

Reihe: Quantitative Ökonomie · Band 137

Herausgegeben von Prof. Dr. Eckart Bomsdorf, Köln, Prof. Dr. Wim Kösters, Bochum, und Prof. Dr. Winfried Matthes, Wuppertal

Dr. Stefan Goller

Klassifikationsverfahren auf Basis von Streuungsmaßen

Simulationen zur Beurteilung der
Ordinalisierung metrischskalierter Merkmale



Inhaltsverzeichnis

Abbildungsverzeichnis	xi
Tabellenverzeichnis	xv
Symbolverzeichnis	xix
1 Grundlagen und Problemstellung	1
1.1 Einführung	1
1.2 Merkmalstypen	4
1.3 Merkmalstypabhängige Probleme bei Klassifikationsverfahren	7
1.4 Zielsetzung und Aufbau der Arbeit	10
2 Gütekriterien zur Partitionierung	13
2.1 Präzisierung des Klassifikationsproblems	13
2.2 Maximum-Likelihood-Kriterien für metrischskalierte Merkmale	14
2.2.1 Allgemeine Modellannahmen	14
2.2.2 Varianzkriterien	16
2.2.3 Determinantenkriterien	17
2.2.4 Klassifikationsverhalten der Maximum-Likelihood-Kriterien	18
2.3 Gütekriterien für nominal- und ordinalskalierte Merkmale auf Basis von Streuungsmaßen	19
2.3.1 Zur Eignung von Streuungsmaßen für die Klassifikation	19

2.3.2	Streuungsmaße für nominalskalierte Merkmale	20
2.3.2.1	Die Entropie	20
2.3.2.2	Das Streuungsmaß von GINI	23
2.3.3	Streuungsmaße für ordinalskalierte Merkmale	26
2.3.3.1	Ein Streuungsmaß auf Basis der Entropie	26
2.3.3.2	Ein Streuungsmaß vom Typ GINI	29
2.3.4	Streuungsmessung bei abhängigen Merkmalen	31
2.3.5	Vergleich der Streuungsmaße	34
2.4	Gütekriterien für gemischtskalierte Merkmale auf Basis von Streuungsmaßen	38
3	Numerische Verfahren zur Partitionierung	41
3.1	Zur Notwendigkeit numerischer Verfahren	41
3.2	Ein heuristisches Verfahren	43
3.2.1	Das Austauschverfahren	43
3.2.2	Die Wahl der Startpartition	46
3.3	Exakte Optimierungsverfahren	47
3.3.1	Dynamische Optimierung	47
3.3.1.1	Der allgemeine Fall	47
3.3.1.2	Der FISHER-Algorithmus als Spezialfall	51
3.3.2	Branch-and-Bound	54
3.3.2.1	Der Grundalgorithmus	54
3.3.2.2	Erweiterungen des Grundalgorithmus	59
4	Verfahren zur Ordinalisierung	65
4.1	Grundsätzliche Vorgehensweise	65
4.2	Graphisch explorative Verfahren	67
4.2.1	Punktdichte und Verteilungsdichte	67

4.2.2	Graphische Darstellung der Punktdichte	69
4.2.2.1	Histogramm	69
4.2.2.2	Kerndichteschätzer	70
4.2.2.3	Die Wahl von Intervallbreite und Bandbreite	72
4.2.3	Ordinalisierung unter Verwendung der Punktdichte	76
4.2.3.1	Subjektive Klassenbildung	76
4.2.3.2	Klassen hoher relativer Punktdichte	77
4.2.3.3	Klassen hoher und niedriger absoluter Punktdichte	79
4.3	Eindimensionale Klassifikationsverfahren	81
4.3.1	Ein geeignetes Gütekriterium	81
4.3.2	Hierarchische Klassifikation	82
4.3.3	Optimale Klassifikation	85
4.4	Naive Verfahren	87
4.4.1	Äquidistante Klassen	87
4.4.2	Klassen gleichen Umfangs	90
4.5	Vergleich der Ordinalisierungsverfahren	92
5	Simulationen	95
5.1	Aufbau der Simulationen	95
5.1.1	Vorgehensweise	95
5.1.2	Konstruktion der Datensätze	97
5.1.3	Messung der Ähnlichkeit zweier Partitionen	111
5.2	Ergebnisse der Simulationen	115
5.2.1	Vorbemerkung	115
5.2.2	Ergebnisse für die Datensätze 1 und 2	115
5.2.3	Ergebnisse für die Datensätze 3 und 4	121
5.2.4	Ergebnisse für die Datensätze 5 und 6	125

5.2.5	Ergebnisse für die Datensätze 7 und 8	127
5.2.6	Ergebnisse für die Datensätze 9 und 10	130
5.2.7	Ergebnisse für Datensatz 11	132
5.2.8	Zusammenfassung und Beurteilung der Ergebnisse	133
6	Schlußbemerkungen	141
A	Abbildungen und Tabellen	143
A.1	Ergebnisse für Datensatz 1	144
A.2	Ergebnisse für Datensatz 2	145
A.3	Ergebnisse für Datensatz 3	148
A.4	Ergebnisse für Datensatz 4	151
A.5	Ergebnisse für Datensatz 5	154
A.6	Ergebnisse für Datensatz 6	157
A.7	Ergebnisse für Datensatz 7	160
A.8	Ergebnisse für Datensatz 8	163
A.9	Ergebnisse für Datensatz 9	166
A.10	Ergebnisse für Datensatz 10	169
A.11	Ergebnisse für Datensatz 11	172
	Literaturverzeichnis	175