Multivariate Analysemethoden

Diskriminanzanalyse

Günter Meinhardt
Johannes Gutenberg Universität Mainz

Discriminant Function Analysis (DFA)

Ziele

- Maximale Trennung von Gruppen auf einem gegebenem Set von p Meßvariablen.
- Auffinden von latenten Diskriminanzfunktionen, die sukzessive maximale Gruppentrennung gewährleisten.
- In der Regel: Auffinden eines niedrig dimensionierten Diskriminanzraumes, in dem die Gruppen separierbar sind.
- Case-Classification in optimalen, niedrig dimensionierten Räumen.
- Bestimmung von Klassifikationsfunktionen für Case-Classification.

Voraussetzung

- Gleiche (homogene) Varianz-Kovarianz Matrizen in allen Gruppen.
- Testungen der Gruppenunterschiede (Centroide), sowie der Homogenität der Σ_j - Matrizen erfordern die Gültigkeit der multivariaten Normalverteilung.

Ansatz

- Optimierung des Verhältnisses der Quadratsummen für „between“ und „within“ Group Varianz.
- Lösung über Eigenwertzerlegung einer aus **B** und **W** Komponenten zusammengesetzten Matrix.

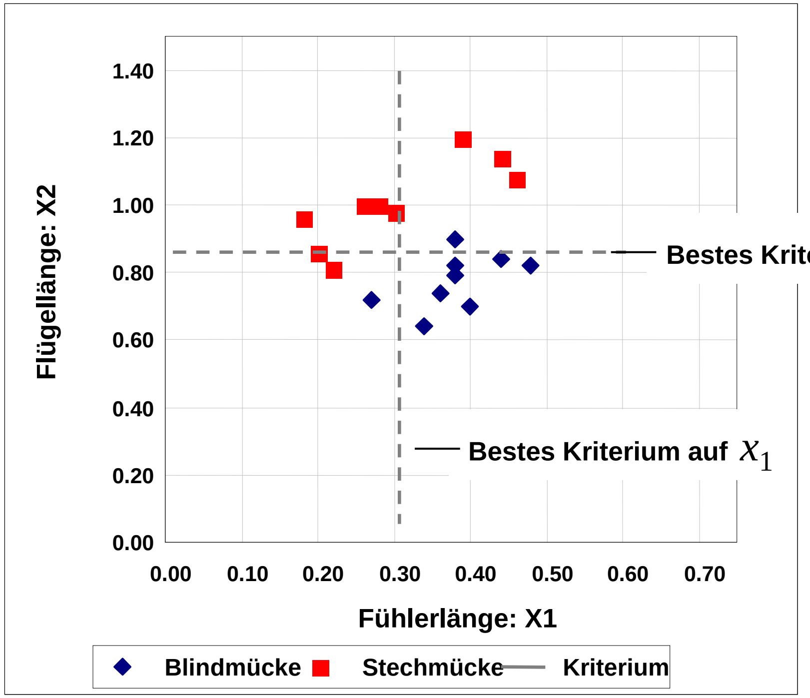
Anwendung

- Diagnostische Trennung schwierig zu trennender Gruppen.
- Bestimmung kritischer diagnostischer Variablen / Reduktion auf relevante diagnostische Variablen in multivariaten Klassifikationen.
- Konstruktion von Algorithmen zur Mustertrennung (Pattern recognition machines) und Bildklassifikation (bildgebende Verf.).
- Qualitätskontrolle und Evaluation von Versuchs- und Kontrollgruppen in multivariaten Designs.

Nachteile

- Restriktion gleicher Varianz-Kovarianz Matrizen in allen Gruppen.
- Case-Classification: Klassifikation im Diskriminanzraum hat gegenüber MDC und Baysian Classifier keine wesentlichen Vorteile (außer Sparsamkeit) und läuft auf dasselbe hinaus.

