

**EIN ALGORITHMUS UND
EIN C++-PROGRAMM ZUR
NICHTPARAMETRISCHEN
DISKRIMINANZANALYSE**

Andreas Möltner

Labor für Klinische Psychophysiologie
Otto-Selz-Institut für Psychologie und Erziehungswissenschaft
Universität Mannheim

Forschungsbericht aus dem Otto-Selz-Institut Nr. 39
1998

Forschungsberichte aus dem Otto-Selz-Institut für Psychologie und Erziehungswissenschaft der Universität Mannheim
Herausgegeben vom geschäftsführenden Direktor des Otto-Selz-Instituts

ISSN 0931-1394

1 Einführung

Das Ziel diskriminanzanalytischer Verfahren ist, aus einer Reihe von bekannten Daten eines Objekts auf deren Zugehörigkeit zu einer Gruppe zu schließen. Beispielsweise seien Personen aufgrund medizinischer Meßwerte als „gesund“ oder „krank“ zu klassifizieren. Zur Durchführung einer Diskriminanzanalyse (DA) muß eine Stichprobe vorliegen, von der sowohl die Daten (Meßwerte) als auch die Gruppenzugehörigkeit bekannt sind. Die Aufgabe der DA ist es nun, eine allgemeine Zuordnungsregel (eine Funktion) zu finden, mit der mit einer möglichst geringen Fehlerwahrscheinlichkeit in der Population aus den Daten die Gruppe bestimmt werden kann.

Der bekannteste Ansatz hierzu ist die lineare Diskriminanzanalyse nach Fisher, in der angenommen wird, jede Gruppe sei durch eine multivariate Normalverteilung mit für alle Gruppen identischer Varianz-Kovarianz-Matrix charakterisiert und die Gruppen unterscheiden sich allein durch ihre Mittelwertsvektoren. Diese Voraussetzung ist das wesentliche Problem der linearen DA nach Fisher, da tatsächliche Meßwerte nur sehr selten auch nur approximativ durch multivariate Normalverteilungen beschreibbar sind. Im Gegensatz zu vielen statistischen Testverfahren, für deren Gültigkeit die Voraussetzung einer wenigstens angenäherten Normalverteilung bei einer hinreichend großen Datenzahl durch den zentralen Grenzwertsatz garantiert ist (z.B. t-Test, multivariate Varianzanalysen), ist bei der linearen DA diese Voraussetzung für die *Rohdaten* wesentlich.

Alternativen zur linearen DA sind seit langen bekannt und praktisch im Einsatz. Einerseits existieren parametrische Verfahren, die statt von Normalverteilungen von anderen Verteilungen ausgehen (z.B. Exponentialverteilungen), andererseits gibt es nicht-parametrische Methoden, in denen überhaupt keine Annahmen über die zugrundeliegenden Verteilungen gemacht werden, sondern explizit (bei den Kernel-Verfahren) oder implizit (K-Nearest-Neighbor-Verfahren) eine Verteilungsschätzung aus der Stichprobe vorgenommen wird. Für die parametrischen Methoden gilt jedoch in der medizinischen oder psychologischen Praxis das gleiche wie für die lineare DA nach Fisher: Meist lassen sich a priori kaum Aussagen über die Verteilungen treffen und im multivariaten Fall auch kaum Transformationen zu handhabbaren Verteilungsklassen finden. Nachteil der nicht-parametrischen Verfahren ist die mit der Variablenzahl exponentiell wachsende Stichprobengröße, die für eine stabile Verteilungsschätzung und einer daraus zu gewinnenden Zuordnungsfunktion notwendig ist.

Eine in Bezug auf die Verteilungsvoraussetzungen „zwischen“ den etablierten parametrischen und nicht-parametrischen Verfahren liegende Klasse von Diskriminanzanalysen stellt die im Folgenden dargestellte Methode einer „nicht-metrischen Diskriminanzanalyse“ dar, in der zwar keine impliziten oder expliziten Annahmen über die genaue Zugehörigkeit der Verteilungen zu bestimmten parametrisierbaren Verteilungsklassen gemacht werden, jedoch die Zuordnungsregel „Daten→Gruppe“ einer wichtigen und inhaltlich häufig plausiblen Einschränkung unterliegt: Werden zwei Objekte mit den Meßwerten (x_1, \dots, x_n) und (y_1, \dots, y_n) derselben Gruppe zugeordnet, so wird auch ein Objekt, dessen Meßwerte aus den Mittelwerten $((x_1 + y_1)/2, \dots, (x_n + y_n)/2)$ bestehen, dieser Gruppe

zugeordnet. Bislang hat sich ein solches Verfahren in der Praxis trotz seiner für medizinische oder psychologische Fragestellungen sehr attraktiven Eigenschaften noch nicht etablieren können. Ein wichtiger Grund hierfür ist, daß die numerischen Algorithmen zur Durchführung einer solchen DA bisher zu langsam sind, um auch auf mittlere Stichprobengrößen (ab 50) anwendbar zu sein. Hierzu ist jedoch anzumerken, daß in der Literatur bislang keine expliziten Resultate zur numerischen Durchführung (Entwicklung und Vergleich von Algorithmen) veröffentlicht sind, weshalb nachfolgend eine kurze Darstellung der Grundlagen und erster vielversprechender Ergebnisse, die in einer Anwendung einer nicht-metrischen Diskriminanzanalyse erzielt wurden, gegeben wird.

