

Bootstrap in der Diskriminanzanalyse

Paul Eberhard Rudolph
Forschungsinstitut für die Biologie
landwirtschaftlicher Nutztiere
FB Genetik und Biometrie
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
rudolph@fbn-dummerstorf.de

Armin Tuchscherer
Forschungsinstitut für die Biologie
landwirtschaftlicher Nutztiere
FB Genetik und Biometrie
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
atuchsch@fbn-dummerstorf.de

Bernd Jäger
Ernst-Moritz-Arndt-Universität
Greifswald
Institut für Biometrie und Medizinische Informatik
Wilhelm-Stahl-Allee 2
17487 Greifswald
bjaege@biometrie.uni-greifswald.de

Margret Tuchscherer
Forschungsinstitut für die Biologie
landwirtschaftlicher Nutztiere
FB Verhaltensphysiologie
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
mtuchsch@fbn-dummerstorf.de

Zusammenfassung

Das Hauptziel in der Diskriminanzanalyse ist die Zuordnung von Objekten unbekannter Klassenzugehörigkeit zu vorgegebenen Klassen. Dabei sind die Objekte durch einen p-dimensionalen Merkmalsvektor und die Klassen durch eine Menge vorgegebener Objekte mit bekannter Klassenzugehörigkeit, die sogenannte Lernstichprobe, charakterisiert. Eine sinnvolle Klassifizierung von Objekten kann nur dann erfolgen, wenn die Wahrscheinlichkeit der Zuordnung eines Objektes zu einer falschen Klasse möglichst klein ist. Das ist bei kleinen Umfängen der Lernstichprobe jedoch häufig nicht der Fall, s. z. B. Rudolph u.a. (1999) und Tuchscherer u.a. (2002).

In diesem Beitrag wird mittels Simulation untersucht, ob sich durch die Vergrößerung des Umfanges einer vorgegebenen Lernstichprobe mittels Bootstrap-Methoden die Fehlklassifikationswahrscheinlichkeiten verringern lassen. In den mit SAS realisierten Simulationsprogrammen finden die in Tuchscherer u.a. (1999) bzw. Rudolph u.a. (2003) beschriebenen Makros zur Erzeugung multivariater Verteilungen bzw. Bootstrap-Stichproben Verwendung.

Die Ergebnisse der Simulationsuntersuchungen werden vorgestellt und für Anwender von Diskriminanzanalyseverfahren Empfehlungen abgeleitet.

Keywords: Diskriminanzanalyse, Fehlklassifikation, Bootstrap

1 Einleitung

In der Regel ist das Ziel einer Diskriminanzanalyse die Zuordnung von Objekten mit unbekannter Zugehörigkeit zu vorgegebenen Klassen, für die eine Lernstichprobe existiert mit Objekten, deren Klassenzugehörigkeit bekannt ist. Die SAS-Prozedur DISCRIM bietet die Möglichkeit, aufgrund einer Lernstichprobe Objekte mit unbekannter Klassenzugehörigkeit den bekannten Klassen zuzuordnen. Dabei stehen parametrische und nichtparametrische Verfahren zur Verfügung. Eine Möglichkeit zur Auswahl eines geeigneten Verfahrens ist die Betrachtung der Fehlklassifikationswahrscheinlichkeiten bei einer Neuordnung der Objekte der Lernstichprobe durch Reklassifizierung mittels crossvalidation, das die Prozedur DISCRIM anbietet.

Von Rudolph u.a. (1999) und Tuchscherer u.a. (2002) wurden für verschiedene Diskriminanzanalyseverfahren mittels Simulation die Fehlklassifikationswahrscheinlichkeiten für kleine Stichprobenumfänge in der Lernstichprobe im Falle ausgewählter mehrdimensional normal- und nicht-normalverteilter Zufallsgrößen untersucht.

