

Inhaltsverzeichnis

1	Einführung	1
1.1	Was ist maschinelles Lernen?	1
1.2	Beispiele für Anwendungen des maschinellen Lernens.....	4
1.2.1	Erlernen von Assoziationen	4
1.2.2	Klassifikation	5
1.2.3	Regression	10
1.2.4	Unüberwachtes Lernen	12
1.2.5	Bestärkendes Lernen	14
1.3	Anmerkungen	15
1.4	Relevante Ressourcen.....	18
1.5	Übungen	19
1.6	Literaturangaben	22
2	Überwachtes Lernen	23
2.1	Erlernen einer Klasse anhand von Beispielen	23
2.2	Vapnik-Chervonenkis(VC)-Dimension	29
2.3	PAC-Lernen	30
2.4	Rauschen.....	32
2.5	Erlernen multipler Klassen.....	34
2.6	Regression	36
2.7	Modellauswahl und Generalisierung	39
2.8	Dimensionen eines Algorithmus für überwachtes Lernen ..	43
2.9	Anmerkungen	45
2.10	Übungen	46
2.11	Literaturangaben	50
3	Bayessche Entscheidungstheorie	51
3.1	Einführung	51
3.2	Klassifikation	53
3.3	Verluste und Risiken	55
3.4	Diskriminanzfunktionen	58
3.5	Assoziationsregeln	59
3.6	Anmerkungen	62
3.7	Übungen	63
3.8	Literaturangaben	67

4	Parametrische Methoden	69
4.1	Einführung	69
4.2	Maximum-Likelihood-Schätzung	70
4.2.1	Bernoulli-Verteilung	71
4.2.2	Multinomiale Dichte	72
4.2.3	Gauß-Verteilung (Normalverteilung)	72
4.3	Bewertung eines Schätzers: Verzerrung und Varianz	73
4.4	Der Bayessche Schätzer	75
4.5	Parametrische Klassifikation	78
4.6	Regression	82
4.7	Das Verzerrung/Varianz-Dilemma	85
4.8	Modellauswahl	88
4.9	Anmerkungen	93
4.10	Übungen	94
4.11	Literaturangaben	96
5	Multivariate Methoden	97
5.1	Multivariate Daten	97
5.2	Parameterschätzung	98
5.3	Schätzung von fehlenden Werten	99
5.4	Multivariate Normalverteilung	100
5.5	Multivariate Klassifikation	104
5.6	Anpassen der Komplexität	110
5.7	Diskrete Merkmale	112
5.8	Multivariate Regression	114
5.9	Anmerkungen	115
5.10	Übungen	116
5.11	Literaturangaben	118
6	Dimensionalitätsreduktion	119
6.1	Einführung	119
6.2	Teilmengenselektion	120
6.3	Hauptkomponentenanalyse	125
6.4	Merkmalseinbettung	131
6.5	Faktorenanalyse	134
6.6	Singulärwertzerlegung und Faktorisierung von Matrizen	139
6.7	Multidimensionale Skalierung	141
6.8	Lineare Diskriminanzanalyse	145
6.9	Kanonische Korrelationsanalyse	149
6.10	Isomap	152
6.11	Lokal lineare Einbettung	154
6.12	Laplacesche Eigenmaps	157
6.13	Anmerkungen	159
6.14	Übungen	161
6.15	Literaturangaben	163

