

Inhaltsverzeichnis

	Über den Autor	5
	Einführung	13
1	Alles, was Sie jemals über Tabellenkalkulationen wissen wollen, sich aber nicht zu fragen getraut haben	23
1.1	Beispieldaten	24
1.2	Sich schnell mit der Steuerungstaste bewegen	25
1.3	Formeln und Daten schnell kopieren	26
1.4	Zellen formatieren	28
1.5	Inhalte einfügen	29
1.6	Diagramme hinzufügen	30
1.7	Die Menüs »Suchen« und »Ersetzen«	32
1.8	Formeln für das Auffinden und Entnehmen von Werten	32
1.9	SVERWEIS verwenden, um Daten zusammenzuführen	34
1.10	Filtern und sortieren	35
1.11	Pivot-Tabellen verwenden	39
1.12	Array-Formeln verwenden	42
1.13	Probleme mit dem Solver lösen	44
1.14	OpenSolver: Ich wünschte, wir würden ihn nicht benötigen. Dem ist aber nicht so	50
1.15	Zusammenfassung	51
2	Clusteranalyse Teil I: Die Kundenbasis mit k-Means aufteilen	53
2.1	Mädchen tanzen mit Mädchen, und Jungens kratzen sich am Kopf . .	55
2.2	Es wird ernst: k-Means-Clusterbildung bei Abonnenten eines E-Mail-Marketings	60
2.2.1	Joey Bag O' Donuts Weinhandel	60
2.2.2	Die Ausgangsdaten	61
2.2.3	Festlegen, was zu bewerten ist	62
2.2.4	Mit vier Clustern beginnen	65
2.2.5	Euklidischer Abstand: Abstandsmessung auf kürzestem Weg	67
2.2.6	Abstände und Clusterzuweisungen für jedermann	69
2.2.7	Clusterzentren bestimmen	71

2.2.8	Aus den Ergebnissen schlau werden	74
2.2.9	Die Top-Verkäufe je Cluster erhalten	75
2.2.10	Die Silhouette: Ein guter Weg, um es unterschiedliche k-Werte unter sich ausfechten zu lassen	79
2.2.11	Was halten Sie von fünf Clustern?	87
2.2.12	Eine Lösung für fünf Cluster	88
2.2.13	Die Top-Verkäufe der fünf Cluster erhalten	89
2.2.14	Die Silhouette für die 5-Means-Clusterbildung berechnen . .	92
2.3	K-Medians-Clusterbildung und asymmetrische Abstandsmessungen . .	93
2.3.1	Die k-Medians-Clusterbildung	94
2.3.2	Eine geeignetere Abstandsmetrik erhalten	94
2.3.3	Bringen Sie das alles in Excel unter	97
2.3.4	Die Top-Verkäufe der 5-Medians-Cluster	98
2.4	Zusammenfassung	102
3	Naives Bayes und wie unglaublich leicht es ist, ein Idiot zu sein .	105
3.1	Wenn Sie ein Produkt »Mandrill« nennen, erhalten Sie Signale und Nebengeräusche	105
3.2	Die kürzeste Einführung in die Wahrscheinlichkeitsrechnung der Welt	108
3.2.1	Bedingte Wahrscheinlichkeiten summieren	108
3.2.2	Die Verbundwahrscheinlichkeit, die Kettenregel und die Unabhängigkeit	109
3.2.3	Was geschieht in einer abhängigen Situation?	110
3.2.4	Die Bayes-Regel	110
3.3	Die Bayes-Regel verwenden, um ein KI-Modell zu erstellen	111
3.3.1	Klassenwahrscheinlichkeiten auf hohem Niveau werden oft miteinander gleichgesetzt	113
3.3.2	Und noch ein paar Kleinigkeiten	114
3.4	Auf geht's mit Excel	116
3.4.1	Für die Sache irrelevante Interpunktion entfernen	117
3.4.2	An Leerzeichen auftrennen	118
3.4.3	Token zählen und Wahrscheinlichkeiten berechnen	122
3.4.4	Wir haben ein Modell! Nutzen wir es!	124
3.5	Zusammenfassung	130
4	Optimierungsmodellierung: Weil der »frisch gepresste« Orangensaft sich nicht selbst herstellt	133
4.1	Warum sollten Data Scientists wissen, was Optimierung bedeutet? . . .	134
4.2	Mit einem einfachen Zielkonflikt geht es los	135
4.2.1	Das Problem als Polytop darstellen	136
4.2.2	Lösen durch Verschieben der Niveaumenge	139
4.2.3	Das Simplex-Verfahren: in den Ecken herumstöbern	140