In Anwenderberatungen der Autoren zur Diskriminanzanalyse tauchte wiederholt die Frage auf, ob es möglich ist, die Fehlklassifikationswahrscheinlichkeiten von Diskriminanzanalyseverfahren dadurch zu verringern, dass man die Umfänge der Lernstichproben z. B. mittels Bootstrapverfahren erhöht. Dieser Fragestellung wird in diesem Beitrag nachgegangen. Dabei sind sich die Autoren durchaus darüber im Klaren, dass es eine allgemein gültige Antwort auf diese Frage nicht geben kann, da diese stets von den zugrunde liegenden Daten abhängen wird.

In diesem Beitrag werden mittels SAS realisierter Simulationen Fehlklassifikationswahrscheinlichkeiten bestimmt. Dabei geht es einmal um die durch Reklassifizierung bestimmten Fehlklassifikationswahrscheinlichkeiten für die Lernstichprobe und die mittels Bootstrap vergrößerten Lernstichproben. Andererseits werden ebenfalls die Fehlklassifikationswahrscheinlichkeiten bei der Zuordnung von Objekten (Testdaten) zu vorgegebenen Klassen mittels Lernstichproben (Ausgangslernstichprobe und mittels Bootstrap vergrößerte Lernstichproben) geschätzt.

Die durchgeführten Simulationen beschränken sich auf die Maximum Likelihood Diskriminanzanalyse und die parameterfreie Methode mit $k=3$ nächsten Nachbarn sowie die bereits von Rudolph u.a. (1999) verwendeten multivariaten Normalverteilungen mit gleichen Kovarianzmatrizen in $n_k = 2$ Klassen.

2 Diskriminanzanalysen mit der SAS Prozedur DISCRIM

In der **Diskriminanzanalyse** geht man davon aus, dass eine Gesamtheit (Menge von Objekten) aus $g \geq 2$ Teilgesamtheiten (Klassen, Gruppen) besteht, wobei jedes Element (Objekt) der Gesamtheit genau einer Teilgesamtheit angehört. Über die Objekte der Gesamtheit sei bekannt, dass sie sich nur sinnvoll durch die gleichzeitige Betrachtung von $p > 1$ Merkmalen, die an den Objekten messbar seien, unterscheiden lassen.

Es sollen nun Messwerte der p Merkmale für Objekte, für die bekannt ist, zu welcher Teilgesamtheit sie gehören, aus allen Teilgesamtheiten vorliegen. Das bedeutet, für jedes derartige Objekt sind sein p -dimensionaler Merkmalsvektor $x \in \mathbb{R}^p$ sowie seine Klassenzugehörigkeit k bekannt. Die Menge der Merkmalsvektoren von Objekten mit bekannter Gruppenzugehörigkeit wird häufig auch als **Lernstichprobe** bezeichnet. Der statistische Hintergrund besteht darin, dass man x als Realisierung einer p -dimensionalen Zufallsgröße $\underline{x}' = (x_1, \dots, x_p)$ auffassen kann.

Das Ziel der Diskriminanzanalyse besteht dann in folgender Aufgabe:

Ein Objekt mit dem Beobachtungsvektor x , dessen Klassenzugehörigkeit unbekannt ist (d. h., dessen Klassenindex k unbekannt ist), soll einer der vorliegenden Gruppen zugeordnet werden.

Die Lösung des Zuordnungsproblems (Klassifizierung) erfolgt mit geeigneten Entscheidungsregeln, die in der Regel auf der Bestimmung einer oder mehrerer optimaler **Trennfunktionen (Diskriminanzfunktionen)** basieren. (Die Anzahl der benötigten Trennfunktionen richtet sich nach der Anzahl der Teilgesamtheiten.)

Unterschiedliche Vorgehensweisen führen zu unterschiedlichen Diskriminanzfunktionen und daraus abgeleiteten Zuordnungsregeln, s. z. B. Huberty (1994).