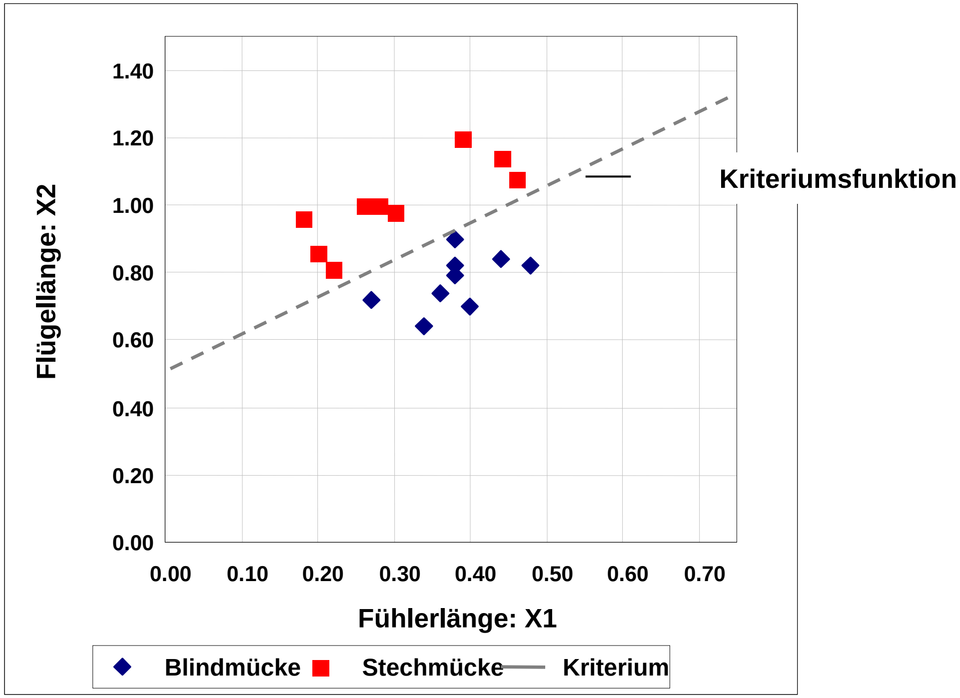
2D-Beispiel



Problem

- Klassifiziere anhand von Fühlerlänge (X_1) und Flügellänge (X_2) möglichst eindeutig in Stechmücke (c_1) und Blindmücke (c_2).
- Das geht mit einem Kriteriumswert auf jeder einzelnen Variable X_1 und X_2 offenbar nicht.

2D-Beispiel



Lösung:

Eine lineare Kriteriumsfunktion teilt den Variablenraum in 2 Gebiete: Oberhalb Stechmücke (c_1), unterhalb Blindmücke (c_2).

$$x_2 = b + ax_1$$

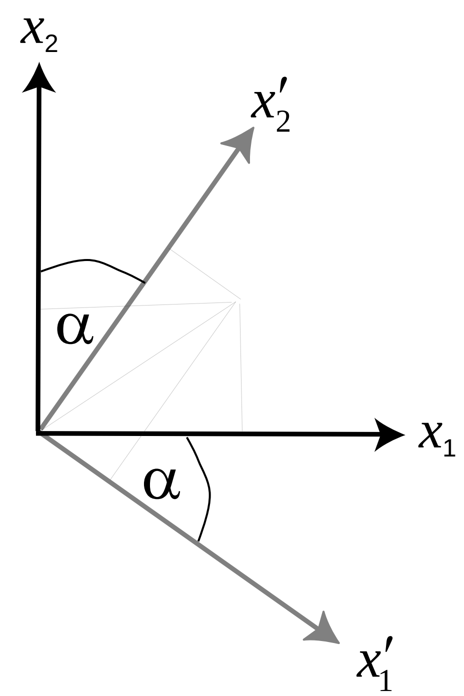
Somit folgt die Klassifikationsfunktion

$$g(x_1, x_2) = \begin{cases} c_1, & \text{wenn } x_2 - ax_1 > b \\ c_2, & \text{wenn } x_2 - ax_1 \leq b \end{cases}$$

Einfache Lösung:

