Verbesserung Ihrer maschinellen

Lernmodelle..... 395

Auswertung von Lernkurven	396
Korrekte Verwendung der Kreuzvalidierung	397
Auswahl der geeigneten Fehler- oder Bewertungsmaße	398
Suche nach den besten Hyperparametern	398
Test von mehreren Modellen.....	399
Bildung des Durchschnitts verschiedener Modelle	399
Mehrstufige Kombination von Modellen	400
Erzeugung neuer Merkmale.....	401
Auswahl von Merkmalen und Beispielen.....	401
Suche nach mehr Daten	402

Stichwortverzeichnis..... 403