Mit der SAS-Prozedur DISCRIM steht dem Anwender ein umfangreiches Werkzeug zur Durchführung von Diskriminanzanalysen zur Verfügung, insbesondere zur Klassifizierung von Objekten mit unbekannter Klassenzugehörigkeit mittels einer Lernstichprobe und den Beobachtungsvektoren der zu klassifizierenden Objekte. Neben parametrischen Verfahren, die eine multivariate Normalverteilung unterstellen, gibt es in der Prozedur DISCRIM auch parameterfreie Verfahren, die keine spezielle Verteilung der Beobachtungsvektoren unterstellen. Diese parameterfreien Verfahren basieren auf nichtparametrischen Schätzungen der klassenspezifischen Wahrscheinlichkeitsdichten, z. B. auf der Basis der Bestimmung der 'k nächsten Nachbarn'. Einzelheiten dazu findet man in SAS/STAT® User's Guide.

Neben der eigentlichen Klassifizierung von Objekten unbekannter Klassenzugehörigkeit stellt die SAS-Prozedur DISCRIM ein gutes Werkzeug zur Ermittlung eines geeigneten Diskriminanzanalyseverfahrens mit der Reklassifizierung der Elemente der Lernstichprobe mittels crossvalidation bereit. Kleine Fehlklassifikationswahrscheinlichkeiten bei der Reklassifizierung sind ein Indikator für wünschenswerte kleine Fehlklassifikationswahrscheinlichkeiten bei der Klassifizierung von Objekten mit unbekannter Klassenzugehörigkeit.

Bei Verwendung einer parametrischen Methode werden durch crossvalidation mit der Prozedur DISCRIM alle Objekte der Lernstichprobe neu klassifiziert, wobei die dabei verwendeten Diskriminanzfunktionen jeweils unter Ausschluß des zu klassifizierenden Objektes berechnet werden. Bei Verwendung einer parameterfreien Methode wird die zur Berechnung von Abständen verwendete Kovarianzmatrix aus allen Daten der Lernstichprobe geschätzt, während die Daten des jeweils zu klassifizierenden Objektes bei der parameterfreien Dichteschätzung unberücksichtigt bleiben.

Die im Kapitel 3 beschriebenen Simulationen beschränken sich auf die Maximum-Likelihood Diskriminanzanalyse als parametrisches Verfahren und als parameterfreies Verfahren auf die Diskriminanzanalyse nach der Methode der $k = 3$ nächsten Nachbarn.

Für multivariat normalverteilte Lernstichproben (Ausgangslernstichprobe und aus dieser mittels Bootstrap erzeugte neue Lernstichproben mit vergrößertem Umfang) und Testdaten sollen die Fehlklassifikationswahr-

scheinlichkeiten bei crossvalidation der Lernstichproben und bei der Klassifizierung von Testdaten ermittelt werden.

3 Simulationsexperiment

Mit einem SAS-Makro zur Realisierung von (näherungsweise) multivariat normalverteilten Zufallsgrößen mit vorgegebenen Erwartungswertvektoren μ_i und vorgegebenen Kovarianzmatrizen COV_i (vgl. Tuchscherer u. a., 1999) wurden Lernstichproben (LSP3) und Testdaten erzeugt. Als Dimension der Beobachtungsvektoren wurde $p=3$ in 2 Klassen ($k \in \{1,2\}$) mit in den Klassen jeweils gleichem Umfang n_k ($n_k = 10, 20, 50, 100$) gewählt. Für die Anzahl der Testdaten wurde $n = 1000$ je Klasse und simulierter Wiederholung (1000) einer Lernstichprobe festgelegt.

Mit dem in Rudolph u.a. (2003) beschriebenen SAS-Makro zur Erzeugung von Bootstrap-Stichproben wurden aus den ursprünglichen Lernstichproben LSP3 neue Lernstichproben LSP3_b mit jeweils vergrößertem Umfang mit $b = 2, 5, 10$ erzeugt. Dabei gibt b den Faktor an, um den der ursprüngliche Lernstichprobenumfang vergrößert wurde.