2 Formale Grundlagen

Gegeben sei eine Mischverteilung H von k n -dimensionalen Verteilungen H_i ($k \geq 2$) mit Dichten h bzw. h_i .

$$H = \sum_{i=1}^k p_i H_i$$

wobei

$$0 < p_i < 1$$

und

$$\sum_{i=1}^k p_i = 1$$

Ziel einer Diskriminanzanalyse ist es, eine Funktion („Zuordnungsregel“)

$$f : \mathbb{R}^n \rightarrow \{1, \dots, k\}$$

zu finden, die ein noch näher zu spezifizierendes Zielkriterium maximiert. Es bezeichne R_i das Teilgebiet des \mathbb{R}^n , das der Gruppe i zugeordnet wird:

$$R_i = \{x \in \mathbb{R}^n : f(x) = i\} = f^{-1}(i)$$

Die Wahrscheinlichkeit e_{ij} , ein Objekt der (tatsächlichen) Gruppe i durch f einer Gruppe j zuzuordnen ist somit

$$e_{ij} = \int_{R_j} h_i(x) dx$$

Insbesondere ist e_{ii} die Wahrscheinlichkeit dafür, ein Objekt aus Gruppe i korrekt zu klassifizieren.

Allgemeines zu Zuordnungsregeln ist der Standardliteratur zur Diskriminanzanalyse zu entnehmen (z.B. McLachlan, 1992); hier sollen nur einige Spezialfälle aufgezählt werden:

a) Totale Summe

Für eine Gruppe i ist der Relativanteil, durch f eine korrekte Klassifikation vorzunehmen, gleich e_{ii} , bezogen auf die Population insgesamt also

$$TS = \sum_{i=1}^k p_i e_{ii}$$

Das Kriterium ist identisch mit dem Bayes'schen Entscheidungskriterium ein Objekt derjenigen Gruppe zuzuordnen, für die die Likelihood $p_i h_i(x)$ maximal unter allen i ist.

Eine für manche Aufgabenstellungen unerwünschte Eigenschaft des Kriteriums besteht in der Einbeziehung der a priori-Wahrscheinlichkeiten der Gruppen p_i . Sei etwa der Anteil einer Gruppe an der Gesamtpopulation sehr gering, so kann es in Hinblick auf das Kriterium optimal sein, kein einziges Objekt der Gruppe zuzuordnen.

b) MAXISUM-Kriterium

Verzichtet man auf die Einbeziehung der a priori-Wahrscheinlichkeiten für Gruppenzugehörigkeiten, so lautet das modifizierte Kriterium

$$MS = \sum_{i=1}^k e_{ii}$$

das äquivalent mit der Annahme $p_i = 1/k$ für die Wahrscheinlichkeitsdichte ist. Nachteilig bei diesem Kriterium kann sein, daß die Anteile der Korrektklassifikationen für manche Gruppen sehr hoch, für andere hingegen relativ niedrig sind.

c) MAXIMIN-Kriterium

Ein in der Literatur häufig behandeltes Kriterium, welches den eben angesprochenen Nachteil nicht besitzt ist

$$MM = \min_i e_{ii}$$

Umgekehrt besteht bei diesem Kriterium der Nachteil, daß die Summe der Korrektklassifikationen u.U. relativ gering sein kann, nur um die Fehlerquote in einer einzigen Gruppe niedrig zu halten.

Zusammenhang mit Sensitivität und Spezifität

In diagnostischen Fragestellungen ist häufig lediglich eine Klassifikation in zwei Gruppen, „Gesund“ und „Patient“, vorzunehmen. Die üblichen Bezeichnungen sind der Tabelle 1 zu entnehmen.

$P=TP+FN$ gibt den Anteil der Patienten in der Population an (Prävalenz), TP/P ist der Anteil der korrekt erkannten Patienten unter den tatsächlichen Patienten (Sensitivität), TN/P' der der korrekt erkannten Gesunden unter den Gesunden (Spezifität). Bezeichnet

Tabelle 1: Bezeichnungen von Anteilen in medizinischen Tests

		Durch Test klassifiziert als		
		Gesund	Patient	
Diagnose	Gesund	TN („true negative“)	FP („false positive“)	P'=1-P
	Patient	FN („false negative“)	TP („true positive“)	P
		Q'=1-Q	Q	1

man die Gesunden mit Gruppe 1 und die Patienten mit Gruppe 2, so ist der Zusammenhang mit den oben eingeführten Termini der DA durch $P' = p_1$, $P = p_2$, $TN = p_1 e_{11}$, ..., $TP = p_2 e_{22}$ gegeben. Insbesondere ist e_{11} die Spezifität und e_{22} die Sensitivität. Zusätzlich zu den bereits genannten Zuordnungskriterien lassen sich daraus weitere formulieren, welche inhaltlich bedeutsam erscheinen, z.B. Maximierung der Spezifität bei vorgegebener Sensitivität u.ä.

