

Auf einen Blick

1	Einführung	17
2	Allgemeines	35
3	K-nächste Nachbarn	75
4	K-Means Clustering.....	87
5	Naiver Bayes-Klassifikator	107
6	Regressionsmethoden.....	115
7	Support-Vektor-Maschinen	123
8	Selbstorganisierende Karten.....	137
9	Entscheidungsbäume	151
10	Neuronale Netze	173
11	Verstärkendes Lernen	199

Inhalt

Vorwort	13
1 Einführung	17
1.1 Maschinelles Lernen	18
1.2 Lernen ist der Schlüssel	19
1.3 Ein wenig Geschichte	20
1.4 Schlüsselmethodiken in diesem Buch	22
1.5 Klassische mathematische Modellierung	26
1.6 Maschinelles Lernen ist anders	28
1.7 Einfachheit führt zu Komplexität	29
1.8 Weiterführende Literatur	33
2 Allgemeines	35
2.1 Jargon und Notation	35
2.2 Skalierung	37
2.3 Distanzmessung	38
2.4 Fluch der Dimensionalität	39
2.5 Hauptkomponentenanalyse	39
2.6 Maximum-Likelihood-Schätzung	40
2.7 Konfusionsmatrix	44
2.8 Kostenfunktion	47
2.9 Gradientenabstieg	52
2.10 Training, Testen und Validieren	54
2.11 Bias und Varianz	57
2.12 Lagrange-Multiplikatoren	63

2.13	Mehrfachklassen	65
2.14	Informationstheorie und Entropie	67
2.15	Verarbeitung natürlicher Sprache (NLP)	70
2.16	Bayes-Theorem	72
2.17	Was nun?	73
2.18	Weiterführende Literatur	74

3 K-nächste Nachbarn 75

3.1	Wofür können wir die Methode verwenden?	75
3.2	Wie die Methode funktioniert	76
3.3	Der Algorithmus	78
3.4	Probleme mit KNN	78
3.5	Beispiel: Körpergröße und -gewicht	79
3.6	Regression	83
3.7	Weiterführende Literatur	85

4 K-Means Clustering 87

4.1	Wofür können wir die Methode verwenden?	87
4.2	Was macht K-Means Clustering?	89
4.3	Scree-Plots	93
4.4	Beispiel: Kriminalität in England, 13 Dimensionen	94
4.5	Beispiel: Volatilität	98
4.6	Beispiel: Zinssatz und Inflation	100
4.7	Beispiel: Zinssätze, Inflation und BIP-Wachstum	103
4.8	Ein paar Kommentare	104
4.9	Weiterführende Literatur	105

5	Naiver Bayes-Klassifikator	107
5.1	Wofür können wir ihn verwenden?	107
5.2	Verwendung des Bayes-Theorems	108
5.3	Anwendung des NBK	108
5.4	In Symbolen	110
5.5	Beispiel: Politische Reden	111
5.6	Weiterführende Literatur	114
6	Regressionsmethoden	115
6.1	Wofür können wir sie verwenden?	115
6.2	Mehrdimensionale lineare Regression	116
6.3	Logistische Regression	117
6.4	Beispiel: Noch einmal politische Reden	119
6.5	Weitere Regressionsmethoden	121
6.6	Weiterführende Literatur	122
7	Support-Vektor-Maschinen	123
7.1	Wofür können wir sie verwenden?	123
7.2	Harte Ränder	123
7.3	Beispiel: Iris (Schwertlilie)	126
7.4	Lagrange-Multiplier-Version	128
7.5	Weiche Ränder	130
7.6	Kernel-Trick	132
7.7	Weiterführende Literatur	136

8	Selbstorganisierende Karten	137
8.1	Wofür können wir sie verwenden?	137
8.2	Die Methode	138
8.3	Der Lernalgorithmus	140
8.4	Beispiel: Gruppierung von Aktien	142
8.5	Beispiel: Abstimmungen im Unterhaus	147
8.6	Weiterführende Literatur	149
9	Entscheidungsbäume	151
9.1	Wofür können wir sie verwenden?	151
9.2	Beispiel: Zeitschriftenabo	153
9.3	Entropie	158
9.4	Überanpassung und Abbruchregeln	161
9.5	Zuschneiden	162
9.6	Numerische Merkmale/Attribute	162
9.7	Regression	164
9.8	Ausblick	171
9.9	Bagging und Random Forest	171
9.10	Weiterführende Literatur	172
10	Neuronale Netze	173
10.1	Wofür können wir sie verwenden?	173
10.2	Ein sehr einfaches Netzwerk	173
10.3	Universelles Approximations-Theorem	174
10.4	Ein noch einfacheres Netzwerk	176
10.5	Die mathematische Manipulation im Detail	177
10.6	Häufige Aktivierungsfunktionen	181

10.7	Das Ziel	182
10.8	Beispiel: Approximation einer Funktion	183
10.9	Kostenfunktion	184
10.10	Backpropagation	185
10.11	Beispiel: Buchstabenerkennung	188
10.12	Training und Testen	190
10.13	Mehr Architekturen	194
10.14	Deep Learning	196
10.15	Weiterführende Literatur	197

11

Verstärkendes Lernen

199

11.1	Wofür können wir es verwenden?	199
11.2	Geländeausfahrt mit Ihrem Lamborghini 400 GT	200
11.3	Jargon	202
11.4	Ein erster Blick auf Blackjack	203
11.5	Der klassische Markow-Entscheidungsprozess für Tic-Tac-Toe	204
11.6	Noch mehr Jargon	206
11.7	Beispiel: Der mehrarmige Bandit	207
11.8	Etwas anspruchsvoller 1: Bekannte Umgebung	211
11.9	Beispiel: Ein Labyrinth	214
11.10	Notation zu Wertefunktionen	218
11.11	Die Bellman-Gleichung	220
11.12	Optimale Policy	221
11.13	Die Bedeutung der Wahrscheinlichkeit	222
11.14	Etwas anspruchsvoller 2	223
11.15	Monte Carlo Policy Evaluation	224
11.16	Temporal-Difference-Lernen	227
11.17	Vor- und Nachteile: MC versus TD	228
11.18	Finden der optimalen Policy	229

11.19 Sarsa	230
11.20 Q-Lernen	232
11.21 Beispiel: Blackjack	233
11.22 Große Zustandsräume	245
11.23 Weiterführende Literatur	245
Datensätze	247
Epilog	251
Index	253