

Ethem Alpaydin

Maschinelles Lernen

3., aktualisierte und erweiterte Auflage

DE GRUYTER
OLDENBOURG

Autor

Prof. Ethem Alpaydin
Department of Computer Engineering
Bogazici University
Istanbul, Turkey
alpaydin@boun.edu.tr

ISBN 978-3-11-074014-1
e-ISBN (PDF) 978-3-11-074019-6
e-ISBN (EPUB) 978-3-11-074027-1

Library of Congress Control Number: 2021948047

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar

© 2022 Walter de Gruyter GmbH, Berlin/Boston
Autorisierte Übersetzung der 3. Englischsprachigen Ausgabe, die bei The MIT Press, Massachusetts Institute of Technology, unter dem Titel: „Introduction to Machine Learning“ erschienen ist.

© 2014 Massachusetts Institute of Technology
Übersetzung 1. Auflage (2008): Simone Linke
Übersetzung 2. Auflage (2018): Dr. Karen Lippert
Übersetzung 3. Auflage (2021): Dr. Karen Lippert
Einbandabbildung: Gettyimages / PhonlamaiPhoto
Druck und Bindung: CPI books GmbH, Leck

www.degruyter.com



Inhaltsverzeichnis

1	Einführung	1
1.1	Was ist maschinelles Lernen?	1
1.2	Beispiele für Anwendungen des maschinellen Lernens	4
1.2.1	Assoziationsregeln	4
1.2.2	Klassifikation	5
1.2.3	Regression	10
1.2.4	Unüberwachtes Lernen	11
1.2.5	Bestärkendes Lernen	12
1.3	Geschichte	13
1.4	Angrenzende Themenfelder	16
1.4.1	Hochleistungsrechnen	16
1.4.2	Datenschutz und Datensicherheit	17
1.4.3	Interpretierbarkeit und Vertrauenswürdigkeit des Modells	18
1.4.4	Data Science	19
1.5	Übungen	19
1.6	Literaturangaben	22
2	Überwachtes Lernen	25
2.1	Lernen einer Klasse aus Beispielen	25
2.2	Vapnik-Chervonenkis-Dimension	31
2.3	PAC-Lernen	32
2.4	Rauschen	34
2.5	Lernen multipler Klassen	36
2.6	Regression	38
2.7	Modellauswahl und Generalisierung	40
2.8	Dimensionen eines Algorithmus für überwachtes Lernen ..	45
2.9	Anmerkungen	47
2.10	Übungen	48
2.11	Literaturangaben	52
3	Bayessche Entscheidungstheorie	53
3.1	Einführung	53
3.2	Klassifikation	55
3.3	Verluste und Risiken	57
3.4	Diskriminanzfunktionen	60
3.5	Assoziationsregeln	61
3.6	Anmerkungen	64

3.7	Übungen	65
3.8	Literaturangaben	69
4	Parametrische Methoden	71
4.1	Einführung	71
4.2	Maximum-Likelihood-Schätzung	72
4.2.1	Bernoulli-Verteilung	73
4.2.2	Multinomiale Dichte	74
4.2.3	Gauß-Verteilung (Normalverteilung)	74
4.3	Bewertung eines Schätzers: Verzerrung und Varianz	75
4.4	Der Bayessche Schätzer	77
4.5	Parametrische Klassifikation	80
4.6	Regression	83
4.7	Das Verzerrung-Varianz-Dilemma	87
4.8	Modellauswahl	91
4.9	Anmerkungen	95
4.10	Übungen	95
4.11	Literaturangaben	98
5	Multivariate Methoden	101
5.1	Multivariate Daten	101
5.2	Parameterschätzung	102
5.3	Schätzung von fehlenden Werten	103
5.4	Multivariate Normalverteilung	104
5.5	Multivariate Klassifikation	108
5.6	Anpassen der Komplexität	114
5.7	Diskrete Merkmale	116
5.8	Multivariate Regression	118
5.9	Anmerkungen	119
5.10	Übungen	120
5.11	Literaturangaben	122
6	Dimensionalitätsreduktion	125
6.1	Einführung	125
6.2	Teilmengenselektion	126
6.3	Hauptkomponentenanalyse	131
6.4	Merkmalseinbettung	137
6.5	Faktorenanalyse	140
6.6	Singulärwertzerlegung und Faktorisierung von Matrizen ..	145
6.7	Multidimensionale Skalierung	147
6.8	Lineare Diskriminanzanalyse	150
6.9	Kanonische Korrelationsanalyse	154
6.10	Isomap	158
6.11	Lokal lineare Einbettung	160
6.12	Laplacesche Eigenmaps	162
6.13	t -verteilte stochastische Nachbareinbettung	165
6.14	Anmerkungen	167

