

# Inhalt

<b>Einleitung</b> .....	<b>XIII</b>
<b>1 Vektoren, Matrizen und Arrays</b> .....	<b>1</b>
1.0 Einführung .....	1
1.1 Einen Vektor erzeugen .....	1
1.2 Eine Matrix erstellen .....	2
1.3 Eine dünn besetzte Matrix erzeugen .....	3
1.4 Elemente auswählen .....	5
1.5 Eine Matrix beschreiben .....	6
1.6 Operationen auf Elemente anwenden .....	7
1.7 Die Größt- und Kleinstwerte suchen .....	8
1.8 Mittelwert, Varianz und Standardabweichung berechnen .....	9
1.9 Die Gestalt von Arrays ändern .....	10
1.10 Einen Vektor oder eine Matrix transponieren .....	11
1.11 Eine Matrix verflachen .....	12
1.12 Den Rang einer Matrix ermitteln .....	12
1.13 Die Determinante berechnen .....	13
1.14 Die Diagonale einer Matrix ermitteln .....	14
1.15 Die Spur einer Matrix berechnen .....	14
1.16 Eigenwerte und Eigenvektoren suchen .....	15
1.17 Punktprodukte berechnen .....	16
1.18 Matrizen addieren und subtrahieren .....	17
1.19 Matrizen multiplizieren .....	18
1.20 Eine Matrix invertieren .....	19
1.21 Zufallswerte erzeugen .....	20
<b>2 Laden von Daten</b> .....	<b>23</b>
2.0 Einführung .....	23
2.1 Einen Beispieldatensatz laden .....	23
2.2 Einen simulierten Datensatz erzeugen .....	25

2.3	Eine CSV-Datei laden . . . . .	27
2.4	Eine Excel-Datei laden . . . . .	28
2.5	Eine JSON-Datei laden . . . . .	29
2.6	Eine SQL-Datenbank abfragen . . . . .	30
<b>3</b>	<b>Datenaufbereitung . . . . .</b>	<b>33</b>
3.0	Einführung . . . . .	33
3.1	Einen Dataframe erstellen . . . . .	34
3.2	Die Daten beschreiben . . . . .	35
3.3	In DataFrames navigieren . . . . .	37
3.4	Zeilen abhängig von Bedingungen auswählen . . . . .	39
3.5	Werte ersetzen . . . . .	40
3.6	Spalten umbenennen . . . . .	41
3.7	Minimum, Maximum, Summe, Mittelwert und Anzahl ermitteln . . . . .	42
3.8	Eindeutige Werte ermitteln . . . . .	43
3.9	Fehlende Werte behandeln . . . . .	45
3.10	Eine Spalte löschen . . . . .	46
3.11	Eine Zeile löschen . . . . .	47
3.12	Doppelte Zeilen löschen . . . . .	49
3.13	Zeilen nach Werten gruppieren . . . . .	50
3.14	Zeilen nach Zeit gruppieren . . . . .	52
3.15	Eine Spalte in einer Schleife durchlaufen . . . . .	54
3.16	Eine Funktion auf alle Elemente in einer Spalte anwenden . . . . .	55
3.17	Eine Funktion auf Gruppen anwenden . . . . .	55
3.18	DataFrames verketteten . . . . .	56
3.19	DataFrames zusammenführen . . . . .	58
<b>4</b>	<b>Numerische Daten verarbeiten . . . . .</b>	<b>61</b>
4.0	Einführung . . . . .	61
4.1	Ein Merkmal neu skalieren . . . . .	61
4.2	Ein Merkmal standardisieren . . . . .	63
4.3	Beobachtungen normalisieren . . . . .	64
4.4	Polynom- und Interaktionsmerkmale erzeugen . . . . .	66
4.5	Merkmale transformieren . . . . .	68
4.6	Ausreißer erkennen . . . . .	69
4.7	Mit Ausreißern umgehen . . . . .	71
4.8	Merkmale diskretisieren . . . . .	73
4.9	Beobachtungen durch Clustern gruppieren . . . . .	75
4.10	Beobachtungen mit fehlenden Werten löschen . . . . .	76
4.11	Fehlende Werte imputieren . . . . .	78

