

Stichwortverzeichnis

Symbole

68-95-99,7-Regel 185

A

ABS() 267

Abstand

– euklidischer 67

Abstandsmatrix 81

Abweichung

– mittlere 185

– Standardabweichung 185

– Standardabweichung berechnen 185

Adjazenz 199

Adjazenzmatrix 199

Adstringens 155

Affinitätsmatrix 200, 216

AIMMS 153

alpha 329, 345

alpha-Wert, Genauigkeit 330

Ausreißer 267, 393

– R 454

Autokorrelation 362

– berechnen 365

B

Backlink 206

Bag-of-Words-Modell 107

Bagged Decision Stump 303

Bagging 306

– Funktionsweise 310

– Vorteile 324

BAR 187

Bayes

– naives 105

– naives, Implementierungsbeispiele 131

– Regel verwenden 111

Begrenzung

– obere äußere 396

– obere innere 396

– untere äußere 396

– untere innere 396

Bereich

– interdeziler 403

– interquantiler 403

– zulässiger 137

BEREICH.VERSCHIEBEN 320

BEREICH.VERSCHIEBEN() 33

BI 14

Big Data 16

Big M 171, 229

BININDEZ() 237

Boosting 324

– Leistungsfähigkeit berechnen 335

– Modell auswerten 333

Boxplot 420, 455

– Werte anzeigen 456

Branch-and-Bound-Algorithmus 148

Brix 155

Business Analytics 14

Business Intelligence 14

C

Cluster

– Schwerpunkt 55

– Zentrum bestimmen 71

– Zuweisung 69

Clusteranalyse 53

Clusterbildung 53

Clustering 53

Clusterverfahren

– agglomeratives 228

– divisives 227

| Stichwortverzeichnis

Community 195
– decodieren 237
Community Detection 196
Competitive Intelligence 14
CPLEX 152
CRAN 433
Cut-Off 281
– festlegen 282

D

Data Analysis And Modeling 14
Data Science 13
– Definition 15
Data-Mining, exploratives 54
Daten
– kategorisierte 256
– kategorisierte, umwandeln 256
– normal verteilte 183
– quantitative 256
– skalieren 457
– standardisieren 64
Daten-Frame 428
Decision Stump 307
– erhalten 313
delta 370
Designmatrix 275
Dokument, Klassifizierung 105
Dummy-Codierung 256

E

Ensemble-Modellierung 303
Entscheidungsstumpf 303
Entscheidungsvariable, binäre 170
Erreichbarkeitsdistanz 415
Euklidische Norm 67
Euklidischer Abstand 67
Excel
– Arbeitsblatt kopieren 45
– Array-Formel 43
– Begriff finden 32
– bewegen in 25
– Diagramm 30
– filtern 35
– Flächendiagramm 388
– Formatierung, bedingte 29
– Formeln kopieren 26

– Grundlagen 23
– Inhalte einfügen 29
– Kopfzeile einfrieren 25
– Makro aufzeichnen 318
– nichtlineare Funktionen 150
– Pivot-Tabellen 39
– PivotTable-Feldliste 40
– Silhouette einrichten 84
– Solver 44
– Spalte einfügen 66, 81
– transponieren 30
– Zeile einfügen 81
– Zellen formatieren 28

F

F-Statistik 271
F-Test 271
F-Verteilung 271
Fan-Chart 388
Feature 252
Fehler
– gewichteter 325
– quadrierter 263
Fehlerberechnung 263
Fehleroptimierung 348
FINDEN() 119
Forecasting 339
Freiheitsgrad 257, 353
FVERT() 272

G

gamma 355
Gephi 199, 242
– Daten exportieren 243
– drucken 208
– Einführung 200
– Graphen gestalten 203
– Installation 201
– Modularität 242
Gewichtung normalisieren 331
Glättung
– additive 115, 123
– exponentielle 343
– exponentielle, doppelte 355
– exponentielle, einfache 343
– exponentielle, einrichten 346