Als Erwartungswertvektoren wurden

$$\mu_1 = (3 \quad 7 \quad 5)', \quad \mu_2 = (2 \quad 8 \quad 4)'$$

und die für beide Klassen identische Kovarianzmatrix

$$COV_1 = COV_2 = \begin{pmatrix} 1.8 & 1.3 & 0.5 \\ 1.3 & 2.1 & 1.2 \\ 0.5 & 1.2 & 2.7 \end{pmatrix}$$

gewählt (vgl. Rudolph u.a., 1999).

Mit diesen simulierten Daten wurden die entsprechenden Fehlklassifikationswahrscheinlichkeiten für crossvalidation der Lernstichproben LP3 (pml, pnn) und LP3_b (pml_b, pnn_b) sowie die Fehlklassifikationswahrscheinlichkeiten für die Testdaten unter Verwendung der Lernstichproben LSP3 (psml, psnn) und LSP_b (psml_b, psnn_b) für die folgenden zwei Diskriminanzanalyseverfahren ermittelt:

1. Verfahren:

Parametrisches Verfahren: Maximum Likelihood Diskriminanzanalyse mit der Voraussetzung gleicher Kovarianzmatrizen in beiden Klassen und gleichen a-priori-Wahrscheinlichkeiten (ml),

2. Verfahren:

Parameterfreies Verfahren: Methode der 'k nächsten Nachbarn' für $k=3$ (nn).

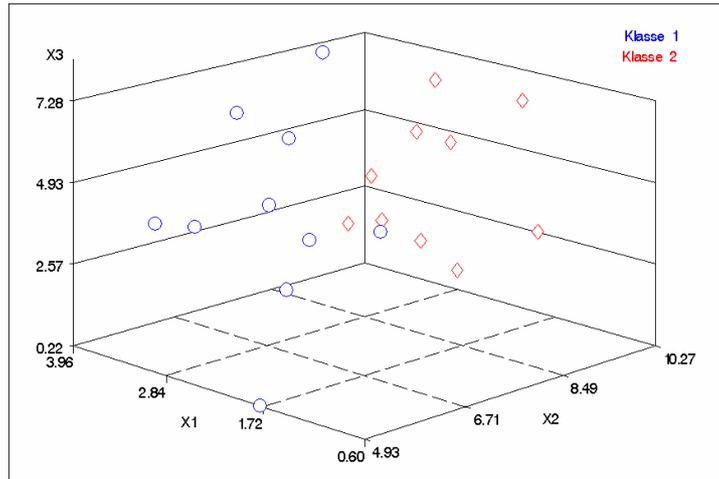


Abb. 1: Lernstichproben mit $n_k = 10$

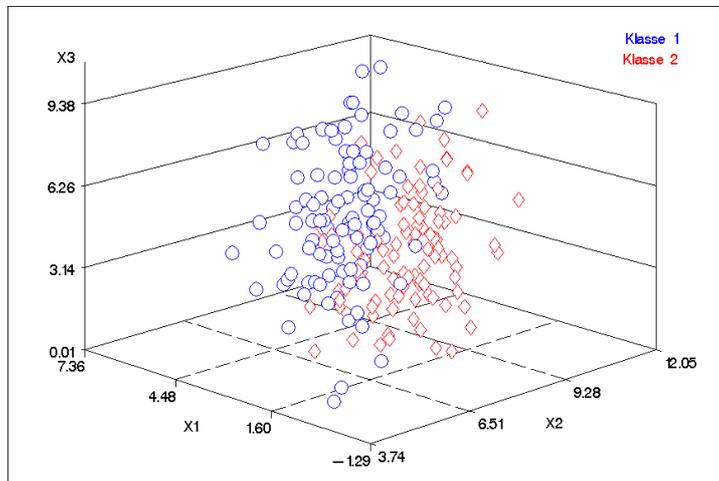


Abb. 2: Lernstichproben mit $n_k = 100$

In den Abbildungen 1 bzw. 2 sind simulierte Lernstichproben für $n_k = 10$ bzw. $n_k = 100$ zur Veranschaulichung der Lage der beiden Klassen dargestellt.

