

Mathe-Basics für Data Scientists

Lineare Algebra, Statistik und
Wahrscheinlichkeitsrechnung für die
Datenanalyse

DAS INHALTS- VERZEICHNIS

» Hier geht's
direkt
zum Buch

Einführung	13
Grundlegende Mathematik und Infinitesimalrechnung	19
Zahlentheorie	20
Reihenfolge der Operationen	21
Variablen	23
Funktionen	24
Summationen	28
Potenzen	30
Logarithmen	33
Eulersche Zahl und natürliche Logarithmen	35
Die eulersche Zahl	35
Natürliche Logarithmen	38
Grenzwerte	39
Ableitungen	41
Partielle Ableitungen	44
Die Kettenregel	47
Integrale	49
Zum Schluss	54
Übungen	54
2 Wahrscheinlichkeit	55
Wahrscheinlichkeit verstehen	55
Wahrscheinlichkeitsrechnung vs. Statistik	57
Wahrscheinlichkeitsmathematik	58
Kombinierte Wahrscheinlichkeiten	58
Vereinigungswahrscheinlichkeiten	59
Bedingte Wahrscheinlichkeit und der Satz von Bayes	61
Kombinierte und vereinigte bedingte Wahrscheinlichkeiten	63
Binomialverteilung	65
Beta-Verteilung	67

Zum Schluss	73
Übungen	74
3 Deskriptive und inferenzielle Statistik	75
Was sind Daten?	75
Deskriptive versus inferenzielle Statistik	78
Grundgesamtheiten, Stichproben und Verzerrungen	78
Deskriptive Statistik	82
Mittelwert und gewichteter Mittelwert	83
Median	84
Modus	85
Varianz und Standardabweichung	86
Die Normalverteilung.	91
Die inverse CDF.	97
z-Werte.	98
Inferenzielle Statistik	100
Der zentrale Grenzwertsatz	101
Konfidenzintervalle	103
Was sind p-Werte?	106
Hypothesentests.	107
Die t-Verteilung: mit kleinen Stichproben umgehen	115
Big Data und der Zielscheibenfehler.	116
Zum Schluss	118
Übungen	119
4 Lineare Algebra	121
Was ist ein Vektor?	121
Vektoren hinzufügen und kombinieren	125
Vektoren skalieren	127
Lineare Hülle (Spann) und lineare Abhängigkeit.	129
Lineare Transformationen	131
Basisvektoren	131
Matrix-Vektor-Multiplikation	134
Matrixmultiplikation	138
Determinanten	141
Spezielle Matrixtypen	144
Quadratische Matrix	144
Identitätsmatrix	145
Inverse Matrix	145
Diagonalmatrix	146
Dreiecksmatrix.	146
Dünnbesetzte Matrix	146
Gleichungssysteme und inverse Matrizen.	147

Eigenvektoren und Eigenwerte	151
Zum Schluss	153
Übungen	154
5 Lineare Regression	155
Eine einfache lineare Regression	156
Einfache lineare Regression mit scikit-learn	159
Residuen und Fehlerquadrate	160
Die beste Anpassungsgerade suchen	163
Gleichung in geschlossener Form	164
Techniken mit inversen Matrizen	165
Gradientenabstieg	167
Überanpassung und Varianz	173
Stochastischer Gradientenabstieg	175
Der Korrelationskoeffizient	177
Statistische Signifikanz	180
Bestimmtheitsmaß	185
Standardfehler der Schätzung	186
Vorhersageintervalle	187
Aufteilung in Trainings- und Testdaten	190
Multiple lineare Regression	196
Zum Schluss	197
Übungen	197
6 Logistische Regression und Klassifikation	199
Logistische Regression verstehen	199
Eine logistische Regression durchführen	202
Logistische Funktion	202
Die logistische Kurve anpassen	204
Multivariable logistische Regression	210
Das Wesen der Log-Odds	213
R-Quadrat	217
p-Werte	221
Aufteilung in Trainings- und Testdaten	223
Wahrheitsmatrizen	224
Der Satz von Bayes und Klassifizierung	227
ROC-Kurve/Fläche unter der Kurve	228
Klassenungleichgewicht	229
Zum Schluss	230
Übungen	231
7 Neuronale Netze	233
Wann man neuronale Netze und Deep Learning verwendet	234

Ein einfaches neuronales Netz	235
Aktivierungsfunktionen	237
Forward Propagation	242
Backpropagation	248
Die Ableitungen von Gewichts- und Schwellenwerten berechnen. . .	248
Stochastischer Gradientenabstieg	252
Die Bibliothek scikit-learn	256
Grenzen von neuronalen Netzen und Deep Learning.	257
Zum Schluss	260
Übung	261
8 Karriereberatung und der Weg in die Zukunft	263
Data Science – neu definiert	264
Data Science – ein geschichtlicher Abriss	266
Ihr eigenes Profil schärfen.	269
SQL-Kenntnisse	269
Programmierkenntnisse	272
Datenvisualisierung	275
Branchenkenntnisse	277
Produktives Lernen	279
Praktiker vs. Ratgeber	279
Worauf Sie bei Data-Science-Jobs achten sollten	282
Rollendefinition	283
Organisatorischer Fokus und Akzeptanz	283
Genügend Ressourcen	285
Vernünftige Ziele	286
Mit bestehenden Systemen konkurrieren.	287
Eine Rolle ist nicht das, was Sie erwartet haben	289
Existiert Ihr Traumjob nicht?	291
Wie geht es weiter?	291
Zum Schluss	292
Anhang A Ergänzende Themen	295
LaTeX-Rendering mit SymPy	295
Binomialverteilung von Grund auf	297
Beta-Verteilung von Grund auf.	298
Den Satz von Bayes ableiten	299
CDF und inverse CDF von Grund auf	301
Ereigniswahrscheinlichkeit mit e im Zeitverlauf vorhersagen.	302
Bergsteigeralgorithmus und lineare Regression	304
Bergsteigeralgorithmus und logistische Regression	306
Eine kurze Einführung in lineare Optimierung.	307
MNIST-Klassifizierer mit scikit-learn	312

Anhang B Lösungen zu den Übungen	315
Index	327