3 Bekannte diskriminanzanalytische Verfahren im Überblick

In der Praxis wird nach wie vor am häufigsten die lineare DA nach Fisher verwendet, wesentlich seltener sind schon Anwendungen zu finden, in denen andere in statistischen Standardpaketen (z.B. SAS) verfügbare Verfahren eingesetzt werden. Zur Beurteilung einer neuen Methode ist ein Vergleich mit den gängigen Methoden notwendig, weshalb nachfolgend eine kurze Beschreibung der in SAS oder SPSS angebotenen Routinen erfolgen soll.

3.1 Lineare und quadratische Diskriminanzanalyse

In der linearen DA nach Fisher wird vorausgesetzt, daß die Verteilungen H_i jeweils n -dimensionale Normalverteilungen mit Mittelwertsvektoren μ_i und *identischer* Kovarianzmatrix Σ sind, d.h. es ist

$$h_i(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma^{-1} (x - \mu_i)\right)$$

Zur Schätzung der Verteilungsparameter wird je Gruppe der Mittelwertsvektor μ_i und die gemeinsame Varianz-Kovarianz-Matrix geschätzt. Die Zuordnung erfolgt anhand des Kriteriums TS, durch Vorgabe gleicher a priori-Wahrscheinlichkeiten ist ebenfalls das Kriterium MS zugänglich. Bedeutsam ist in diesem Zusammenhang, daß die resultierende Aufteilung des \mathbb{R}^n durch $(n-1)$ -dimensionale Hyperebenen erfolgt (im bivariaten Fall sind die Grenzlinien zwischen den Bereichen R_i Geradenstücke). Die Zuordnung von Objekten

erfolgt anhand von linearen Klassifikationsfunktionen

$$\begin{aligned}
 c_1(x) &= a_{10} + a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\
 &\vdots \\
 c_i(x) &= a_{i0} + a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \\
 &\vdots \\
 c_k(x) &= a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kn}x_n
 \end{aligned}$$

Ein Objekt mit Datenvektor (x_1, \dots, x_n) wird der Gruppe i zugeordnet, für die c_i maximal ist.

Eine einfache Erweiterung des Verfahrens ist die quadratische Diskriminanzanalyse, in der auch für jede Gruppe eine eigene Varianz-Kovarianz-Matrix Σ_i angenommen wird:

$$h_i(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right)$$

„Quadratisch“ heißt diese DA deshalb, weil die Klassifikationsfunktionen c_i in diesem Fall quadratische Formen sind, die Grenzlinien zwischen den R_i sind somit z.B. parabolisch oder elliptisch geformt.

3.2 Kernel-Verfahren

In Kernel-Verfahren wird durch Wahl einer n -dimensionalen Kernfunktion eine Schätzung der Verteilungsdichte für die einzelnen Dichtefunktionen h_i vorgenommen und danach – normalerweise anhand des Kriteriums TS oder MS – der \mathbb{R}^n aufgeteilt. Die resultierende Aufteilung des Raums besitzt keine weiteren spezifischen Eigenschaften, durch die Wahl des Kerns (insbesondere seiner „Breite“) kann eine mehr oder weniger „glatte“ Aufteilung erreicht werden.

3.3 K -Nearest-Neighbor

Das K -Nearest-Neighbor-Verfahren werden für jeden Punkt des \mathbb{R}^n anhand einer zu wählenden Metrik die K nächsten Nachbarn bestimmt und jeder Punkt der Gruppe zugeordnet, die den höchsten Anteil unter diesen K Punkten besitzt. Das Verfahren kann als Modifikation des Kernel-Verfahrens verstanden werden und besitzt für die resultierende Aufteilung in die R_i ähnliche Eigenschaften: Ein kleines K ergibt eine Aufteilung in viele einzelne Bereiche, mit steigendem K werden die Bereiche wieder zusammenhängender. Durch mit der Größe der Stichprobe in einer bestimmten Relation wachsendem K gewinnt man analog zur Kernel-Methode eine Schätzung der Verteilungsdichten der h_i .

4 Nicht-metrische Diskriminanzanalyse

4.1 Konvexe Partitionierung

Eine Eigenschaft der linearen DA nach Fisher besteht darin, daß die bei einer Diskriminanzanalyse erfolgende Aufteilung des \mathbb{R}^n zu *konvexen* Mengen führt, d.h. zu Mengen, bei denen mit je zwei Punkten auch die gesamte, die beiden Punkte verbindende Strecke mit in der Menge enthalten ist:

$$M \text{ ist konvex} \Leftrightarrow \forall x, y \in M, 0 \leq a \leq 1 : ax + (1 - a)y \in M$$

Man zeigt leicht, daß daraus für eine beliebige Menge x_i in M jede konvexe Linearkombination

$$\sum a_i x_i$$

mit

$$\begin{aligned} a_i &\geq 0 \\ \sum_i a_i &= 1 \end{aligned}$$

in M enthalten ist.

Unter einer *konvexen Partitionierung* einer Menge M versteht man eine Aufteilung der Menge in *konvexe* Teilmengen M_i mit

$$\begin{aligned} \bigcup M_i &= M \\ \forall_{i \neq j} M_i \cap M_j &= \emptyset \end{aligned}$$

Eine konvexe Partitionierung des \mathbb{R}^n besitzt die Eigenschaft, daß jeweils zwei Mengen M_i und M_j *linear separabel* sind, d.h. es existiert eine Funktion w , so daß

$$\begin{aligned} wx &> 0 \quad \text{für } x \in M_i^\circ \\ wx &< 0 \quad \text{für } x \in M_j^\circ \end{aligned}$$

wobei M° die Menge M ohne ihren Rand bezeichnet (für diesen gilt = statt < bzw. >). Umgekehrt gilt auch, daß zwei linear separable Mengen konvex sind.