Aus den jeweils 1000 mal wiederholten Bestimmungen der Fehlklassifikationswahrscheinlichkeiten wurden Minima, Maxima und Mittelwerte berechnet. Die Mittelwerte der Fehlklassifikationswahrscheinlichkeiten sind in Kapitel 4 zusammengestellt. Die entsprechenden Tabellen mit Minima und Maxima sind im Anhang zu finden.

Aufgrund der Größe wurde auf eine Wiedergabe des SAS-Programms an dieser Stelle verzichtet. Interessenten können es beim Erstautor anfordern. Für einen Anwender mit SAS-Programmierkenntnissen ist es außerdem sicher einfach, aus dem in Rudolph u.a. (1999) zu findenden Simulationsprogramm und dem in Rudolph u.a. (2003) beschriebenen Makro zur Erzeugung von Bootstrap-Stichproben ein entsprechendes Programm zu erstellen.

4 Ergebnisse und Diskussion

In den folgenden Tabellen 1 und 2 sind die Mittelwerte über jeweils 1000 Wiederholungen der ermittelten Fehlklassifikationswahrscheinlichkeiten zusammengestellt.

Erwartungsgemäß kann zunächst festgestellt werden, dass die mittleren Fehlklassifikationswahrscheinlichkeiten (pml, pnn) bei Reklassifizierung der ursprünglichen Lernstichproben für die Maximum Likelihood Diskriminanzanalyse geringer sind als für Diskriminanzanalysen nach der Methode der (3) nächsten Nachbarn. Das Gleiche gilt für die mittleren Fehlklassifikationswahrscheinlichkeiten (psml, psnn) bei der Klassifizierung von Testdaten mittels der ursprünglichen Lernstichproben. Dies ist damit zu begründen, dass die simulierte(n) Daten (näherungsweise) den Voraussetzungen für die Maximum Likelihood Diskriminanzanalyse genügen (multivariate Normalverteilung).

Es ist außerdem zu sehen, dass sich die zueinander gehörenden mittleren Fehlklassifikationswahrscheinlichkeiten pml und psml bzw. pnn und psnn kaum voneinander unterscheiden.

Ein höherer Umfang der Lernstichprobe führt offensichtlich bei beiden Diskriminanzanalyseverfahren auch zu niedrigeren mittleren Fehlklassifikationswahrscheinlichkeiten pml und psml bzw. pnn und psnn.

Tabelle 1: Mittelwerte aus 1000 Wiederholungen für die Methode ml

k	nk	b	pml	psml	pml_b	psml_b
1	10	2	0.166	0.162	0.121	0.167
2	10	2	0.166	0.159	0.130	0.165
1	10	5	0.154	0.160	0.104	0.163
2	10	5	0.161	0.161	0.103	0.165
1	10	10	0.164	0.160	0.102	0.162
2	10	10	0.165	0.161	0.101	0.163
1	20	2	0.147	0.145	0.130	0.147
2	20	2	0.147	0.143	0.127	0.147
1	20	5	0.142	0.144	0.116	0.145
2	20	5	0.144	0.145	0.117	0.148
1	20	10	0.144	0.142	0.116	0.143
2	20	10	0.145	0.147	0.118	0.148
1	50	2	0.137	0.135	0.131	0.136
2	50	2	0.134	0.136	0.127	0.137
1	50	5	0.136	0.135	0.126	0.136
2	50	5	0.137	0.135	0.127	0.136
1	50	10	0.136	0.136	0.126	0.136
2	50	10	0.132	0.135	0.121	0.135
1	100	2	0.132	0.134	0.128	0.134
2	100	2	0.131	0.132	0.128	0.133
1	100	5	0.133	0.132	0.129	0.132
2	100	5	0.131	0.133	0.126	0.133
1	100	10	0.133	0.133	0.127	0.134
2	100	10	0.133	0.133	0.128	0.133

