

Statistische Methoden der Datenanalyse

Vorlesung im Sommersemester 2010

H. Kolanoski

Humboldt-Universität zu Berlin

Inhaltsverzeichnis

Literaturverzeichnis	iv
1 Grundlagen der Statistik	3
1.1 Wahrscheinlichkeit	3
1.1.1 Definition über die Häufigkeit	3
1.1.2 Kombinatorische Definition	3
1.1.3 Axiomatische Definition der Wahrscheinlichkeit	4
1.2 Verteilungen von Zufallsvariablen	7
1.2.1 Eigenschaften von Verteilungen	8
1.2.2 Erwartungswerte	11
1.2.3 Wahrscheinlichster Wert und Median	12
1.2.4 Stichproben und Schätzwerte	13
1.3 Simulation von Verteilungen	14
1.3.1 Umkehrung der Verteilungsfunktion	14
1.3.2 ‘Hit and Miss’ Methode	16
2 Spezielle Verteilungen einer Variablen	19
2.1 Binomial-Verteilung	19
2.2 Multinomial-Verteilung	24
2.3 Poisson-Verteilung	25
2.4 Gleichverteilung	27
2.5 Normalverteilung	30
2.5.1 Vertrauensintervalle:	32
2.6 Zentraler Grenzwertsatz	34
3 Verteilungen mehrerer Variablen	37
3.1 Eigenschaften von Verteilungen mehrerer Variablen	37
3.1.1 Wahrscheinlichkeitsdichte, Verteilungsfunktion, Randverteilung	37
3.1.2 Bedingte Wahrscheinlichkeitsdichten, Selektionsschnitte	38
3.2 Erwartungswerte	39
3.3 Kovarianzmatrix	40
3.3.1 Definition und Eigenschaften der Kovarianzmatrix	40
3.3.2 Beispiel: Multi-dimensionale Gaussverteilung	40
3.3.3 Kovarianzen von Stichproben	41
3.3.4 Kovarianzmatrix von unabhängigen Variablen	41
3.3.5 Korrelationen	42
3.4 Lineare Funktionen von mehreren Zufallsvariablen	44

3.5	Nicht-lineare Funktionen von Zufallsvariablen	46
3.5.1	Eine Funktion von einem Satz von Zufallsvariablen	46
3.5.2	Mehrere Funktionen von einem Satz von Zufallszahlen	47
3.6	Transformationen von Zufallsvariablen	50
4	Stichproben und Schätzungen	53
4.1	Stichproben, Verteilungen und Schätzwerte	53
4.2	Eigenschaften von Schätzwerten	54
4.3	Stichproben aus Normalverteilungen; χ^2 -Verteilung	56
5	Monte-Carlo-Methoden	63
5.1	Einführung	63
5.2	Zufallszahlengeneratoren	65
5.2.1	Multiplikativ kongruentielle Generatoren	65
5.2.2	Mersenne-Twister	66
5.2.3	Quasi-Zufallszahlen	66
5.3	Monte-Carlo-Erzeugung von Ereignissen	67
5.3.1	Inversionsmethode	67
5.3.2	‘Hit-and-Miss’-Verfahren	70
5.3.3	Majorantenmethode	70
5.3.4	Wichtung der Ereignisse	71
5.4	Monte-Carlo-Integration	72
5.4.1	Majoranten-Methode mit Hit-or-Miss	72
5.4.2	MC-Integration mit Ereigniswichtung	73
5.4.3	Varianz-reduzierende Verfahren	73
5.4.4	Stratified Sampling (‘Geschichtete Stichproben’)	74
6	Die Maximum-Likelihood-Methode	77
6.1	Das Maximum-Likelihood-Prinzip	77
6.2	ML-Methode für Histogramme	80
6.3	Berücksichtigung von Zwangsbedingungen	81
6.3.1	Methode der Lagrange-Multiplikatoren	82
6.3.2	Zwangsbedingungen als Zufallsverteilungen	82
6.3.3	Erweiterte ML-Methode	83
6.3.4	Freiheitsgrade und Zwangsbedingungen	84
6.4	Fehlerbestimmung für ML-Schätzungen	85
6.4.1	Allgemeine Methoden der Varianzabschätzung	85
6.4.2	Varianzabschätzung durch Entwicklung um das Maximum	86
6.4.3	Vertrauensintervalle und Likelihood-Kontouren	86
6.5	Eigenschaften von ML-Schätzungen	88
7	Methode der kleinsten Quadrate	91
7.1	Prinzip der Methode der kleinsten Quadrate	91
7.2	Lineare Anpassung	92
7.2.1	Anpassung der Messwerte an eine Gerade	92
7.2.2	Anpassung einer allgemeinen linearen Funktion der Parameter	95
7.3	Anpassung nicht-linearer Funktionen der Parameter	100

8	Signifikanzanalysen	103
8.1	Einführung	103
8.2	Prüfung von Hypothesen	104
8.2.1	χ^2 -Test	104
8.2.2	Studentsche t-Verteilung	104
8.2.3	F-Verteilung	106
8.2.4	Kolmogorov-Smirnov-Test	108
8.3	Vertrauensintervalle	110
8.3.1	Bayes-Vertrauensintervalle	110
8.3.2	‘Klassische’ Vertrauensintervalle	110
9	Klassifikation und statistisches Lernen	115
9.1	Einführung	115
9.2	Schätzung von Wahrscheinlichkeitsdichten	117
9.3	Lineare Diskriminanten	118
9.3.1	Klassentrennung durch Hyperebenen	118
9.3.2	Fisher-Diskriminante	119
9.4	Neuronale Netze zur Datenklassifikation	120
9.4.1	Einleitung: Neuronale Modelle	120
9.4.2	Natürliche neuronale Netze	122
9.4.3	Künstliche neuronale Netze (KNN)	126
9.4.4	Das einfache Perzeptron	128
9.4.5	Das Mehrlagen-Perzeptron	132
9.4.6	Lernen	136
9.4.7	Typische Anwendungen für Feed-Forward-Netze	141
9.4.8	BP-Lernen und der Bayes-Diskriminator	144
9.5	Entscheidungsbäume	149
9.5.1	Aufwachsen eines Baumes	149
9.5.2	Verstärkte Entscheidungsbäume	151
9.6	Stützvektormaschinen	152
9.6.1	Lineare SVM-Klassifikation	153
9.6.2	Nichtlineare Erweiterung mit Kernelfunktionen	155
10	Entfaltung	159
10.1	Einführung	159
10.2	Entfaltung von Histogrammen	159