4.2 Anwendung in der Diskriminanzanalyse

Raveh (1989) schlug unter der Bezeichnung „nonmetric discriminant analysis“ (NDA) ein Verfahren vor, bei dem in Anwendung auf zwei Gruppen eine lineare Funktion w

$$w \cdot x = \sum_{i=1}^n w_i x_i$$

gesucht wird, für die die Zahl der *Scores*

$$z(x) := w \cdot x$$

der Objekte $x^{(1)}$ in Gruppe 1 möglichst oft *größer* als für Objekte $x^{(2)}$ in Gruppe 2 ist. Raveh bezeichnet das Verfahren als *nicht-metrisch*, weil das Zielkriterium, welches maximiert wird, allein auf der Anzahl der erfüllten Ungleichungen

$$z(x_i^{(1)}) \geq z(x_j^{(2)}), \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2$$

beruht (n_1 und n_2 seien die Gruppengrößen von Gruppe 1 bzw. 2). Als tatsächlichen Separationsindex verwendet Raveh

$$S = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (z(x_i^{(1)}) - z(x_j^{(2)}))}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |z(x_i^{(1)}) - z(x_j^{(2)})|}$$

Zur numerischen Bestimmung der optimalen Zielfunktion w schlägt Raveh das Powell'sche konjugierte Gradientenverfahren vor. Als Zuordnungsregel wird schließlich ein von den Gruppengrößen abhängiger Cutoff-Punkt c bestimmt (im Fall $n_1 = n_2$ der Median aller $z(x)$) und ein Objekt x der Gruppe 1 zugeordnet, falls $z(x) \geq c$ ist, ansonsten der Gruppe 2.

Durch die lineare Funktion $z(x)$ wird der \mathbb{R}^n und damit auch die Menge der Objekte x in zwei konvexe Mengen partitioniert. Eine wichtige Eigenschaft des Verfahrens ist, daß das Verfahren zu den gleichen Resultaten nach beliebigen affinen Transformationen der Meßwertvektore x führt, insbesondere ist es *skalierungsinvariant*. Ein Nachteil des Verfahrens ist, daß die Zielfunktion S keinen direkten Zusammenhang mit den üblichen diskriminanzanalytischen Zuordnungsregeln besitzt, darüberhinaus ist auch das numerische Verfahren nur approximativ, da die Zielfunktion u.U. verschiedene lokale Maxima aufweist.

Einen anderen Ansatz verfolgt die Arbeit von Burshtein et al. (1992), in der der Begriff der „minimum impurity partitions“ eingeführt wird und in der gezeigt wird, daß für bestimmte Zielfunktionsklassen die Lösung über einen finiten Objektraum im \mathbb{R}^n statt durch vollständiger Enumeration aller Partitionen auf die Betrachtung aller linear separablen Teilräume beschränkt werden kann. Bekanntlich gibt es

$$\frac{1}{k!} \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^n$$

Partitionen eine m -elementigen Menge in k nichtleere Teilmengen. Durch die Beschränkung auf Partitionen in linear separable (und damit konvexe) Teilmengen, genügt es im \mathbb{R}^n , sich auf

$$2 \sum_{i=0}^n \binom{m-1}{i}$$

Partitionen beschränken; dieser Ausdruck ist nur noch polynomial abhängig von m (Burshtein, 1991, siehe hierzu auch Cover (1965), in der die Anwendung auf diskriminanzanalytische Fragestellungen bereits angedeutet wird).

Es liegt nahe, die beiden Ansätze von Raveh und Burshtein et al. zu verknüpfen und unter Zugrundelegung der üblichen Zielkriterien (z.B. MAXIMIN oder MAXISUM) eine optimale Partition der Objektmenge in linear separable (also konvexe) Teilmengen vorzunehmen, die „minimal verunreinigt“ (im Sinne des Zielkriteriums) sind. Die Zuordnungsregel besteht darin, ein Objekt x der Menge zuzuordnen, in deren konvexen Hülle sie liegt. Im Gegensatz zu parametrischen Diskriminanzanalysen ist es dabei jedoch möglich, daß Teilgebiete des \mathbb{R}^n keiner Gruppe zugeordnet sind, da diese außerhalb der konvexen Hüllen der verschiedenen Gruppen liegen. Eine ähnliche Eigenschaft besitzt auch das K -Nearest-Neighbor-Verfahren. Aus inhaltlichen Gründen kann dies sogar wünschenswert sein: Wenn aus den Daten und den Modellannahmen keine Aussagen über die (wahrscheinliche) Grundmenge zu treffen sind, so ist das selbst eine u.U. relevante Information (man vgl. z.B. die Problematik von Normbereichen, s. Ackermann 1985). Ist man an einer exhaustiven Partitionierung des gesamten \mathbb{R}^n interessiert, so kann für die nicht durch die konvexen Hüllen abgedeckten Bereiche eine zusätzliche Zuordnungsregel eingeführt werden, die der Konvexitätsannahme nicht widerspricht.