6.15	Übungen	169
6.16	Literaturangaben	171
7	Clusteranalyse	173
7.1	Einführung	173
7.2	Mischungsdichten	174
7.3	k -Means-Clusteranalyse	175
7.4	Expectation-Maximization-Algorithmus	179
7.5	Mischungsmodelle mit verborgenen Variablen	185
7.6	Überwachtes Lernen nach einer Clusteranalyse	186
7.7	Spektrale Clusteranalyse	187
7.8	Hierarchische Clusteranalyse	189
7.9	Wahl der Clusteranzahl	191
7.10	Anmerkungen	192
7.11	Übungen	193
7.12	Literaturangaben	195
8	Nichtparametrische Methoden	199
8.1	Einführung	199
8.2	Nichtparametrische Dichteschätzung	200
8.2.1	Histogrammschätzer	201
8.2.2	Kernel-Schätzer	203
8.2.3	k -Nächste-Nachbarn-Schätzer	204
8.3	Verallgemeinerung auf multivariate Daten	205
8.4	Nichtparametrische Klassifikation	207
8.5	Verdichtete Nächste-Nachbarn-Methode	208
8.6	Abstandsbasierte Klassifikation	210
8.7	Ausreißerererkennung	213
8.8	Nichtparametrische Regression: Glättungsmodelle	214
8.8.1	Gleitende Mittelwertglättung	215
8.8.2	Glättung durch Kernel-Funktion	217
8.8.3	Gleitende Linienglättung	217
8.9	Wahl des glättenden Hyperparameters	218
8.10	Anmerkungen	220
8.11	Übungen	222
8.12	Literaturangaben	225
9	Entscheidungsbäume	227
9.1	Einführung	227
9.2	Univariate Bäume	229
9.2.1	Klassifikationsbäume	230
9.2.2	Regressionsbäume	234
9.3	Pruning	237
9.4	Regelextraktion aus Bäumen	239
9.5	Lernen von Regeln anhand von Daten	241
9.6	Multivariate Bäume	245
9.7	Anmerkungen	247

9.8	Übungen	250
9.9	Literaturangaben	252
10	Lineare Diskriminanz	255
10.1	Einführung	255
10.2	Generalisierung des linearen Modells	257
10.3	Geometrie der linearen Diskriminanz	258
10.3.1	Zwei Klassen	258
10.3.2	Multiple Klassen	260
10.4	Paarweise Trennung	261
10.5	Neubetrachtung der parametrischen Diskriminanz	263
10.6	Gradientenabstieg	265
10.7	Logistische Diskriminanz	266
10.7.1	Zwei Klassen	266
10.7.2	Multiple Klassen	270
10.7.3	Multiple Labels	275
10.8	Lernen von Rangordnungen	276
10.9	Anmerkungen	279
10.10	Übungen	279
10.11	Literaturangaben	282
11	Mehrlagige Perzeptronen	285
11.1	Einführung	285
11.1.1	Das Gehirn verstehen	286
11.1.2	Neuronale Netze und Parallelverarbeitung	287
11.2	Das Perzeptron	289
11.3	Training eines Perzeptrons	293
11.4	Lernen von Booleschen Funktionen	296
11.5	Mehrlagige Perzeptronen	298
11.6	Das MLP als universelle Näherungsfunktion	300
11.7	Backpropagation	302
11.7.1	Nichtlineare Regression	302
11.7.2	Zweiklassendiskriminanz	307
11.7.3	Diskriminanz bei multiplen Klassen	308
11.7.4	Diskriminanz bei multiplen Labels	308
11.8	Übertraining	309
11.9	Lernen verborgener Darstellungen	311
11.10	Autoencoder	315
11.11	Die word2vec-Architektur	317
11.12	Anmerkungen	321
11.13	Übungen	323
11.14	Literaturangaben	324
12	Tiefes Lernen	327
12.1	Einführung	327
12.2	Training mehrerer verborgener Schichten	331
12.2.1	Rektifizierte lineare Einheit	331