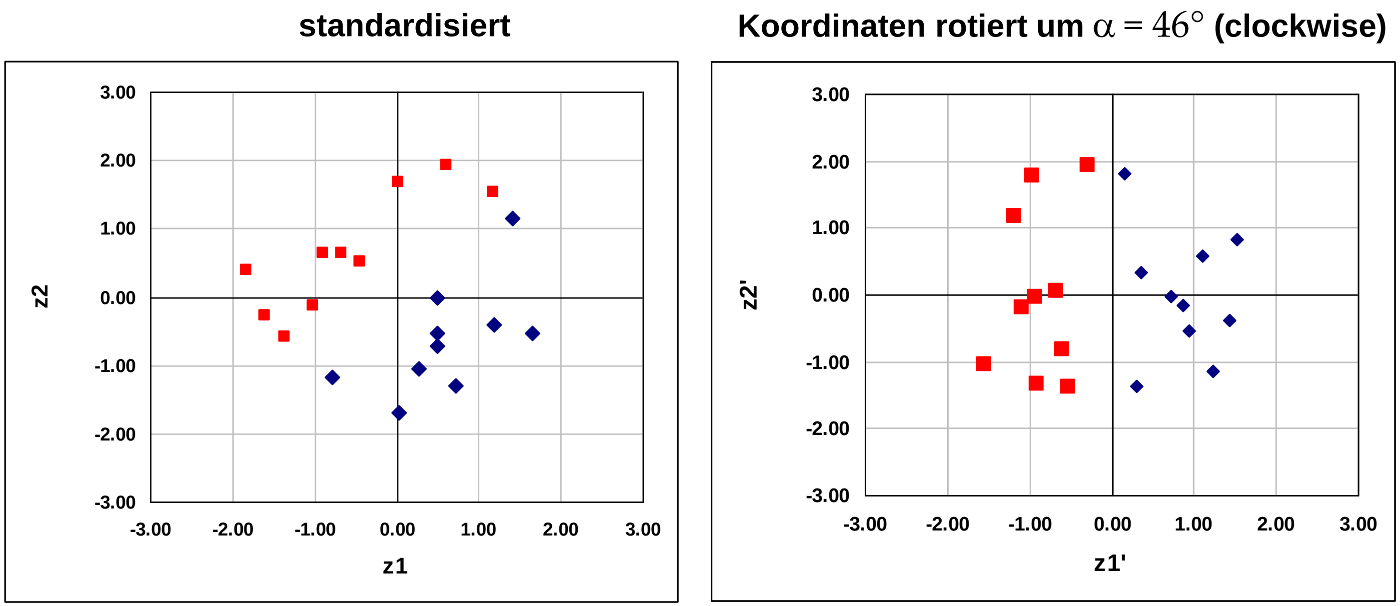
Zuerst die Daten im Nullpunkt zentrieren und dann um den optimalen Winkel α drehen !

Zentrierung & Rotation



Die Varianz zwischen den Gruppen wird auf der Achse x'_1 maximiert, und x'_2 steht senkrecht x'_1. Eine Parallele zu x'_2 liefert das optimale Trennkriterium.

z-Standard

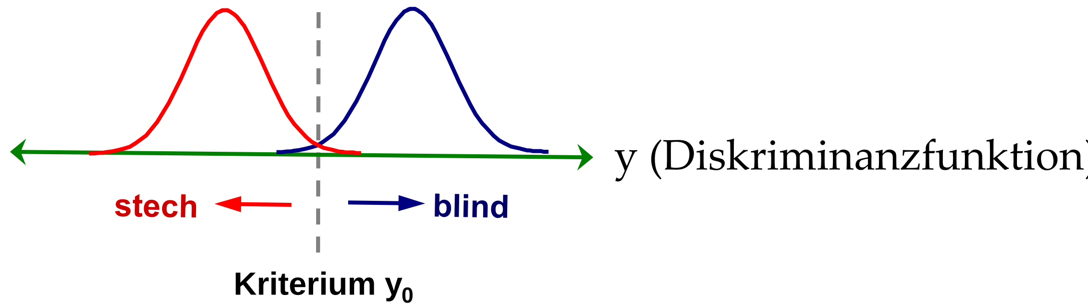


Diskriminanz-funktion

- Die neue x- Achse z_1' ist die Diskriminanzfunktion y. Auf ihr läßt sich ein Kriterium zur optimalen Trennung beider Gruppen finden.
- Da eine Drehoperation auf die Diskriminanzfunktion geführt hat, ist sie darstellbar als eine Linearkombination der alten Koordinaten:

$$y = b_1 x_1 + b_2 x_2$$

y: Linear-kombination



$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} z_1' \\ z_2' \end{pmatrix} \Rightarrow \begin{aligned} z_1 \cos\alpha - z_2 \sin\alpha &= z_1' \\ z_1 \sin\alpha + z_2 \cos\alpha &= z_2' \end{aligned}$$