7	Clusteranalyse	165
7.1	Einführung	165
7.2	Mischungsdichten	166
7.3	k -Means-Clusteranalyse	167
7.4	Expectation-Maximization-Algorithmus	171
7.5	Mischungsmodelle mit verborgenen Variablen	177
7.6	Überwachtes Lernen nach einer Clusteranalyse	178
7.7	Spektrale Clusteranalyse	179
7.8	Hierarchische Clusteranalyse	181
7.9	Auswahl der Anzahl an Clustern	183
7.10	Anmerkungen	184
7.11	Übungen	185
7.12	Literaturangaben	187
8	Nichtparametrische Methoden	189
8.1	Einführung	189
8.2	Nichtparametrische Dichteschätzung	191
8.2.1	Histogrammschätzer	191
8.2.2	Kernel-Schätzer	193
8.2.3	k -Nächste-Nachbarn-Schätzer	195
8.3	Verallgemeinerung auf multivariate Daten	196
8.4	Nichtparametrische Klassifikation	197
8.5	Verdichtete Nächste-Nachbarn-Methode	198
8.6	Abstands-basierte Klassifikation	200
8.7	Ausreißererkennung	203
8.8	Nichtparametrische Regression: Glättungsmodelle	205
8.8.1	Gleitende Mittelwertglättung	206
8.8.2	Glättung durch Kernel-Funktion	208
8.8.3	Gleitende Linienglättung	208
8.9	Wahl des Glättungsparameters	209
8.10	Anmerkungen	210
8.11	Übungen	213
8.12	Literaturangaben	215
9	Entscheidungsbäume	217
9.1	Einführung	217
9.2	Univariate Bäume	219
9.2.1	Klassifikationsbäume	220
9.2.2	Regressionsbäume	225
9.3	Pruning	227
9.4	Regelextraktion aus Bäumen	229
9.5	Lernen von Regeln anhand von Daten	231
9.6	Multivariate Bäume	235
9.7	Anmerkungen	237
9.8	Übungen	240
9.9	Literaturangaben	242

10	Lineare Diskriminanz	245
10.1	Einführung	245
10.2	Generalisierung des linearen Modells	247
10.3	Geometrie der linearen Diskriminanz	248
10.3.1	Zwei Klassen	248
10.3.2	Multiple Klassen	250
10.4	Paarweise Trennung	251
10.5	Neubetrachtung der parametrischen Diskriminanz	253
10.6	Gradientenabstieg	255
10.7	Logistische Diskriminanz	256
10.7.1	Zwei Klassen	256
10.7.2	Multiple Klassen	259
10.8	Diskriminanz durch Regression	265
10.9	Lernen von Rangordnungen	266
10.10	Anmerkungen	269
10.11	Übungen	270
10.12	Literaturangaben	273
11	Mehrlagige Perzeptronen	275
11.1	Einführung	275
11.1.1	Das Gehirn verstehen	276
11.1.2	Neuronale Netze und Parallelverarbeitung	277
11.2	Das Perzeptron	279
11.3	Training eines Perzeptrons	283
11.4	Lernen von Booleschen Funktionen	286
11.5	Mehrlagige Perzeptronen	287
11.6	Das MLP als universelle Näherungsfunktion	290
11.7	Backpropagation-Algorithmus	291
11.7.1	Nichtlineare Regression	292
11.7.2	Zweiklassendiskriminanz	296
11.7.3	Diskriminanz bei multiplen Klassen	297
11.7.4	Multiple verborgene Schichten	297
11.8	Trainingsprozeduren	298
11.8.1	Verbesserung der Konvergenz	298
11.8.2	Übertraining	299
11.8.3	Strukturieren des Netzes	301
11.8.4	Hinweise	303
11.9	Anpassung der Netzgröße	305
11.10	Bayessche Betrachtungsweise des Lernens	308
11.11	Dimensionalitätsreduktion	309
11.12	Lernen mit Zeitreihen	312
11.12.1	Time Delay Neural Networks	313
11.12.2	Rekurrente Netze	313
11.13	Tiefes Lernen	315
11.14	Anmerkungen	317
11.15	Übungen	319
11.16	Literaturangaben	322