Das entsprechende C++-Programm zur Durchführung der nichtparametrischen Diskriminanzanalyse mit den Zielkriterien Totale Summe, MAXIMIN und MAXISUM verwendet einen Branch and Bound Algorithmus zur Enumeration aller k voneinander linear separablen Teilmengen im \mathbb{R}^n .

4.3 Erweiterungen

Die Beschränkung auf konvexe Partitionen mag in manchen Fällen als zu restriktiv erscheinen, analog zur quadratischen DA wäre die Aufteilung in Gebiete, die durch quadratische Formen abgegrenzt sind (z.B. entlang einer Parabel, Ausschneiden einer Ellipse o.ä.), ebenfalls in Betracht zu ziehen. Eine solche Erweiterung ist in der vorgestellten DA jedoch bereits implizit enthalten: Werden im bivariaten Fall statt der Variablen x_1 und x_2 zusätzlich x_1^2 , y_2^2 und $x_1 \cdot x_2$ mit aufgenommen, so sind die resultierenden konvexen Partitionen im 5-dimensionalen Raum im ursprünglich zweidimensionalen Raum Gebiete, die durch quadratische Formen (Hyperbeln, Ellipsen, Parabeln) abgegrenzt sind (s. insbesondere Cover, 1965). Praktisch stößt dieses Verfahren jedoch schnell an seine Grenzen, da die Zahl der Variablen sich quadratisch mit der der Ursprungsvariablen erhöht und der z.Z. bestehende Algorithmus noch zu langsam ist, um in akzeptabler Zeit das Ergebnis zu liefern.

5 Schmerzwahrnehmung bei chronischen Schmerzpatienten

In einer Untersuchung zur Schmerzverarbeitung bei chronischen Schmerzpatienten (Kopfschmerzen und Wirbelsäulensyndrom) wurden u.a. mehrere psychophysikalische Kenngrößen der Schmerzwahrnehmung erhoben, von denen angenommen wird, daß sie bei chronischen Schmerzpatienten gegenüber Gesunden *charakteristisch* verändert sind (Kleinböhl et al., 1998). Dementsprechend sollte eine Trennung der Patientengruppen von gesunden Kontrollpersonen durch eine DA möglich sein. Das hier dargestellte Beispiel beschränkt sich auf drei Variable: (a) Mit Grenzwertmethode erhobene Schmerzschwelle PT_{lim} , (b) Schmerzschwelle für tonische Hitzereize PT_{ton} und (c) mittlere Sensitivierung über neun um die Schmerzschwelle dargebotene Hitzereize DT_{mean} (eine eingehende Beschreibung der Prozedur und der Maße findet sich in Kleinböhl, 1998). Insgesamt liegen der DA die Daten von 15 chronischen Schmerzpatienten mit Kopfschmerz (Gruppe H: Headache), mit 15 Patienten mit Rückenschmerzen (Gruppe M: Muskuloskeletal pain) und 23 Kontrollpersonen (Gruppe C: Control) zugrunde. Tabelle 2 zeigt die Basisstatistiken der Schmerzmaße für die drei Gruppen.

Tabelle 2: Basistatistiken der Schmerzgrößen.

	Control (C)	Headache (H)	Muskuloskeletal pain (M)
N	23	15	15
PT_{lim}	45.50 ± 2.06	43.85 ± 1.94	45.56 ± 2.11
PT_{ton}	44.82 ± 1.45	44.00 ± 1.91	43.77 ± 1.67
DT_{mean}	0.192 ± 0.35	-0.22 ± 0.81	-0.98 ± 2.08

Die Abbildungen 1 (a), (b) und (c) zeigen aus verschiedenen Perspektiven den Datenquader für die drei Variablen.

In Tabelle 3 sind für die lineare Diskriminationsanalyse nach Fisher, der nichtparametrischen DA mit dem MAXIMIN- und dem MAXISUM-Kriterium die Klassifikationsmatrizen mit den bivariaten Variablenkombinationen aus PT_{lim} , PT_{ton} und DT_{mean} sowie für alle drei Variablen aufgeführt.

Die Ergebnisse zeigen die deutliche Überlegenheit der nichtparametrischen DA über die auf der multivariaten Normalverteilungsannahme beruhenden DA nach Fisher. In beiden Kriterien MAXIMIN und MAXISUM ist der Anteil der korrekt klassifizierten Datenpunkte deutlich höher. So ist etwa beim Variablenpaar (PT_{lim} , PT_{ton}) in der linearen DA nach Fisher für die Gruppe M lediglich ein Anteil von 40% korrekter Klassifikationen erreicht, unter Zugrundelegung des MINIMAX-Kriteriums läßt sich dieser Anteil auf 60% steigern. Für beide nichtparametrischen DA ist die Summe der korrekten Klassifikationen mit über 193% nahezu um 50% höher als bei der klassischen Diskriminanzanalyse. Für das Variablenpaar PT_{lim} und PT_{ton} zeigen die Abbildungen 2 (a) und (b) das Klassifikationsergebnis für das MAXIMIN-Kriterium.

