

Inhaltsverzeichnis

1	Einleitung	1
2	Datenstruktur und Notation	3
2.1	Datenstruktur	3
2.1.1	Generelle Nächste Nachbarn	3
2.1.2	Richtungsbezogene Nächste Nachbarn	5
2.1.3	Summen als Prädiktor	7
2.2	Modifikation der Nächste Nachbarn Ermittlung	12
2.2.1	Distanzmaße	12
2.2.2	Standardisierung	13
2.2.3	Bindungsbrechung	14
2.3	Gütemaße	15
2.4	Notation	17
3	Klassifikationsverfahren	19
3.1	Nächste Nachbarn	20
3.1.1	k-Nächste Nachbarn	20
3.1.2	Gewichteter k-Nächste Nachbarn Algorithmus	25
3.2	Lineare Diskriminanzanalyse	29
3.3	Quadratische Diskriminanzanalyse	33
3.4	Logistische Regression	35
3.5	Lasso	39
3.6	Random Forests	45
3.7	Boosting	50
3.7.1	AdaBoost	50
3.7.2	Gradient Boosting	52
3.8	Ensemble-Lasso	56

4	Simulierte Datensätze	61
4.1	mlbench	62
4.2	2 dimensionale Gaußverteilung	66
4.3	2 dimensionale Gaußverteilung mit Störvariablen	69
4.4	Einfaches Klassifikationsproblem	72
4.5	Schwieriges Klassifikationsproblem	76
4.6	Ergebnisübersicht	78
5	Reale Datensätze	83
5.1	Glas Identifikation	84
5.2	Brustkrebs	86
5.3	Ionosphäre	89
5.4	Australische Kreditkartenanträge	92
5.5	Glaukom	94
5.6	Ergebnisübersicht	97
6	Fazit	101
Anhang A	Fluch der Dimensionen	105
Anhang B	Zusatzgrafiken der simulierten Datensätze	107
B.1	(mlbench) - Gütemaße	107
B.2	(mlbench) - Variablenwichtigkeit	109
B.3	(mlbench) - Wilcoxon Vorzeichen Rang Tests	112
B.4	(HT1) - Gütemaße	115
B.5	(HT1) - Variablenwichtigkeit	117
B.6	(HT1) - Wilcoxon Vorzeichen Rang Tests	120
B.7	(HT2) - Gütemaße	123
B.8	(HT2) - Variablenwichtigkeit	125
B.9	(HT2) - Wilcoxon Vorzeichen Rang Tests	132
B.10	(easy) - Gütemaße	135
B.11	(easy) - Variablenwichtigkeit	137
B.12	(easy) - Wilcoxon Vorzeichen Rang Tests	144
B.13	(difficult) - Gütemaße	147
B.14	(difficult) - Variablenwichtigkeit	149
B.15	(difficult) - Wilcoxon Vorzeichen Rang Tests	156
Anhang C	Zusatzgrafiken der realen Datensätze	159
C.1	(Glas Identifikation) - Gütemaße	159
C.2	(Glas Identifikation) - Variablenwichtigkeit	161
C.3	(Glas Identifikation) - Wilcoxon Vorzeichen Rang Tests	168

C.4 (Brustkrebs) - Gütemaße	171
C.5 (Brustkrebs) - Variablenwichtigkeit	173
C.6 (Brustkrebs) - Wilcoxon Vorzeichen Rang Tests	180
C.7 (Ionosphäre) - Gütemaße	183
C.8 (Ionosphäre) - Variablenwichtigkeit	185
C.9 (Ionosphäre) - Wilcoxon Vorzeichen Rang Tests	191
C.10 (Kreditkartenanträge) - Gütemaße	194
C.11 (Kreditkartenanträge) - Variablenwichtigkeit	196
C.12 (Kreditkartenanträge) - Wilcoxon Vorzeichen Rang Tests	205
C.13 (Glaukom) - Gütemaße	208
C.14 (Glaukom) - Variablenwichtigkeit	210
C.15 (Glaukom) - Wilcoxon Vorzeichen Rang Tests	217
Literaturverzeichnis	221

Verbesserung von Klassifikationsverfahren
Informationsgehalt der k-Nächsten-Nachbarn nutzen

Koch, D.

2016, XXII, 224 S. 278 Abb., Softcover

ISBN: 978-3-658-11475-6