4.2.4	Mit Excel arbeiten	141
4.2.5	Am Ende dieses Kapitels wartet ein Monster	152
4.3	Frisch vom Baum in Ihr Glas ... mit einem kurzen Boxenstopp fürs Mischen	153
4.3.1	Sie verwenden für das Mischen ein Modell	154
4.3.2	Beginnen wir mit ein paar Spezifikationen	154
4.3.3	Zurück zum gleichbleibenden Geschmack	156
4.3.4	Die Daten in Excel eintragen	157
4.3.5	Das Problem in Solver eingeben	161
4.3.6	Die Standards herabsetzen	163
4.3.7	Ein totes Eichhörnchen loswerden: der Minimax-Ansatz . . .	167
4.3.8	Wenn-Dann- und die Big-M-Bedingung	169
4.3.9	Variablen vervielfachen: das Volumen bis auf 11 hochtreiben	174
4.4	Modellierungsrisiko	182
4.4.1	Normal verteilte Daten	183
4.5	Zusammenfassung	192
5	Clusteranalyse Teil II: Netzwerkdiagramme und die Entdeckung der Community	195
5.1	Was ist ein Netzwerkdiagramm?	196
5.2	Einen einfachen Graphen darstellen	197
5.3	Eine kurze Einführung in Gephi	200
5.3.1	Die Installation von Gephi und die Vorbereitung der Dateien	201
5.3.2	Den Graphen gestalten	203
5.3.3	Rangfolge von Knoten	205
5.3.4	Drucken	208
5.3.5	Dem Graphen an die Daten gehen	209
5.4	Aus den Daten des Weinhandels einen Graphen bilden	210
5.4.1	Eine Kosinus-Ähnlichkeitsmatrix erstellen	213
5.4.2	Einen r-Nachbarschaftsgraphen entwickeln	216
5.5	Wie viel ist eine Kante wert? Normale Punkte und Penaltys bei der Modularität von Graphen	221
5.5.1	Was ist ein Punkt und woraus besteht ein Penalty?	221
5.5.2	Das Arbeitsblatt für die Bewertungen einrichten	225
5.6	Lassen Sie uns Cluster bilden!	227
5.6.1	Aufteilung Nummer 1	228
5.6.2	Aufteilung 2: Electric Boogaloo	234
5.6.3	Und ... Aufteilung 3: Aufteilung mit Vergeltung	236
5.6.4	Die Communitys decodieren und analysieren	237
5.7	Einmal hin und wieder zurück: eine Gephi-Tabelle	242
5.8	Zusammenfassung	247