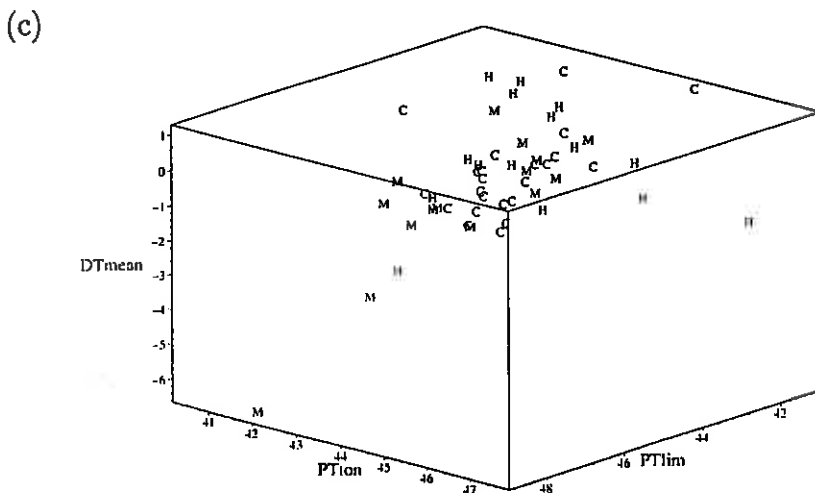
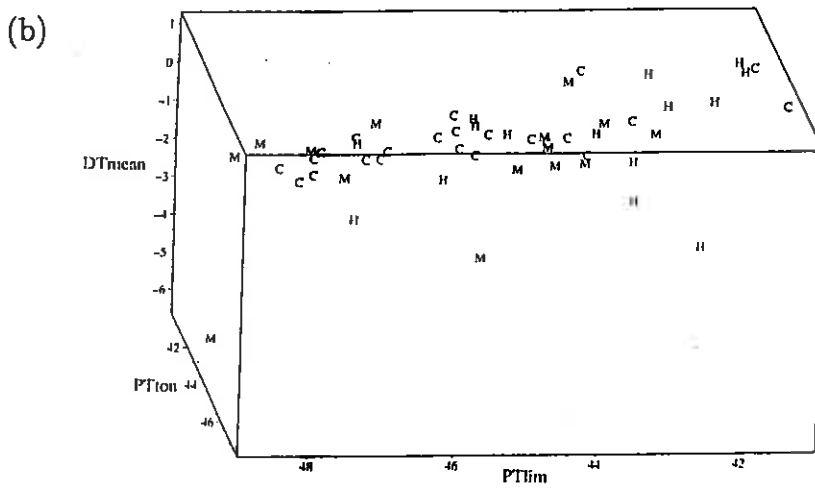
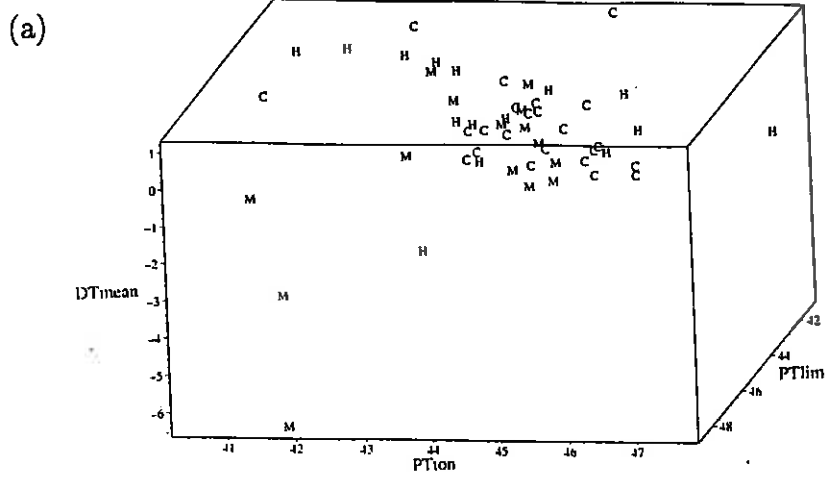


Abbildung 1: Daten der Schmerzuntersuchung.

Die drei Datenquader zeigen aus verschiedenen Perspektiven die Rohdaten für die Untersuchungsgruppen Kopfschmerz (H, N=15), Rückenschmerz (M, N=15) und Kontrollen (C, N=23).

Tabelle 3: Vergleich der Diskriminanzanalysen.

Die Tabelle zeigt die Klassifikationsmatrizen der linearen DA nach Fisher und der nichtparamterischen linearen DA für die Zielkriterien MAXISUM und MAXIMIN für die Variablenkombinationen (PTlim, PTton), (PTlim, DTmean), (PTton, DTmean) und (PTlim, PTton, DTmean).