12.2.2	Initialisierung	332
12.2.3	Backpropagation mit mehreren verborgenen Schichten....	332
12.3	Verbessern der Trainingskonvergenz.....	335
12.3.1	Momentum.....	335
12.3.2	Adaptiver Lernfaktor	336
12.3.3	Batch-Normalisierung	338
12.4	Regularisierung.....	339
12.4.1	Hinweise	340
12.4.2	Gewichtsabbau	341
12.4.3	Dropout.....	344
12.5	Faltungsschichten.....	346
12.5.1	Die Idee	346
12.5.2	Formalisierung	347
12.5.3	Beispiele: LeNet-5 und AlexNet	351
12.5.4	Erweiterungen.....	352
12.5.5	Multimodale tiefe Netze	353
12.6	Einstellen der Netzwerkstruktur	355
12.6.1	Struktursuche und Hyperparametersuche	355
12.6.2	Sprungverbindungen.....	356
12.6.3	Gattereinheiten.....	358
12.7	Lernen von Sequenzen	358
12.7.1	Beispielaufgaben	358
12.7.2	Neuronale Netze mit Zeitverzögerung	359
12.7.3	Rekurrente Netze.....	359
12.7.4	Langes Kurzzeitgedächtnis.....	362
12.7.5	Gated Recurrent Unit (GRU).....	364
12.8	Generative gegnerische Netze	365
12.9	Anmerkungen	367
12.10	Übungen	369
12.11	Literaturangaben	371
13	Lokale Modelle	375
13.1	Einführung.....	375
13.2	Kompetitives Lernen	375
13.2.1	Online- k -Means-Algorithmus	376
13.2.2	Adaptive Resonanztheorie	381
13.2.3	Selbstorganisierende Merkmalskarten	382
13.3	Radiale Basisfunktionen	383
13.4	Regelbasiertes Wissen	390
13.5	Normalisierte Basisfunktionen	391
13.6	Kompetitive Basisfunktionen	393
13.7	Lernen mit Vektorquantisierung	396
13.8	Das MoE-Modell	396
13.8.1	Kooperative Experten	399
13.8.2	Kompetitive Experten	399
13.9	Hierarchische MoE und weiche Entscheidungsbäume....	400
13.10	Anmerkungen	402

13.11	Übungen	403
13.12	Literaturangaben	406
14	Kernel-Maschinen	409
14.1	Einführung	409
14.2	Die optimal trennende Hyperebene	411
14.3	Der nicht trennbare Fall: Soft-Margin-Trennebenen	415
14.4	ν -SVM	418
14.5	Kernel-Trick	419
14.6	Vektorielle Kernel	421
14.7	Definition von Kernen	423
14.8	Multiple-Kernel-Lernen	425
14.9	Mehrklassen-Kernel-Maschinen	427
14.10	Kernel-Maschinen und Regression	428
14.11	Kernel-Maschinen und Ranking	433
14.12	Einklassen-Kernel-Maschinen	434
14.13	Breiter-Margin-Nächster-Nachbar-Klassifikator	438
14.14	Dimensionalitätsreduktion mit Kernel	440
14.15	Anmerkungen	441
14.16	Übungen	442
14.17	Literaturangaben	444
15	Graphenmodelle	447
15.1	Einführung	447
15.2	Kanonische Fälle für bedingte Unabhängigkeit	449
15.3	Generative Modelle	457
15.4	d -Separation	459
15.5	Belief-Propagation	460
15.5.1	Ketten	460
15.5.2	Bäume	462
15.5.3	Mehrfachbäume	464
15.5.4	Verbindungsbaume	466
15.6	Ungerichtete Graphen: Markovsche Zufallsfelder	467
15.7	Lernen der Struktur eines Graphenmodells	470
15.8	Einflussdiagramme	471
15.9	Anmerkungen	472
15.10	Übungen	473
15.11	Literaturangaben	476
16	Hidden-Markov-Modelle	479
16.1	Einführung	479
16.2	Diskrete Markov-Prozesse	480
16.3	Hidden-Markov-Modelle	483
16.4	Drei Problemstellungen für HMMs	485
16.5	Evaluierungsproblem	485
16.6	Herausfinden der Zustandssequenz	489
16.7	Lernen von Modellparametern	491