12	Lokale Modelle	327
12.1	Einführung	327
12.2	Kompetitives Lernen	327
12.2.1	Online k-Means-Algorithmus	328
12.2.2	Adaptive Resonanztheorie	333
12.2.3	Selbstorganisierende Merkmalskarten	334
12.3	Radiale Basisfunktionen	336
12.4	Einbindung von regelbasiertem Wissen	342
12.5	Normalisierte Basisfunktionen	343
12.6	Kompetitive Basisfunktionen	345
12.7	Lernen mit Vektorquantisierung	348
12.8	Gemischte Expertensysteme	349
12.8.1	Kooperative Expertensysteme	351
12.8.2	Kompetitive Expertensysteme	352
12.9	Hierarchisch gemischte Expertensysteme	353
12.10	Anmerkungen	354
12.11	Übungen	355
12.12	Literaturangaben	358
13	Kernel-Maschinen	361
13.1	Einführung	361
13.2	Die optimal trennende Hyperebene	363
13.3	Der nicht trennbare Fall: Soft-Margin-Trennebenen	367
13.4	ν -SVM	370
13.5	Kernel-Trick	371
13.6	Vektorielle Kernel	373
13.7	Definition von Kernen	375
13.8	Multiple-Kernel-Lernen	377
13.9	Mehrklassen-Kernel-Maschinen	379
13.10	Kernel-Maschinen und Regression	380
13.11	Kernel-Maschinen und Ranking	385
13.12	Einklassen-Kernel-Maschinen	386
13.13	Breiter-Margin-Nächster-Nachbar-Klassifikator	389
13.14	Dimensionalitätsreduktion mit Kernel	391
13.15	Anmerkungen	393
13.16	Übungen	394
13.17	Literaturangaben	396
14	Graphenmodelle	399
14.1	Einführung	399
14.2	Kanonische Fälle für bedingte Unabhängigkeit	401
14.3	Generative Modelle	409
14.4	d-Separation	411
14.5	Belief-Propagation	412
14.5.1	Ketten	412
14.5.2	Bäume	414
14.5.3	Mehrfachbäume	416

14.5.4	Verbindungsbäume	418
14.6	Ungerichtete Graphen: Markovsche Zufallsfelder	419
14.7	Lernen der Struktur eines Graphenmodells	422
14.8	Einflussdiagramme	423
14.9	Anmerkungen	424
14.10	Übungen	425
14.11	Literaturangaben	428
15	Hidden-Markov-Modelle	431
15.1	Einführung	431
15.2	Diskrete Markov-Prozesse	432
15.3	Hidden-Markov-Modelle	435
15.4	Drei grundsätzliche Probleme eines HMM	437
15.5	Evaluierungsproblem	438
15.6	Herausfinden der Zustandssequenz	442
15.7	Lernen von Modellparametern	443
15.8	Kontinuierliche Beobachtungen	446
15.9	Das HMM als Graphenmodell	447
15.10	Modellauswahl im HMM	451
15.11	Anmerkungen	453
15.12	Übungen	455
15.13	Literaturangaben	458
16	Bayessche Schätzung	461
16.1	Einführung	461
16.2	Bayessche Schätzung der Parameter diskreter Verteilungen	465
16.2.1	$K > 2$ -Zustände: Dirichlet-Verteilung	465
16.2.2	$K = 2$ -Zustände: Betaverteilung	467
16.3	Bayessche Schätzung der Parameter einer Gauß-Verteilung	467
16.3.1	Univariater Fall: Unbekannter Mittelwert, bekannte Varianz	467
16.3.2	Univariater Fall: Unbekannter Mittelwert, unbekannte Varianz	470
16.3.3	Multivariater Fall: Unbekannter Mittelwert, unbekannte Kovarianz	472
16.4	Bayessche Schätzung der Parameter einer Funktion	473
16.4.1	Regression	473
16.4.2	Regression mit Prior für die Präzision des Rauschens	477
16.4.3	Der Gebrauch von Basis/Kernel-Funktionen	479
16.4.4	Bayessche Klassifikation	481
16.5	Wahl eines Priors	483
16.6	Bayesscher Modellvergleich	484
16.7	Bayessche Schätzung für ein Mischungsmodell	487
16.8	Nichtparametrische Bayessche Modelle	490
16.9	Gaußsche Prozesse	491