PTlim, PTton	lineare DA nach Fisher				nichtpar. DA (MAXIMIN)				nichtpar. DA (MAXISUM)			
	C	H	M	korr	C	H	M	korr	C	H	M	korr
C	11	7	5	47.83	17	3	3	73.91	10	3	10	43.48
H	2	9	4	60.00	4	9	2	60.00	1	9	5	60.00
M	3	6	6	40.00	5	1	9	60.00	0	1	14	93.33
Min.				40.00				60.00				43.48
Summe				147.83				193.91				196.81
PTlim, DTmean												
	C	H	M	korr	C	H	M	korr	C	H	M	korr
C	15	7	1	65.22	14	4	5	60.87	18	2	3	78.26
H	5	9	1	60.00	3	10	2	66.67	1	9	5	60.00
M	6	5	4	26.67	1	4	10	66.67	3	2	10	66.67
Min.				26.67				60.87				60.00
Summe				151.89				194.21				204.93
PTton, DTmean												
	C	H	M	korr	C	H	M	korr	C	H	M	korr
C	16	6	1	69.56	18	4	1	78.26	16	2	5	69.56
H	5	8	2	53.33	1	9	5	60.00	2	8	5	53.33
M	5	5	5	33.33	4	2	9	60.00	2	1	12	80.00
Min.				33.33				60.00				53.33
Summe				156.22				198.26				202.89
PTlim, PTton DTmean												
	C	H	M	korr	C	H	M	korr	C	H	M	korr
C	14	7	2	60.87	18	4	1	78.26	18	2	3	78.26
H	5	9	1	60.00	1	10	4	66.67	2	9	4	60.00
M	5	6	4	26.67	4	1	10	66.67	2	1	12	80.00
Min.				26.67				66.67				60.00
Summe				147.54				211.60				218.26

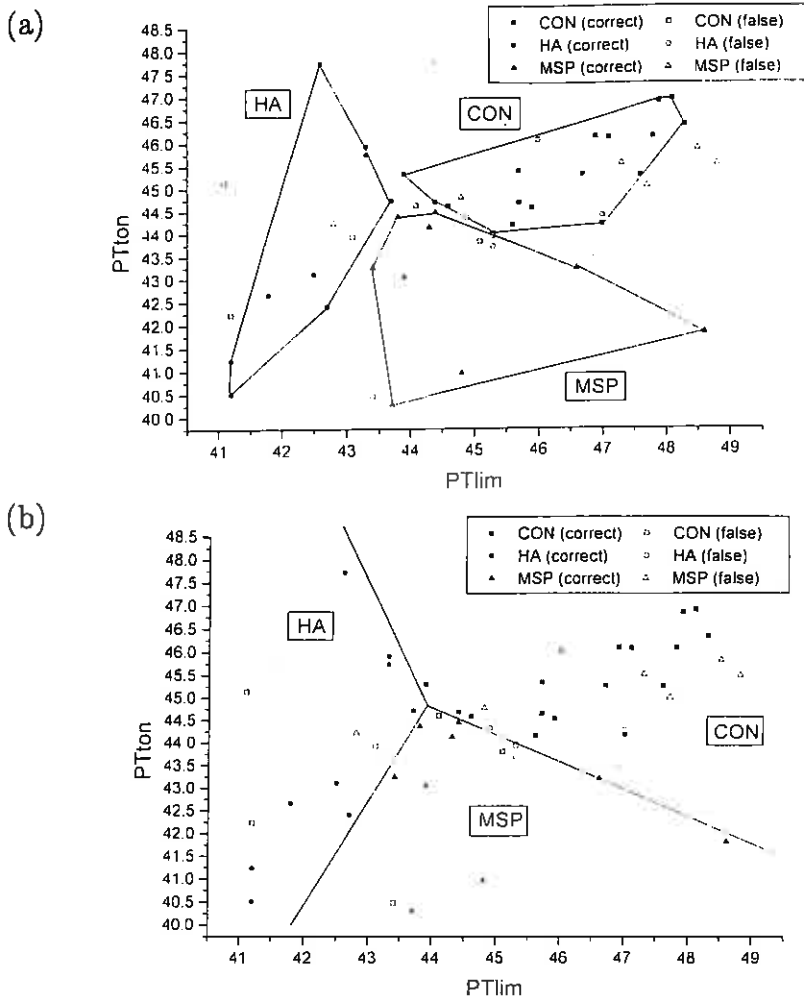


Abbildung 2: Nichtparametrische DA für das Variablenpaar (PTlim, PTton).

In Abb. (a) sind die durch die nichtparametrische DA gefundenen konvexen Gebiete der drei Gruppen dargestellt. Die konvexen Gebiete stellen „minimal verunreinigte“ Bereiche der Objektmengen im Sinne von Burshtein et al. (1992) dar. In Abb. (b) ist die entsprechende Partitionierung des \mathbb{R}^2 angegeben. Die in (a) gezeigten konvexen Gebiete sind durch Geradenstücke voneinander getrennt (linear separabel).

Noch deutlicher ist die Überlegenheit bei Verwendung aller drei Variablen. Die Summe der Trefferraten steigt bei der linearen DA nach Fisher gegenüber den bivariaten Trefferraten nicht an, die nichtparametrische DA erlaubt beim MINIMAX Kriterium in allen Gruppen eine Mindesttrefferrate von 66.7%, in der Summe werden 212% bzw. 218% (MAXIMIN bzw. MAXISUM) erreicht.

In den Abbildungen 3 ist die Trennung der konvexen Gebiete der Gruppenpaare (C,H), (C,M) und (H,M) dargestellt. Aufgeführt sind in den Abbildungen nur die korrekt klassifizierten Datenpunkte, die Projektionen sind dabei so gewählt, daß die lineare Separabilität deutlich wird.