6	Der Großvater der betreuten künstlichen Intelligenz – die Regression	249
6.1	He, was bist du? Schwanger?	249
6.2	Machen Sie sich nicht selbst verrückt	250
6.3	Die Schwangerschaft von Kundinnen bei RetailMart mithilfe der linearen Regression vorhersagen	251
6.3.1	Welche Funktionen benötigt werden	252
6.3.2	Die Trainingsdaten zusammenstellen	253
6.3.3	Dummy-Variablen erzeugen	255
6.3.4	Backen wir uns unsere eigene lineare Regression	258
6.3.5	Statistiken und lineare Regression: R-Quadrat, F-Test und t-Tests	268
6.3.6	Vorhersagen anhand neuer Daten tätigen und die Leistungs- fähigkeit messen	279
6.4	Mit einer logistischen Regression Schwangerschaften in Kundenhaus- halten vorhersagen	290
6.4.1	Als Erstes benötigen Sie eine Verknüpfungsfunktion	290
6.4.2	Die logistische Funktion einbinden und alles neu optimieren	292
6.4.3	Eine echte logistische Regression zusammenbauen	294
6.4.4	Modellauswahl – die Leistungsfähigkeit des linearen mit der des logistischen Modells vergleichen	297
6.5	Wenn Sie mehr wissen wollen	300
6.6	Zusammenfassung	301
7	Ensemble-Modelle: eine Menge mieser Pizza	303
7.1	Die Daten aus Kapitel 6 verwenden	304
7.2	Bagging: zufällig anordnen, trainieren, wiederholen	306
7.2.1	Decision Stump ist keine sehr sexy Bezeichnung für eine blöde Vorhersage	307
7.2.2	Das sieht für mich gar nicht mal so dumm aus!	308
7.2.3	Das Modell untersuchen	319
7.3	Boosting: Wenn das Ergebnis falsch ist, verstärken Sie es und versuchen es auf ein Neues	324
7.3.1	Das Modell trainieren – jedes Merkmal wird angesprochen	325
7.3.2	Das verstärkte Modell auswerten	333
7.4	Zusammenfassung	337
8	Prognosen: Atmen Sie tief durch, Sie können nicht gewinnen	339
8.1	Der Handel mit Schwertern stottert	340
8.2	Mit Zeitreihen vertraut werden	341
8.3	Langsam Fahrt aufnehmen mit einer einfachen exponentiellen Glättung	343
8.3.1	Prognosen mit der einfachen exponentiellen Glättung einrichten	346

8.4	Es könnte ein Trend vorliegen	351
8.5	Die lineare exponentielle Glättung nach Holt	355
8.5.1	Die lineare exponentielle Glättung nach Holt in einem Arbeitsblatt einrichten	356
8.5.2	Sind Sie nun fertig? Einen Blick auf Autokorrelationen werfen	362
8.6	Die multiplikative Glättung nach Holt-Winters	369
8.6.1	Die Anfangswerte für Niveau, Trend und Saisonabhängigkeit festlegen	371
8.6.2	Die Prognose ins Rollen bringen	376
8.6.3	Optimieren!	381
8.6.4	Bestätigen Sie mir jetzt bitte, dass wir fertig sind	383
8.6.5	Um die Prognose einen Vorhersagebereich legen	383
8.6.6	Für die Galerie: Ein Fan-Chart anlegen	388
8.7	Zusammenfassung	390
9	Die Entdeckung von Ausreißern: Nur weil sie sonderbar sind, heißt das nicht, dass sie auch unwichtig sind	393
9.1	Auch Ausreißer sind nur (schlechte?) Menschen	394
9.2	Der faszinierende Fall von Hadlum gegen Hadlum	395
9.2.1	Tukey-Begrenzungen	396
9.2.2	Tukey-Begrenzungen in einem Arbeitsblatt anwenden	397
9.2.3	Die Grenzen dieser einfachen Vorgehensweise	399
9.3	In nichts wirklich schlecht, aber auch nirgends wirklich gut	401
9.3.1	Daten für einen Graphen vorbereiten	402
9.3.2	Einen Graphen erstellen	405
9.3.3	Die k nächsten Nachbarn erhalten	407
9.3.4	Methode 1 zum Entdecken von Ausreißern in einem Graphen: Verwenden Sie einfach den Indegree	408
9.3.5	Methode 2 zum Entdecken von Ausreißern in einem Graphen: Differenzierte Ergebnisse mit k-Abstand erhalten .	412
9.3.6	Methode 3 zum Entdecken von Ausreißern in einem Gra- phen: Local Outlier Factors sind dort, wo die Musik spielt .	414
9.4	Zusammenfassung	419
10	Von der Tabellenkalkulation zu R wechseln	421
10.1	Mit R loslegen	422
10.1.1	Ein paar einfache Fingerübungen	423
10.1.2	Daten in R einlesen	431
10.2	Sich aktiv mit Data Science beschäftigen	433
10.2.1	Ein paar Zeilen sphärisches k-Means für Wein-Daten	433

10.3	Mit den Schwangerschaftsdaten ein KI-Modell entwickeln	440
10.3.1	Prognosen in R tätigen	449
10.3.2	Sich um das Entdecken von Ausreißern kümmern	454
10.4	Zusammenfassung	458
	Stichwortverzeichnis	459