<b>5</b>	<b>Kategorische Daten behandeln</b> .....	<b>81</b>
5.0	Einführung .....	81
5.1	Nominale kategorische Merkmale codieren .....	82
5.2	Ordinale kategorische Merkmale codieren .....	84
5.3	Wörterbücher von Merkmalen codieren .....	86
5.4	Fehlende Klassenwerte imputieren .....	88
5.5	Umgang mit unausgewogenen Klassen .....	90
<b>6</b>	<b>Text verarbeiten</b> .....	<b>95</b>
6.0	Einführung .....	95
6.1	Text bereinigen .....	95
6.2	HTML parsen und bereinigen .....	97
6.3	Satzzeichen entfernen .....	98
6.4	Text in Tokens zerlegen .....	99
6.5	Stoppwörter entfernen .....	99
6.6	Stemming .....	100
6.7	Wortarten markieren .....	101
6.8	Text als Bag of Words codieren .....	104
6.9	Wortbedeutung gewichten .....	106
<b>7</b>	<b>Datum und Uhrzeit</b> .....	<b>109</b>
7.0	Einführung .....	109
7.1	Strings in Datumswerte konvertieren .....	109
7.2	Umgang mit Zeitzonen .....	111
7.3	Datum und Uhrzeit auswählen .....	112
7.4	Datumsdaten in mehrere Merkmale aufgliedern .....	113
7.5	Die Differenz zwischen Datumswerten bilden .....	114
7.6	Wochentage codieren .....	115
7.7	Ein verzögertes Merkmal erstellen .....	116
7.8	Gleitende Zeitfenster verwenden .....	117
7.9	Fehlende Daten in Zeitreihen behandeln .....	118
<b>8</b>	<b>Bilder verarbeiten</b> .....	<b>121</b>
8.0	Einführung .....	121
8.1	Bilder laden .....	122
8.2	Bilder speichern .....	124
8.3	Bilder in der Größe ändern .....	125
8.4	Bilder beschneiden .....	126
8.5	Bilder weichzeichnen .....	127
8.6	Bilder scharfzeichnen .....	129
8.7	Kontrast erhöhen .....	130

8.8	Farben isolieren . . . . .	132
8.9	Binärbilder erzeugen . . . . .	134
8.10	Hintergründe entfernen . . . . .	136
8.11	Kantendetektion . . . . .	138
8.12	Eckenerkennung . . . . .	140
8.13	Merkmale für maschinelles Lernen erzeugen . . . . .	143
8.14	Die mittlere Farbe als Merkmal codieren . . . . .	145
8.15	Farbhistogramme als Merkmale codieren . . . . .	146
<b>9</b>	<b>Verringerung der Dimensionalität durch Merkmalsextraktion . . . . .</b>	<b>151</b>
9.0	Einführung . . . . .	151
9.1	Merkmale mithilfe von Hauptkomponenten reduzieren . . . . .	152
9.2	Merkmale reduzieren, wenn Daten nicht linear separierbar sind . . . . .	154
9.3	Merkmale durch Maximierung der Klassenseparierbarkeit reduzieren . . . . .	156
9.4	Merkmale durch Matrixfaktorisierung reduzieren . . . . .	159
9.5	Merkmale auf schwach besetzten Daten reduzieren . . . . .	160
<b>10</b>	<b>Dimensionalität durch Merkmalsauswahl reduzieren . . . . .</b>	<b>163</b>
10.0	Einführung . . . . .	163
10.1	Numerische Merkmale nach dem Schwellenwert von Varianzen auswählen . . . . .	163
10.2	Binäre Merkmale nach dem Schwellenwert von Varianzen auswählen . . . . .	165
10.3	Stark korrelierte Merkmale verarbeiten . . . . .	166
10.4	Für eine Klassifizierung irrelevante Merkmale entfernen . . . . .	168
10.5	Merkmale rekursiv entfernen . . . . .	170
<b>11</b>	<b>Modellbewertung . . . . .</b>	<b>173</b>
11.0	Einführung . . . . .	173
11.1	Kreuzvalidierungsmodelle . . . . .	173
11.2	Ein Baseline-Regressionsmodell erstellen . . . . .	177
11.3	Ein Baseline-Klassifizierungsmodell erstellen . . . . .	179
11.4	Vorhersagen binärer Klassifikatoren bewerten . . . . .	180
11.5	Schwellenwerte von binären Klassifikatoren bewerten . . . . .	183
11.6	Mehrklassige Klassifikatorvorhersagen bewerten . . . . .	187
11.7	Die Performance eines Klassifikators visualisieren . . . . .	188
11.8	Regressionsmodelle bewerten . . . . .	191
11.9	Clustermodelle bewerten . . . . .	192
11.10	Eine benutzerdefinierte Bewertungsmetrik erstellen . . . . .	194
11.11	Die Wirkung der Trainingsmengengröße visualisieren . . . . .	195