Tabelle 2: Mittelwerte aus 1000 Wiederholungen für die Methode nn

k	nk	b	pnn	psnn	pnn_b	psnn_b
1	10	2	0.187	0.187	0.124	0.247
2	10	2	0.190	0.184	0.129	0.243
1	10	5	0.179	0.183	0.022	0.212
2	10	5	0.184	0.188	0.023	0.215
1	10	10	0.191	0.184	0.000	0.213
2	10	10	0.192	0.188	0.000	0.216
1	20	2	0.178	0.176	0.115	0.228
2	20	2	0.177	0.174	0.121	0.228
1	20	5	0.178	0.174	0.023	0.205
2	20	5	0.176	0.175	0.024	0.205
1	20	10	0.176	0.170	0.000	0.201
2	20	10	0.178	0.176	0.000	0.208
1	50	2	0.168	0.165	0.110	0.213
2	50	2	0.164	0.165	0.109	0.216
1	50	5	0.168	0.166	0.022	0.197
2	50	5	0.167	0.165	0.021	0.197
1	50	10	0.165	0.165	0.000	0.194
2	50	10	0.163	0.165	0.000	0.195
1	100	2	0.162	0.162	0.108	0.208
2	100	2	0.161	0.162	0.106	0.209
1	100	5	0.163	0.160	0.021	0.192
2	100	5	0.163	0.163	0.022	0.194
1	100	10	0.164	0.163	0.000	0.193
2	100	10	0.164	0.162	0.000	0.193

Bei der Vergrößerung der Umfänge der Lernstichproben mittels Bootstrap wäre eigentlich zu erwarten, dass die mittleren Fehlklassifikationswahrscheinlichkeiten p_{ml_b} bzw. p_{nn_b} in etwa gleich bleiben, da sich die Elemente der erweiterten Lernstichprobe nur aus den Elementen der ursprünglichen Lernstichprobe rekrutieren, d. h., Elemente der ursprünglichen Lernstichprobe tauchen in der vergrößerten Lernstichprobe mehrfach auf. Entgegen dieser Vermutung verringern sich die mittleren Fehlklassifikationswahrscheinlichkeiten p_{ml_b} bzw. p_{nn_b} mit wachsendem Bootstrapfaktor b . Diese Verringerung ist bei den Diskriminanzanalysen nach der Methode der (3) nächsten Nachbarn wesentlich stärker als bei den Maximum Likelihood Diskriminanzanalysen. Das ist damit zu begründen, dass eventuell das am besten trennende Element in der vergrößerten Lernstichprobe gleich mehrfach vorhanden ist.

Im Gegensatz zur Verringerung der mittleren Fehlklassifikationswahrscheinlichkeiten p_{ml_b} bzw. p_{nn_b} vergrößern sich die mittleren Fehlklassifikationswahrscheinlichkeiten p_{sml_b} bzw. p_{snn_b} . Während diese Vergrößerung bei den Maximum Likelihood Diskriminanzanalysen nur gering ausfällt, ist sie bei den Diskriminanzanalysen nach der Methode der (3) nächsten Nachbarn wesentlich deutlicher.

Zusammenfassend kann folgendes festgestellt werden:

Die Vergrößerung von Lernstichproben in der Diskriminanzanalyse mittels Bootstrap kann zu einer Verringerung der Fehlklassifikationswahrscheinlichkeiten bei Reklassifizierung der erweiterten Lernstichproben führen. Diese Fehlklassifikationswahrscheinlichkeiten sind aber dann nicht notwendig ein Maß für Fehlklassifikationswahrscheinlichkeiten bei der Klassifizierung von Testdaten. Diese Fehlklassifikationswahrscheinlichkeiten können sogar größer sein als bei Verwendung der ursprünglichen Lernstichprobe. Deshalb kann die Anwendung von Bootstrapmethoden zur Vergrößerung der Lernstichproben in Diskriminanzanalysen nicht empfohlen werden.