Da $y = z_1'$ gilt

$$\boxed{y = b_1 z_1 + b_2 z_2}$$

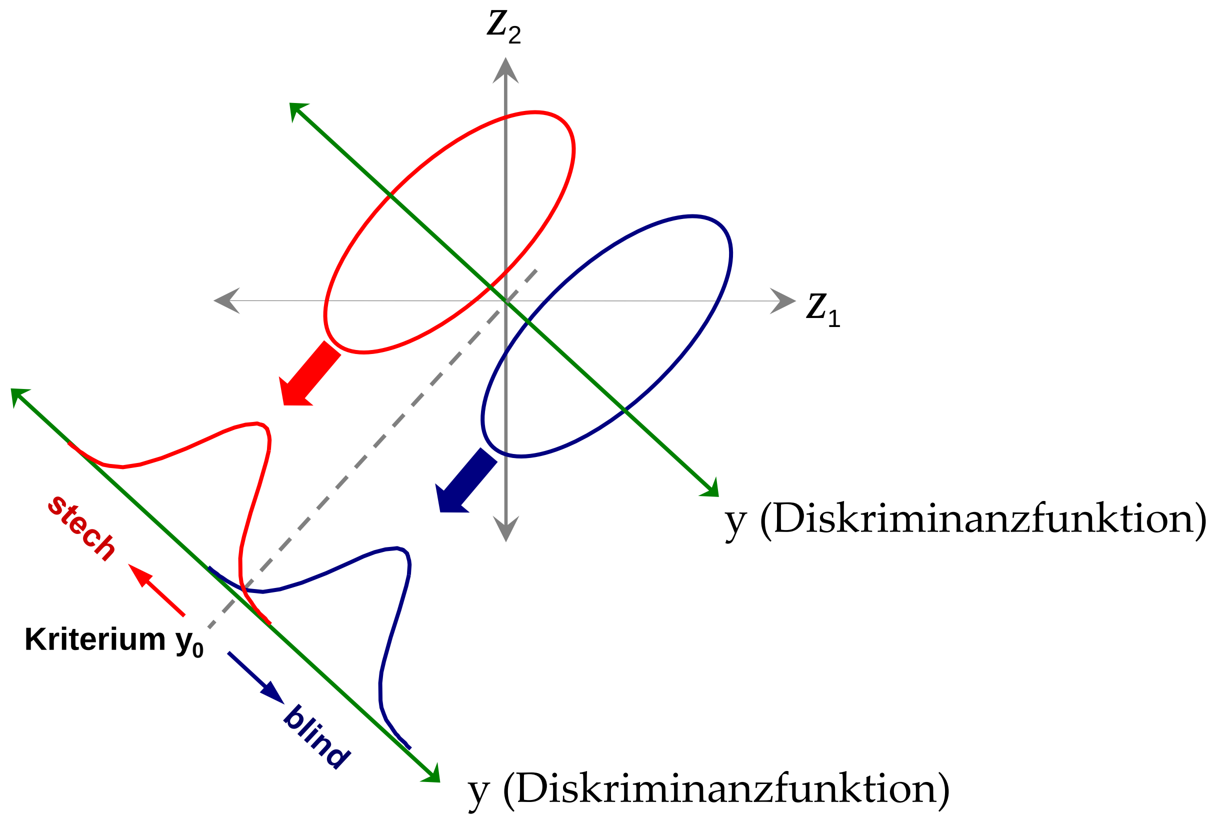
mit $b_1 = \cos\alpha$ und $b_2 = -\sin\alpha$

Koeffizienten von y

Das Auffinden der Koeffizienten b_1 und b_2 ist also identisch mit dem Problem, den optimalen Drehwinkel α zu bestimmen. Hierfür braucht man ein Kriterium der gewünschten maximalen Trennung, und die Lösung des dahinter stehenden Maximierungsproblems.

[Excel-Beispiel]

Rotation zur y - Funktion



Klassifikation

- Case-Classification durch einfachen Vergleich mit dem Kriterium y_0.
- Prüfung des Gruppenunterschieds mit einem einfachen t - Test auf y.
- Voraussetzung: homogene Varianz-Kovarianz Matrizen.