11.12	Einen Textbericht der Bewertungsmetriken erstellen . . . . .	198
11.13	Die Wirkung von Hyperparameter-Werten visualisieren . . . . .	199
<b>12</b>	<b>Modellauswahl . . . . .</b>	<b>203</b>
12.0	Einführung . . . . .	203
12.1	Beste Modelle mittels erschöpfender Suche auswählen . . . . .	204
12.2	Beste Modelle mittels Zufallssuche auswählen . . . . .	206
12.3	Beste Modelle von mehreren Lernalgorithmen auswählen . . . . .	208
12.4	Beste Modelle bei Vorverarbeitung auswählen . . . . .	210
12.5	Die Modellauswahl durch Parallelisierung beschleunigen . . . . .	212
12.6	Die Modellauswahl mit algorithmusspezifischen Methoden beschleunigen . . . . .	213
12.7	Performance nach der Modellauswahl bewerten . . . . .	215
<b>13</b>	<b>Lineare Regression . . . . .</b>	<b>217</b>
13.0	Einführung . . . . .	217
13.1	Eine Gerade anpassen . . . . .	217
13.2	Interaktive Effekte verarbeiten . . . . .	219
13.3	Eine nichtlineare Beziehung anpassen . . . . .	221
13.4	Varianz durch Regularisierung verringern . . . . .	223
13.5	Merkmale mit Lasso-Regression reduzieren . . . . .	225
<b>14</b>	<b>Bäume und Wälder . . . . .</b>	<b>227</b>
14.0	Einführung . . . . .	227
14.1	Einen Klassifikator mittels Entscheidungsbaum trainieren . . . . .	227
14.2	Einen Entscheidungsbaumregressor trainieren . . . . .	229
14.3	Ein Entscheidungsbaummodell visualisieren . . . . .	230
14.4	Einen Random-Forest-Klassifikator trainieren . . . . .	232
14.5	Einen Random-Forest-Regressor trainieren . . . . .	234
14.6	Wichtige Merkmale in Random Forests identifizieren . . . . .	235
14.7	Wichtige Merkmale in Random Forests auswählen . . . . .	237
14.8	Unausgewogene Klassen behandeln . . . . .	239
14.9	Die Baumgröße steuern . . . . .	240
14.10	Performance durch Boosting verbessern . . . . .	242
14.11	Random Forests mit Out-of-Bag-Fehlern bewerten . . . . .	243
<b>15</b>	<b>K-nächste Nachbarn . . . . .</b>	<b>245</b>
15.0	Einführung . . . . .	245
15.1	Die nächsten Nachbarn einer Beobachtung suchen . . . . .	245
15.2	Einen k-nächste-Nachbarn-Klassifikator erzeugen . . . . .	248
15.3	Die Größe der besten Nachbarschaft ermitteln . . . . .	250
15.4	Einen radiusbasierten Nächste-Nachbarn-Klassifikator erstellen . . . . .	251

<b>16</b>	<b>Logistische Regression</b> .....	<b>253</b>
16.0	Einführung .....	253
16.1	Einen binären Klassifikator trainieren .....	253
16.2	Einen Multiklassenklassifikator trainieren .....	255
16.3	Varianz durch Regularisierung verringern .....	256
16.4	Einen Klassifikator auf sehr großen Daten trainieren .....	257
16.5	Unausgewogene Klassen verarbeiten .....	258
<b>17</b>	<b>Support Vector Machines</b> .....	<b>261</b>
17.0	Einführung .....	261
17.1	Einen linearen Klassifikator trainieren .....	261
17.2	Nicht linear separierbare Klassen mit Kernels behandeln .....	264
17.3	Vorhergesagte Wahrscheinlichkeiten erzeugen .....	267
17.4	Stützvektoren bestimmen .....	269
17.5	Unausgewogene Klassen verarbeiten .....	270
<b>18</b>	<b>Naive Bayes-Klassifikatoren</b> .....	<b>273</b>
18.0	Einführung .....	273
18.1	Einen Klassifikator für kontinuierliche Merkmale trainieren ...	274
18.2	Einen Klassifikator für diskrete und abzählbare Merkmale trainieren .....	276
18.3	Einen naiven Bayes-Klassifikator für binäre Merkmale trainieren .....	277
18.4	Vorhergesagte Wahrscheinlichkeiten kalibrieren .....	278
<b>19</b>	<b>Clustering</b> .....	<b>281</b>
19.0	Einführung .....	281
19.1	K-Means-Clustering .....	281
19.2	K-Means-Clustering beschleunigen .....	284
19.3	Clustering mittels Mean-Shift .....	285
19.4	Clustering durch hierarchisches Zusammenführen .....	288
<b>20</b>	<b>Neuronale Netze</b> .....	<b>291</b>
20.0	Einführung .....	291
20.1	Vorverarbeitung der Daten für neuronale Netze .....	292
20.2	Ein neuronales Netz entwerfen .....	294
20.3	Einen binären Klassifikator trainieren .....	297
20.4	Einen Klassifikator für mehrere Klassen trainieren .....	299
20.5	Einen Regressor trainieren .....	301
20.6	Vorhersagen treffen .....	303
20.7	Den Trainingsverlauf visualisieren .....	304

20.8	Überanpassung durch Regularisierung der Gewichte vermindern. . . . .	307
20.9	Überanpassung durch frühes Stoppen verringern . . . . .	309
20.10	Überanpassung durch Dropout verringern . . . . .	311
20.11	Den Fortschritt beim Training eines Modells speichern . . . . .	313
20.12	Neuronale Netze mit k-facher Kreuzvalidierung . . . . .	315
20.13	Neuronale Netze optimieren . . . . .	317
20.14	Neuronale Netze visualisieren . . . . .	319
20.15	Bilder klassifizieren . . . . .	320
20.16	Höhere Performance durch Bildverbesserung . . . . .	324
20.17	Text klassifizieren . . . . .	326
<b>21</b>	<b>Trainierte Modelle speichern und laden . . . . .</b>	<b>331</b>
21.0	Einführung . . . . .	331
21.1	Ein scikit-learn-Modell speichern und laden . . . . .	331
21.2	Ein Keras-Modell speichern und laden . . . . .	333
	<b>Index . . . . .</b>	<b>335</b>