Literatur

- [1] Huberty, C. J.(1994): Applied Discriminant Analysis, John Wiley & Sons. Inc., New York.
- [2] Rudolph, P. E. u.a.(1999): Beurteilung von Diskriminanzanalyseverfahren in und mit SAS. In: Proceedings der 3. Konferenz der SAS-Anwender in Forschung und Entwicklung, Ed. Ortseifen, C. : Ruprecht-Karls-Univ. Heidelberg, S. 245-258.
- [3] Rudolph, P. E.; Tuchscherer, A.; Jäger, B.; Tuchscherer, M. (2003): Bootstrap mit SAS. In: Proceedings der 7. Konferenz der SAS-Anwender in Forschung und Entwicklung, Shaker Verlag, Aachen.
- [4] SAS Institute Inc. (1999): SAS/STAT® User's Guide, Version 8, Cary, NC: SAS Institute Inc.
- [5] Tuchscherer, A.; Rudolph, P. E.; Jäger, B.; Tuchscherer, M. (1999). Ein SAS-Makro zur Erzeugung multivariat normalverteilter Zufallsgrößen. In: Proceedings der 3. Konferenz der SAS-Anwender in Forschung und Entwicklung, Ed. Ortseifen, C. : Ruprecht-Karls-Univ. Heidelberg, S. 293-306.
- [6] Tuchscherer, A.; Rudolph, P. E.; Jäger, B.; Tuchscherer, M. (2002). Bewertung von Diskriminanzanalyseverfahren in SAS bei Nicht-normalität. In: Proceedings der 6. Konferenz der SAS-Anwender in Forschung und Entwicklung, Eds. Christmann, A.;Weihs, C., Dortmund, 315-325.

Anhang: Tabelle 3: Minima aus 1000 Wiederholungen für die Methode ml

k	nk	b	pml	psml	pml_b	psml_b
1	10	2	0.00	0.035	0.00	0.038
2	10	2	0.00	0.039	0.00	0.038
1	10	5	0.00	0.028	0.00	0.027
2	10	5	0.00	0.043	0.00	0.041
1	10	10	0.00	0.035	0.00	0.033
2	10	10	0.00	0.036	0.00	0.044
1	20	2	0.00	0.059	0.00	0.052
2	20	2	0.00	0.041	0.00	0.031
1	20	5	0.00	0.043	0.00	0.038
2	20	5	0.00	0.047	0.00	0.046
1	20	10	0.00	0.043	0.00	0.042
2	20	10	0.00	0.053	0.00	0.046
1	50	2	0.00	0.073	0.00	0.069
2	50	2	0.02	0.064	0.00	0.065
1	50	5	0.00	0.065	0.00	0.067
2	50	5	0.02	0.075	0.02	0.069
1	50	10	0.00	0.055	0.00	0.065
2	50	10	0.04	0.051	0.00	0.054
1	100	2	0.04	0.074	0.04	0.075
2	100	2	0.05	0.083	0.04	0.080
1	100	5	0.05	0.078	0.03	0.081
2	100	5	0.04	0.073	0.03	0.069
1	100	10	0.06	0.090	0.05	0.088
2	100	10	0.05	0.085	0.04	0.082