- exponentielle, Funktionsweise 344
 - Holt 355
 - multiplikative 369
 - Parameter festlegen 381
 - Glättungskonstante 343, 345
 - Gleitender Mittelwert 372
 - Gleitkomma-Unterlauf 115
 - glm() 440
 - Glockenkurve 183
 - Google Drive 20
 - Graph
 - gewichteter 216
 - kNN 217
 - kolorieren 243
 - Modularität 221
 - r-Nachbarschaftsgraphen 217, 407
 - Grenzwertoptimierungskurve 287
 - Grenzwertsatz, zentraler 183
 - GRG 265
 - Gurobi 152
- H**
- Hamming-Abstand 94
 - HBase 16
 - Holt
 - Glättung 355
 - Glättung einrichten 356
 - Holt-Winters 369
- I**
- Imputation 305
 - Indegree 409
 - Independenz 109
 - INDEX() 33, 85
 - Ingraph Library 248
 - Interzept 259
 - IQR 396
 - ISTNV() 127
- K**
- k 65
 - k-Abstand 412
 - k-Means 54
 - R 433
 - sphärisches 97, 433
 - Vorgehensweise 65
 - k-Medians 94
 - Kante 196
 - KDD 14, 55
 - KI 15, 249
 - KI-Modell 133
 - entwickeln 440
 - KKLEINSTE() 33
 - Klassifikator, schwacher 307
 - KLEIN() 117
 - KLEINSTE() 84
 - kNN 394
 - kNN-Graph 217
 - erstellen 408
 - Knoten 196
 - Eingangsgrad 206
 - Rang 205
 - Knotenunreinheit 308
 - Knowledge Discovery 14, 55
 - Knowledge Extraction 14
 - Koeffizient 259
 - Korrelogramm 366
 - Kosinus-Ähnlichkeit 95
 - Kosinus-Ähnlichkeitsmatrix 213
 - Kosinus-Abstand 95
- L**
- Leistungsfähigkeit 335
 - Level 427
 - LibreOffice 19, 24
 - Lineare Programmierung 134
 - Lineares Programm 136
 - Linkspam 206
 - Liste, R 428
 - LOF 414, 457
 - Logistische Funktion 291
 - Louvain-Methode 227
- M**
- MAD 403
 - Mahout 16
 - Manhattan-Abstand 94
 - MAP 112
 - Marktsegmentierung 53
 - Maschinelles Lernen
 - betreutes 54
 - unbetreutes 54

| Stichwortverzeichnis

Matrix
– multiplizieren 275
– R 427
– zweidimensionale 427
Matrixmultiplikation 273
– Excel 273
Maximum-a-posteriori-Regel 112
Median 306
– absolute Abweichung 403
Minimax 167
MINV() 273
Mittelwert
– gekappter/gestutzter 403
– interquantiler 403
– winsorisierter 403
MITTELWERTWENN() 84
MMULT() 275
Modell
– lineares 258
– Merkmal 252
– trainieren 262
– vergleichen 290
Modellierungsrisiko 182
Modellvergleich 297
Modularität
– Maximierung 196, 220
– Summe der Bewertungen 244
MongoDB 16
Monte-Carlo-Simulation 186, 383
MTRANS() 43, 229

N

Natural Language Processing 116
Neigung, Standardfehler 353
Neo4J 248
Netzwerkdiagramm 196
– Daten codieren 248
– gerichtetes 199
– ungerichtetes 199
Netzwerkgraph 195
Niveau 343
– berechnen 344
Niveaumenge 139
NLP 116
Node Impurity 308
NodeXL 199

Normalverteilung 183
NORMINV() 187, 384
NORMVERT() 395
Nullhypothese 271
NumPy 16
NV 127

O

OpenSolver 50, 152
Operations Research 14
OPL 153
Optimierung 44, 134
Optimierungsmodell 133
Optimierungsproblem 44
Outlier 393
Oversampling 255

P

p() 108
p-Wert 271
Partitionierung, hierarchische 227
Penalty 221, 325
Performance-Objekt 446
Polygon 137
Polytop 137
Prädikator 252
Präzision 283
prediction() 446
Prognose, R 449
Prognosefehler 346

Q

Quantil
– drittes 396
– erstes 396
QUANTIL() 386, 396
Quantilmittel 403

R

R 422
– aggregate() 437
– Arbeitsverzeichnis 431
– Ausreißer 454
– boxplot() 455
– c() 425