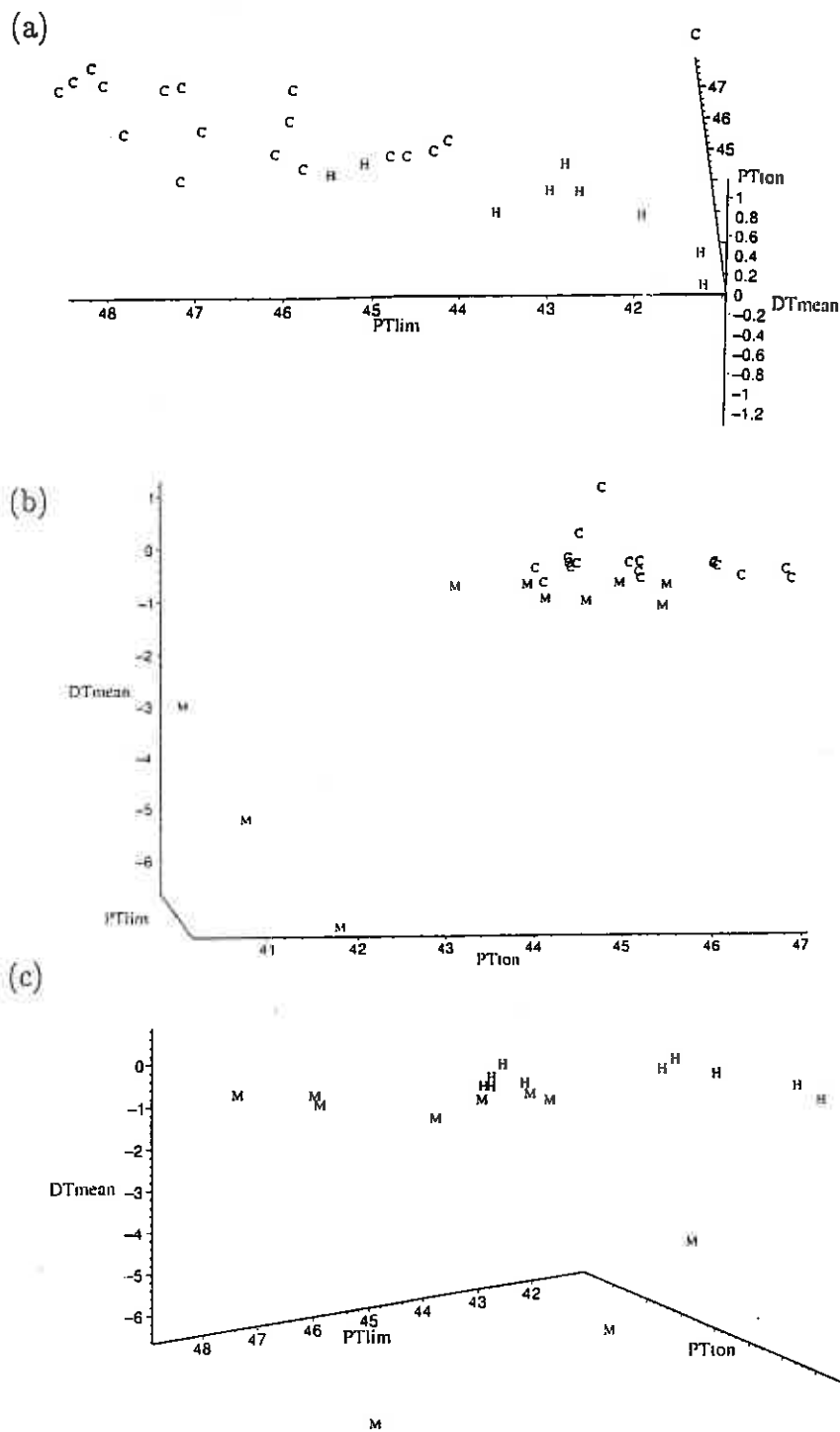


Abbildung 3: Nichtparametrische DA für das Variablen triplet (PTlim, PTton, DTmean). Die Abbildungen zeigen die lineare Separabilität der korrekt klassifizierten Personen im Datenquader (s. Abb. 1). In (a) ist die Projektion gezeigt, bei der die Gruppen C und H voneinander durch eine Gerade getrennt werden können. Analog zeigt (b) die Trennung von C und M sowie (c) die der Gruppen M und H.

Literatur

- Ackermann, H (1985). Mehrdimensionale nicht-parametrische Normbereiche: methodologische und medizinische Aspekte (Medizinische Informatik und Statistik, Bd 57). Springer, Berlin.
- Burshtein, D, Della Pietra, V, Kenevsky D, Nádas, A (1991). Splitting theorems for classification and regression trees. C.R. Acad. Sci. Paris., 313, S. I, 537–540.
- Burshtein, D, Della Pietra, V, Kenevsky D, Nádas, A (1992). Minimum impurity partitions. Ann. Stat., 20, 1637–1646.
- Cover, TM (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. Comput., EC-14, 326–334.
- Kleinböhl, D, Hölzl, R, Möltner, A, Rommel, C, Osswald, PM, Weber, C (1998). Psychophysical measures of sensitization to tonic heat discriminate chronic pain patients. Eingereicht bei Pain.
- McLachlan, GJ (1992). Discriminant analysis and statistical pattern recognition. Wiley, New York.
- Raveh, A (1989). A nonmetric approach to linear discriminant analysis. J. Amer. Stat. Ass., 84, 176–183.