Tabelle 4: Minima aus 1000 Wiederholungen für die Methode nn

k	nk	b	pnn	psnn	pnn_b	psnn_b
1	10	2	0.00	0.030	0.00	0.029
2	10	2	0.00	0.034	0.00	0.041
1	10	5	0.00	0.032	0.00	0.029
2	10	5	0.00	0.045	0.00	0.043
1	10	10	0.00	0.031	0.00	0.025
2	10	10	0.00	0.030	0.00	0.040
1	20	2	0.00	0.043	0.00	0.032
2	20	2	0.00	0.049	0.00	0.049
1	20	5	0.00	0.051	0.00	0.063
2	20	5	0.00	0.049	0.00	0.061
1	20	10	0.00	0.037	0.00	0.064
2	20	10	0.00	0.051	0.00	0.044
1	50	2	0.02	0.078	0.00	0.099
2	50	2	0.02	0.081	0.00	0.085
1	50	5	0.02	0.080	0.00	0.082
2	50	5	0.00	0.077	0.00	0.095
1	50	10	0.02	0.049	0.00	0.070
2	50	10	0.04	0.075	0.00	0.097
1	100	2	0.04	0.091	0.02	0.119
2	100	2	0.04	0.090	0.02	0.122
1	100	5	0.06	0.099	0.00	0.114
2	100	5	0.04	0.100	0.00	0.111
1	100	10	0.07	0.080	0.00	0.118
2	100	10	0.03	0.086	0.00	0.117

Tabelle 5: Maxima aus 1000 Wiederholungen für die Methode ml

k	nk	b	pml	psml	pml_b	psml_b
1	10	2	0.60	0.467	0.600	0.578
2	10	2	0.60	0.490	0.550	0.469
1	10	5	0.60	0.424	0.460	0.477
2	10	5	0.50	0.457	0.460	0.445
1	10	10	0.60	0.473	0.400	0.517
2	10	10	0.60	0.430	0.440	0.418
1	20	2	0.40	0.290	0.475	0.337
2	20	2	0.40	0.287	0.475	0.306
1	20	5	0.40	0.326	0.420	0.312
2	20	5	0.35	0.334	0.350	0.332
1	20	10	0.45	0.304	0.395	0.310
2	20	10	0.45	0.286	0.350	0.303
1	50	2	0.30	0.231	0.310	0.250
2	50	2	0.30	0.224	0.320	0.224
1	50	5	0.32	0.218	0.328	0.224
2	50	5	0.28	0.248	0.284	0.244
1	50	10	0.30	0.257	0.322	0.249
2	50	10	0.30	0.238	0.328	0.231
1	100	2	0.25	0.204	0.240	0.200
2	100	2	0.23	0.187	0.255	0.193
1	100	5	0.23	0.204	0.234	0.200
2	100	5	0.22	0.201	0.234	0.206
1	100	10	0.23	0.200	0.223	0.203
2	100	10	0.25	0.203	0.245	0.215

Tabelle 6: Maxima aus 1000 Wiederholungen für die Methode nn

K	nk	b	pnn	psnn	pnn_b	psnn_b
1	10	2	0.80	0.548	0.650	0.722
2	10	2	0.70	0.532	0.900	0.676
1	10	5	0.70	0.461	0.180	0.547
2	10	5	0.70	0.637	0.180	0.566
1	10	10	0.70	0.492	0.030	0.506
2	10	10	0.70	0.506	0.030	0.536
1	20	2	0.55	0.391	0.425	0.559
2	20	2	0.65	0.361	0.350	0.563
1	20	5	0.55	0.363	0.140	0.399
2	20	5	0.50	0.441	0.130	0.408
1	20	10	0.55	0.381	0.015	0.388
2	20	10	0.55	0.408	0.030	0.420
1	50	2	0.40	0.306	0.280	0.387
2	50	2	0.32	0.290	0.290	0.353
1	50	5	0.38	0.289	0.092	0.333
2	50	5	0.38	0.288	0.068	0.361
1	50	10	0.34	0.313	0.010	0.323
2	50	10	0.38	0.281	0.006	0.328
1	100	2	0.28	0.261	0.210	0.342
2	100	2	0.30	0.252	0.225	0.307
1	100	5	0.31	0.252	0.064	0.275
2	100	5	0.31	0.263	0.070	0.292
1	100	10	0.30	0.231	0.007	0.292
2	100	10	0.31	0.247	0.009	0.270