Kriterium der Maximierung

Maximiert wird das Verhältnis der Quadratsummen für die Variation auf y zwischen Gruppen QS_B und der Variation innerhalb Gruppen QS_W.

$$\boxed{\Gamma\left(\vec{b}\right)=\frac{QS_{Between}}{QS_{Within}}=\frac{\text{erklärte Variation}}{\text{nicht erklärte Variation}}=\max}$$

(Wähle die Koeffizienten b so, daß Γ(b) maximal wird)

Quadrat-summen-zerlegung

Wie in der Varianzanalyse gilt die Quadratsummenzerlegung

$$QS_{Total}=QS_{Between}+QS_{Within}$$

mit

$$QS_{Total}=\sum_{l=1}^{K}\sum_{i=1}^{n_l}\left(y_{il}-\bar{y}\right)^2$$

$$QS_{Between}=\sum_{l=1}^{K}n_l\left(\bar{y}_l-\bar{y}\right)^2$$

$$QS_{Within}=\sum_{l=1}^{K}\sum_{i=1}^{n_l}\left(y_{il}-\bar{y}_l\right)^2$$

K = Anzahl Gruppen
n_l = Umfang Gruppe l

Kenngrößen der Güte

γ: (Eigenwert der Maximierung)

$$\gamma = \frac{\text{erklärte Var}}{\text{nicht erklärte Var}} \quad \Rightarrow \quad \frac{\gamma}{1+\gamma} = \frac{QS_B}{QS_B + QS_W} = \frac{\text{erklärte Var}}{\text{Gesamt Var}}$$

$$c^2 = \frac{\gamma}{1+\gamma} = \frac{\text{erklärte Var}}{\text{Gesamt Var}} \quad , \quad c = \text{"Kanonische Korrelation"}$$

$$\Lambda = \frac{1}{1+\gamma} = \frac{\text{nicht erklärte Var}}{\text{Gesamt Var}} \quad , \quad \Lambda = \text{"Wilk's Lambda"}$$

Offenbar gilt

$$c^2 + \Lambda = \frac{QS_B}{QS_B + QS_W} + \frac{QS_W}{QS_B + QS_W} = 1$$

χ^2 - Test der Trennleistung

$$\chi^2 = -\left(N - \frac{m+K}{2} - 1 \right) \ln \Lambda$$

K = Anzahl Gruppen
N = $\Sigma n_l = n_1 + n_2 + \ldots n_K$
m = Anzahl Variablen

ist χ^2 verteilt mit m(K-1) Freiheitsgraden

→ Die Trennleistung wird mit einem χ^2 Test auf Signifikanz getestet

Gepoolte Varianz der y - Funktion

Die Varianz innerhalb der Gruppen wird zu einer gepoolt:

$$\hat{s}_y^2 = \frac{\sum_l QS_{W(l)}}{df_l} = \frac{QS_{W(1)} + QS_{W(2)} + \ldots + QS_{W(K)}}{(n_1 - 1) + (n_2 - 1) + \ldots + (n_K - 1)} = \frac{\sum_l n_l \cdot s_l^2}{\sum_l (n_l - 1)}$$

Normierte y - Funktion

Damit wird die Varianz der Diskriminanzfunktion auf 1 normiert:

$$\tilde{y} = \frac{y}{\hat{s}_y} \Rightarrow \operatorname{var}(\tilde{y}) = \left(\frac{1}{\hat{s}_y} \right)^2 \cdot \operatorname{var}(y) = \frac{\hat{s}_y^2}{\hat{s}_y^2} = 1$$

MANOVA

Additivität der Variation

Es gilt: $\mathbf{M} = \mathbf{B} + \mathbf{W}$

Totale QS und Kreuzprodukte

Between Group QS und Kreuzprodukte

Within Group QS und Kreuzprodukte

Kompakte Darstellung

$$\sum_l \sum_i (\mathbf{x}_{il} - \bar{\mathbf{x}})(\mathbf{x}_{il} - \bar{\mathbf{x}})^t = \sum_l n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^t + \sum_l \sum_i (\mathbf{x}_{il} - \bar{\mathbf{x}}_l)(\mathbf{x}_{il} - \bar{\mathbf{x}}_l)^t$$