- cbind() 428, 438
 - Console 422
 - data.frame() 429
 - Datei einlesen 431
 - Daten einlesen 431
 - Daten hinzufügen 428
 - Datenexport 436
 - Datenzugriff 428
 - DMwR 457
 - factor() 427
 - forecast() 451
 - forecast, Paket 450
 - getwd() 431
 - glm() 441
 - Hilfe 423
 - install.package() 434
 - IQR() 454
 - is.na() 432
 - Kurve zeichnen 447
 - Length() 425
 - library() 434
 - lofactor() 457
 - logische Anweisungen 426
 - logistische Regression 441
 - Matrix 427
 - matrix() 427
 - ncol() 435
 - order() 439
 - Paket laden 433
 - Paket suchen 433
 - performance() 446
 - predict() 443
 - print 422
 - randomForest() 443
 - rbind() 428
 - rechnen mit 422
 - ROCR 441
 - ROCR, Paket 446
 - row.name() 436
 - scale() 457
 - sortieren 439
 - summary() 426
 - t() 427
 - Textausgabe 422
 - Textdaten 426
 - transponieren 427
 - ts() 450
 - varImpPlot() 443
 - which() 426, 436, 455
 - Working Directory 431
 - write.csv() 436
 - Zeilen zählen 427
 - r-Nachbarschaftsgraph 217
 - R-Quadrat 270
 - Rückverweis 206
 - Random Forest 303, 323
 - randomForest 323
 - Paket 441
 - RANG() 408
 - Reachability Distance 415
 - read.csv() 431
 - Recall 287
 - Regression 250
 - lineare 251
 - logistische 267, 290
 - logistische, Tests 297
 - mediane 267
 - Relation 195
 - RGP() 267, 352
 - ROC-Kurve 287, 323
 - R 441
- S**
- Saisonabhängiger Einfluss 362
 - Sensitivität 287
 - SES 343
 - setwd() 431
 - Silhouette 79
 - berechnen 92
 - Definition 80
 - Simplex-Verfahren 140
 - skmeans 433
 - Solver
 - evolutionärer Algorithmus 149
 - Excel 44
 - GRG-Nichtlinearer Algorithmus 149
 - Spezifität 284
 - SSCP 275
 - Standardabweichung 65, 185
 - berechnen 185
 - Ersatz für 403
 - Szenarien 186

| Stichwortverzeichnis

Standardfehler 348
– Neigung 353
Standardisieren 402
STANDARDISIERUNG() 404
Standardnormalverteilung 402
Stichprobe, zufallsgesteuerte 310
Stoppwort 121
str() 430
Stub 222
summa() 430
SUMMENPRODUKT() 43
SUMMEWENN() 76
SUMQUADABW() 269
SVERWEIS() 34, 35

T

t-Test 274, 352
– aufrufen 278
– zweiseitiger 278
t-Verteilung 278
TEIL() 120
Testtrennwert 281
Titan 248
Token 111
– isolieren 118
– zählen 122
Trainingsdaten zusammenstellen 253
Trend 351
Trimean 403
Triple Exponential Smoothing 369
Tukey
– Begrenzungen 396
– Begrenzungen anwenden 397
– Grenzen 399
TVERT() 278, 352, 354
Typ-I-Fehler 282
Typ-II-Fehler 282

U

Überanpassung 324

V

Validierungsdaten 279

Variable
– abhängige 252
– Dummy-Variable 255
– unabhängige 252
VERGLEICH() 33
Verknüpfungsfunktion 290
Verteilung, multimodale 400
Verteilungsfunktion, kumulative 183, 395
Voronoi-Diagramm 57

W

Wahrscheinlichkeit
– bedingte 108
– Gesetz der totalen 108
– Kettenregel 109
– konditionale 108
– Multiplikationsregel 109
– Verbundwahrscheinlichkeit 109
Wahrscheinlichkeitsrechnung 108
Wahrscheinlichkeitsverteilung 183
Weak Learner 307
WECHSELN() 117
WENNFEHLER() 97
Wortmengenmodell 107
WURZEL() 97
WVERWEIS() 34

Y

y-Achsenabschnitt 259

Z

ZÄHLENWENN() 409
ZÄHLENWENNNS() 284, 314, 326
Zeitreihe 341
– R 450
Zentralwert 306
Zielkonflikt 135
Zielvorgabe 44
Zufallswald 303, 323
ZUFALLSZAHL() 310
Zurechnung
– mehrfache 306
– nächster Nachbar 306