16.8	Kontinuierliche Beobachtungen	494
16.9	Das HMM als Graphenmodell	495
16.10	Modellauswahl im HMM.....	498
16.11	Anmerkungen	501
16.12	Übungen	503
16.13	Literaturangaben	505
17	Bayessche Schätzung	507
17.1	Einführung	507
17.2	Parameter diskreter Verteilungen	511
17.2.1	$K > 2$ -Zustände: Dirichlet-Verteilung	511
17.2.2	$K = 2$ -Zustände: Beta-Verteilung	513
17.3	Parameter einer Gauß-Verteilung	513
17.3.1	Univariater Fall: Unbekannter Mittelwert, bekannte Varianz	513
17.3.2	Univariater Fall: Unbekannter Mittelwert, unbekannte Varianz	516
17.3.3	Multivariater Fall: Unbekannter Mittelwert, unbekannte Kovarianz	518
17.4	Bayessche Schätzung der Parameter einer Funktion	519
17.4.1	Regression	519
17.4.2	Regression mit Prior für die Präzision des Rauschens	523
17.4.3	Basis/Kernel-Funktionen	525
17.4.4	Bayessche Klassifikation	527
17.5	Wahl eines Priors.....	529
17.6	Bayesscher Modellvergleich	530
17.7	Bayessche Schätzung für ein Mischungsmodell.....	533
17.8	Nichtparametrische Bayessche Modelle	536
17.9	Gaußsche Prozesse	537
17.10	Dirichlet-Prozesse und Chinaestaurants	541
17.11	Latente Dirichlet-Allokation	542
17.12	Beta-Prozesse und indische Büffets.....	545
17.13	Anmerkungen	546
17.14	Übungen	547
17.15	Literaturangaben	548
18	Kombination mehrerer Lerner	551
18.1	Grundprinzip	551
18.2	Generierung diverser Lerner	552
18.3	Methoden der Modellkombination	555
18.4	Voting	556
18.5	Fehlerkorrekturcodes	560
18.6	Bagging	562
18.7	Boosting	563
18.8	Neubetrachtung des MoE-Modells	566
18.9	Geschachtelte Generalisierung	568
18.10	Feinabstimmung eines Ensembles	569