Hierin sind die **x** Vektoren mit *m* Komponenten (Variablen):

$$\mathbf{x}_{il} = \begin{pmatrix} x_{il1} \\ x_{il2} \\ \vdots \\ x_{ilm} \end{pmatrix} \qquad \bar{\mathbf{x}}_l = \begin{pmatrix} \bar{x}_{l1} \\ \bar{x}_{l2} \\ \vdots \\ \bar{x}_{lm} \end{pmatrix} \qquad \bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_m \end{pmatrix}$$

Regel

Die Matrizen **B** und **W** werden als inneres Produkt (Zeilen- mal Spalten) der Variablen-Vektoren aufgebaut und dann über Fälle und Gruppen summiert.

B und W Matrix der x-Variablen

B-Matrix (p=2 Vars)

Treatment (Group) Quadratsummen & Kreuzprodukte

$$\begin{array}{cc} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \mathbf{B} = & \begin{pmatrix} TQS_1 & TQS_{12} \\ TQS_{12} & TQS_2 \end{pmatrix} \begin{array}{c} x_1 \\ x_2 \end{array} \end{array}$$

Komponenten

$$TQS_1 = n_1\left(\bar{x}_{11} - \bar{x}_1\right)^2 + n_2\left(\bar{x}_{21} - \bar{x}_1\right)^2$$

Group Var

$$TQS_2 = n_1\left(\bar{x}_{12} - \bar{x}_2\right)^2 + n_2\left(\bar{x}_{22} - \bar{x}_2\right)^2$$

$$TQS_{12} = n_1\left(\bar{x}_{11} - \bar{x}_1\right)\left(\bar{x}_{12} - \bar{x}_2\right) + n_2\left(\bar{x}_{21} - \bar{x}_1\right)\left(\bar{x}_{22} - \bar{x}_2\right)$$

W-Matrix (p=2 Vars)

Within group Quadratsummen & Kreuzprodukte (gepoolt)

$$W = \begin{array}{cc} & \begin{array}{cc} x_1 & x_2 \end{array} \\ & \begin{pmatrix} WQS_1 & WQS_{12} \\ WQS_{12} & WQS_2 \end{pmatrix} \begin{array}{c} x_1 \\ x_2 \end{array} \end{array}$$

Komponenten

$$WQS_1 = \sum_{i=1}^{n_1} \left(x_{i11} - \bar{x}_{11}\right)^2 + \sum_{i=1}^{n_2} \left(x_{i21} - \bar{x}_{\underset{\text{Group}}{2}\underset{\text{Var}}{1}}\right)^2$$

$$WQS_2 = \sum_{i=1}^{n_1} \left(x_{i12} - \bar{x}_{12}\right)^2 + \sum_{i=1}^{n_2} \left(x_{i22} - \bar{x}_{22}\right)^2$$

$$WQS_{12} = \sum_{i=1}^{n_1} \left(x_{i11} - \bar{x}_{11}\right)\left(x_{i12} - \bar{x}_{12}\right) + \sum_{i=1}^{n_2} \left(x_{i21} - \bar{x}_{21}\right)\left(x_{i22} - \bar{x}_{22}\right)$$

W aus gepoolten Σ -Matrizen

$$\mathbf{W} = n_1\Sigma_1 + n_2\Sigma_2 + \ldots + n_K\Sigma_K$$

mit $\mathbf{\Sigma}_l$ der Varianz-Kovarianz Matrix in Gruppe l.

Gang der Lösung (DFA)

Max-Bedingung

$$\Gamma = \frac{QS_B}{QS_W} = \frac{\mathbf{v}^t\mathbf{B}\mathbf{v}}{\mathbf{v}^t\mathbf{W}\mathbf{v}} \quad \text{mit} \quad \mathbf{v} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

ist die Darstellung der Quadratsummen der Diskriminanzfunktion y über die quadratische Form mit dem Vektor der b - Koeffizienten

---

Maximierung

$\max(\Gamma)\big|_{\mathbf{v}}$ führt auf $\dfrac{\partial\Gamma}{\partial\mathbf{v}} = 0$ und dies auf

Eigenwert-bedingung

$(\mathbf{B} - \gamma\mathbf{W})\mathbf{v} = \mathbf{0}$ nach Vormultiplizieren mit $\mathbf{W}^{-1}$ auf

$$(\mathbf{W}^{-1}\mathbf{B} - \gamma\mathbf{I})\mathbf{v} = \mathbf{0}$$

was eine Eigenwertbedingung für die Matrix $\mathbf{A} = \mathbf{W}^{-1}\mathbf{B}$ ist.