16.10	Dirichlet-Prozesse und Chinaestaurants	495
16.11	Latente Dirichlet-Allokation	496
16.12	Betaprotzesse und indische Büffets	499
16.13	Anmerkungen	500
16.14	Übungen	501
16.15	Literaturangaben	502
17	Kombination mehrerer Lerner	505
17.1	Grundprinzip	505
17.2	Generierung diverser Lerner	506
17.3	Methoden der Modellkombination	509
17.4	Voting	510
17.5	Fehlerkorrekturcodes	514
17.6	Bagging	517
17.7	Boosting	517
17.8	Neubetrachtung der gemischten Expertensysteme	520
17.9	Geschachtelte Generalisierung	522
17.10	Feinabstimmung eines Ensembles	523
17.10.1	Wahl einer Teilmenge des Ensembles	524
17.10.2	Konstruktion von Metalernern	524
17.11	Kaskadierung	525
17.12	Anmerkungen	527
17.13	Übungen	530
17.14	Literaturangaben	532
18	Bestärkendes Lernen	535
18.1	Einführung	535
18.2	Fälle mit einem Zustand: K -armiger Bandit	537
18.3	Elemente des bestärkenden Lernens	538
18.4	Modellbasiertes Lernen	541
18.4.1	Wertiteration	541
18.4.2	Taktikiteration	542
18.5	Lernen mit temporaler Differenz	542
18.5.1	Explorationsstrategien	543
18.5.2	Deterministische Belohnungen und Aktionen	544
18.5.3	Nichtdeterministische Belohnungen und Aktionen	545
18.5.4	Eignungsprotokolle	548
18.6	Generalisierung	549
18.7	Teilweise beobachtbare Zustände	552
18.7.1	Beispiel: Das Tigerproblem	554
18.8	Anmerkungen	559
18.9	Übungen	560
18.10	Literaturangaben	563
19	Experimente mit maschinellern Lernen	565
19.1	Einführung	565
19.2	Faktoren, Antwort und Strategie beim Experimentieren ..	569

19.3	Antwortflächenmethode	571
19.4	Randomisieren, Wiederholen und Blocken	572
19.5	Richtlinien für Experimente mit maschinellem Lernen	573
19.6	Kreuzvalidierung und Resampling-Methoden	577
19.6.1	K -fache Kreuzvalidierung	578
19.6.2	5×2 -Kreuzvalidierung	579
19.6.3	Bootstrapping	580
19.7	Leistungsmessung für Klassifikatoren	580
19.8	Intervallschätzung	584
19.9	Hypothesenprüfung	587
19.10	Bewertung der Leistungsfähigkeit von Klassifikationsal-	
	gorithmen	590
19.10.1	Binomialtest	590
19.10.2	Test der approximierten Normalverteilung	591
19.10.3	t -Test	591
19.11	Vergleich von zwei Klassifikationsalgorithmen	592
19.11.1	Der McNemarsche Test	592
19.11.2	Gepaarter t -Test mit K -facher Kreuzvalidierung	592
19.11.3	Gepaarter t -Test mit 5×2 Kreuzvalidierung	593
19.11.4	Gepaarter F -Test mit 5×2 Kreuzvalidierung	594
19.12	Vergleich mehrerer Algorithmen: Varianzanalyse	595
19.13	Vergleich über mehrere Datenmengen	600
19.13.1	Vergleich zweier Algorithmen	601
19.13.2	Mehrere Algorithmen	603
19.14	Multivariate Tests	604
19.14.1	Vergleich zweier Algorithmen	605
19.14.2	Vergleich mehrerer Algorithmen	606
19.15	Anmerkungen	607
19.16	Übungen	609
19.17	Literaturangaben	610
Anhang		613
A.1	Elemente der Wahrscheinlichkeit	613
A.1.1	Axiome der Wahrscheinlichkeit	614
A.1.2	Bedingte Wahrscheinlichkeit	614
A.2	Zufallsvariablen	615
A.2.1	Funktionen der Wahrscheinlichkeitsverteilung und Wahr-	
	scheinlichkeitsdichte	615
A.2.2	Gemeinsame Verteilungs- und Dichtefunktionen	616
A.2.3	Bedingte Verteilungen	617
A.2.4	Satz von Bayes	617
A.2.5	Erwartung	618
A.2.6	Varianz	618
A.2.7	Das schwache Gesetz großer Zahlen	619
A.3	Spezielle Zufallsvariablen	620
A.3.1	Bernoulli-Verteilung	620
A.3.2	Binomialverteilung	620

A.3.3	Multinomiale Verteilung	620
A.3.4	Gleichverteilung	621
A.3.5	Normalverteilung (Gauß-Verteilung)	621
A.3.6	Chi-Quadrat-Verteilung.....	623
A.3.7	t -Verteilung	623
A.3.8	F -Verteilung	624
A.4	Literaturangaben	624
Index		625