18.10.1	Wahl einer Teilmenge des Ensembles	569
18.10.2	Konstruktion von Metalernern	570
18.11	Kaskadierung	571
18.12	Anmerkungen	573
18.13	Übungen	575
18.14	Literaturangaben	577
19	Bestärkendes Lernen	581
19.1	Einführung	581
19.2	Fälle mit einem Zustand: K -armiger Bandit	583
19.3	Elemente des bestärkenden Lernens	584
19.4	Modellbasiertes Lernen	587
19.4.1	Wertiteration	587
19.4.2	Taktikiteration	588
19.5	Lernen mit temporaler Differenz	588
19.5.1	Explorationsstrategien	589
19.5.2	Deterministische Belohnungen und Aktionen	590
19.5.3	Nichtdeterministische Belohnungen und Aktionen	591
19.5.4	Eignungsprotokolle	594
19.6	Generalisierung	595
19.7	Teilweise beobachtbare Zustände	598
19.7.1	Setting	598
19.7.2	Beispiel: Das Tigerproblem	600
19.8	Tiefes Q -Lernen	605
19.9	Taktikgradienten	606
19.10	Backgammon und Go	609
19.11	Anmerkungen	610
19.12	Übungen	611
19.13	Literaturangaben	614
20	Experimente mit maschinellem Lernen	617
20.1	Einführung	617
20.2	Faktoren, Antwort und Strategie beim Experimentieren	621
20.3	Antwortflächenmethode	623
20.4	Randomisieren, Wiederholen und Blocken	624
20.5	Richtlinien für Experimente mit maschinellem Lernen	625
20.6	Kreuzvalidierung und Resampling-Methoden	629
20.6.1	K -fache Kreuzvalidierung	630
20.6.2	5×2 -Kreuzvalidierung	631
20.6.3	Bootstrapping	632
20.7	Leistungsmessung für Klassifikatoren	632
20.8	Intervallschätzung	636
20.9	Hypothesenprüfung	639
20.10	Leistungsbewertung für Klassifikationsalgorithmen	641
20.10.1	Binomialtest	642
20.10.2	Test der approximierten Normalverteilung	643
20.10.3	t -Test	643

20.11	Vergleich von zwei Klassifikationsalgorithmen	644
20.11.1	Der McNemarsche Test	644
20.11.2	Gepaarter t -Test mit K -facher Kreuzvalidierung	644
20.11.3	Gepaarter t -Test mit 5×2 Kreuzvalidierung	645
20.11.4	Gepaarter F -Test mit 5×2 Kreuzvalidierung	646
20.12	Vergleich mehrerer Algorithmen: Varianzanalyse	647
20.13	Vergleich über mehrere Datensätze	651
20.13.1	Vergleich zweier Algorithmen	653
20.13.2	Vergleich mehrerer Algorithmen	655
20.14	Multivariate Tests	656
20.14.1	Vergleich zweier Algorithmen	656
20.14.2	Vergleich mehrerer Algorithmen	658
20.15	Anmerkungen	659
20.16	Übungen	660
20.17	Literaturangaben	662
A	Wahrscheinlichkeit	663
A.1	Elemente der Wahrscheinlichkeit	663
A.1.1	Axiome der Wahrscheinlichkeit	663
A.1.2	Bedingte Wahrscheinlichkeit	664
A.2	Zufallsvariablen	665
A.2.1	Verteilungsfunktion und Wahrscheinlichkeitsdichte	665
A.2.2	Gemeinsame Verteilungsfunktion, gemeinsame Dichte	666
A.2.3	Bedingte Verteilungen	666
A.2.4	Satz von Bayes	666
A.2.5	Erwartungswert	667
A.2.6	Varianz	668
A.2.7	Das schwache Gesetz großer Zahlen	669
A.3	Spezielle Verteilungen von Zufallsvariablen	669
A.3.1	Bernoulli-Verteilung	669
A.3.2	Binomialverteilung	670
A.3.3	Multinomiale Verteilung	670
A.3.4	Gleichverteilung	670
A.3.5	Normalverteilung (Gauß-Verteilung)	671
A.3.6	Chi-Quadrat-Verteilung	672
A.3.7	t -Verteilung	673
A.3.8	F -Verteilung	673
A.4	Literaturangaben	673
B	Lineare Algebra	675
B.1	Vektoren	675
B.2	Matrizen	677
B.3	Ähnlichkeit zwischen Vektoren	678
B.4	Quadratische Matrizen	679
B.5	Lineare Abhängigkeit und Rang	679
B.6	Die inverse Matrix	680
B.7	Positiv definite Matrizen	680

B.8	Spur und Determinante.....	680
B.9	Eigenwerte und Eigenvektoren	681
B.10	Spektralzerlegung	682
B.11	Singulärwertzerlegung	682
B.12	Literaturangaben.....	683
C	Optimierung	685
C.1	Einführung.....	685
C.2	Lineare Optimierung	687
C.3	Konvexe Optimierung	687
C.4	Dualität	688
C.5	Lokale Optimierung	690
C.6	Literaturangaben.....	691
Index		693