---

v ist Eigenvektor von A

A ist eine m x m Matrix, also ist **v** allgemein m- stellig. Zu jedem Eigenwert γ ungleich 0 existiert ein Eigenvektor **v**. Die Stellen des **v** Vektors sind die gesuchten Diskriminanzkoeffizienten jeder Diskriminanzfunktion.

Eigenvektoren v

$$\mathbf{v}_1^{\mathbf{t}} = \begin{pmatrix} b_{11} & b_{21} & \cdots & b_{m1} \end{pmatrix} \ldots \mathbf{v}_m^{\mathbf{t}} = \begin{pmatrix} b_{1k} & b_{2k} & \cdots & b_{mk} \end{pmatrix}$$

$$\text{mit} \quad k = \min(K-1, m)$$

Anzahl von v

Es gibt so viele Eigenvektoren **v**, und damit auch so viele Diskriminanzfunktionen, wie die kleinere Zahl aus der Anzahl der Gruppen-1 und der Anzahl der Variablen, m.

Normierung der Diskriminanz-funktion y

Die gepoolte Varianz einer Diskriminanzfunktion erhält man direkt aus der quadratischen Form

$$\hat{s}_y^2 = \frac{1}{N-K}\mathbf{v}^{\mathbf{t}}\mathbf{W}\mathbf{v}$$

Damit kann y direkt nach der Bestimmung normiert werden, indem man

$$\tilde{\mathbf{v}} = \frac{1}{\hat{s}_y}\mathbf{v}$$ als Koeffizientenvektor der normierten Diskriminanzfunktion verwendet: $$\tilde{y} = \frac{1}{\hat{s}_y}\sum_{j=1}^{m} b_j x_j$$

Nicht standardisiert

Sind die Variablen x nicht standardisiert worden, kommt eine additive Konstante hinzu:

$$y = b_0 + \sum_{j=1}^{m} b_j x_j \quad \text{mit} \quad b_0 = -\sum_{j=1}^{m} b_j \bar{x}_j$$

Mehrere DFs (Diskriminanz-raum)

- Sukzessive extrahierte Diskriminanzfunktionen klären absteigend geordnet Diskriminationsvarianz auf.
- Es gilt für die anteilige Varianzaufklärung durch Funktion y_i

$$\frac{QS_{B(i)}}{QS_W + QS_B} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots + \lambda_i + \ldots + \lambda_k}$$

- Alle Diskriminanzfunktionen können auf signifikante Diskriminationsleistung getestet werden (s. z.B. Bortz, 2005, S. 610)
- Alle sukzessiven Diskriminanzfunktionen sind orthogonal.
- Das Prinzip der Aufteilung der Diskriminationsvarianz auf sukzessiv nach Beitrag geordnete und orthogonale Diskriminanz-faktoren ist mit der PCA gut vergleichbar.
- Daraus ergibt sich auch ein vergleichbarer Anwendungszusammen-hang (s.n.)

Anwendung

- Ermittlung relevanter Diskriminationsvariablen.
- Wenn man an einer Reduktion der kritischen Varablen interessiert ist.
- Wenn der Vergleich / die Trennung von Populationen im Vordergrund steht: Benutzt man k -Diskriminanzfunktionen als Eingabedaten für MANOVA oder T^2 Kontraste, wird eine maximale Trennschärfe erreicht, die größer ist als die der k einzelnen Variablen des Sets für k < m.

Einzelfall-Klassifikation

- Kann im Diskriminanzraum mit denselben Verfahren (MDC, QCR, Baysian Classifier) wie üblich gemacht werden.
- Die Einzelfall-Klassifikation wird im vollständigen Diskriminanz- nicht besser als im Variablenraum mit allen Variablen. Vorteile ergeben sich nur, wenn weniger Variablen verwendet werden sollen.
- Die DFA gestattet die Herleitung einfacher Klassifikationsfunktionen mit denen die Fallklassifikation besonders ökonomisch ist.