Aufgaben zu Statistik PLUS

Aufgabe 95

PLUS: HKA

Gegeben ist die folgende Datenmatrix:

$$A = \begin{pmatrix} -9 & -9 & 1 \\ 3 & 1 & 7 \\ 16 & 16 & -2 \\ -12 & -11 & 20 \\ 0 & -2 & -1 \end{pmatrix}$$

Daraus ergibt sich für die Kovarianzmatrix $S = \mathrm{Cov}(A)$ sowie für die Eigenwerte λ_i und Eigenvektoren v_i von S:

$$S = \begin{pmatrix} 97.84 & 94.0 & s_{13} \\ 94.00 & 91.6 & -45.4 \\ s_{31} & -45.4 & s_{33} \end{pmatrix}, \quad \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \approx \begin{pmatrix} 218.6 \\ 36.3 \\ 0.6 \end{pmatrix},$$

$$V = (v_1, v_2, v_3) \approx \begin{pmatrix} -0.66 & 0.25 & 0.71 \\ -0.63 & 0.32 & -0.70 \\ 0.40 & 0.91 & 0.05 \end{pmatrix}$$

a) Berechnen Sie s_{13}, s_{31} und s_{33}.

b) Zeigen Sie, dass näherungsweise λ_2 ein Eigenwert von S zum Eigenvektor v_2 ist.

c) Wieviel Prozent der Information können in den ersten beiden Hauptkomponenten konzentriert werden?

d) Bestimmen Sie die Faktorwertematrix und stellen Sie eine Repräsentation der Daten in einem Koordinatensystem mit den ersten beiden Hauptkomponenten als Achsen dar.

e) Betten Sie die ursprünglichen Merkmalsvektoren in den Faktorwerteplot ein.

a) $S = \begin{pmatrix} \cdot & \cdot & -50 \\ \cdot & \cdot & \cdot \\ -50 & \cdot & 66 \end{pmatrix}$

b) v_2 ist EV von $S \Leftrightarrow S \cdot v_2 = \lambda_2 v_2$

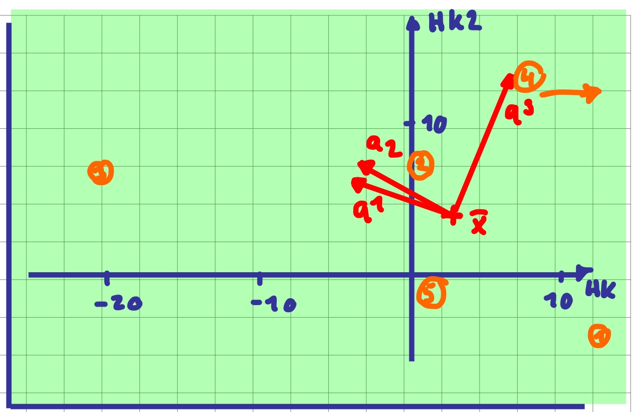
$S \cdot \begin{pmatrix} 0.25 \\ 0.32 \\ 0.91 \end{pmatrix} = \begin{pmatrix} 9.040 \\ 11.498 \\ 33.032 \end{pmatrix} = 36.3 \cdot \begin{pmatrix} 0.2497 \\ 0.3176 \\ 0.9125 \end{pmatrix}$

c) $\dfrac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \dfrac{218.6 + 36.3}{218.6 + 36.3 + 0.6} \approx 99.78\%$

d) $X = A \cdot F = A \cdot \begin{pmatrix} -0.66 & 0.25 \\ -0.63 & 0.32 \\ 0.40 & 0.91 \end{pmatrix} \approx \begin{pmatrix} 12.04 & -4.24 \\ 0.22 & 7.47 \\ -21.50 & 7.33 \\ 22.97 & 11.71 \\ 0.86 & -1.55 \end{pmatrix}$ 1 … 5

$\bar{x}$: 2.92 4.14

e) urspr. Merkmale: $\left.\begin{array}{l} (1\ 0\ 0) \cdot F \\ (0\ 1\ 0) \cdot F \\ (0\ 0\ 1) \cdot F \end{array}\right\} \left.\begin{array}{l} (-0.66\ \ 0.25)\ a^1 \\ (-0.63\ \ 0.32)\ a^2 \\ (0.40\ \ 0.91)\ a^3 \end{array}\right\}$ wg. Skala mal 10



Aufgabe 96

Gegeben ist die folgende Datenmatrix:

$$A = \begin{pmatrix} 1 & 5 \\ 3 & 4 \\ 2 & 2 \\ 4 & 1 \\ 5 & 3 \end{pmatrix}$$

a) Berechnen Sie die Kovarianzmatrix S zu A.

b) Bestimmen Sie die Eigenwerte und Eigenvektoren zu S.

c) Wieviel Prozent der Information kann durch die erste Hauptkomponenten dargestellt werden?

d) Bestimmen Sie die Faktorwertematrix und stellen Sie eine Repräsentation der Daten in einem Koordinatensystem mit beiden Hauptkomponenten als Achsen dar.

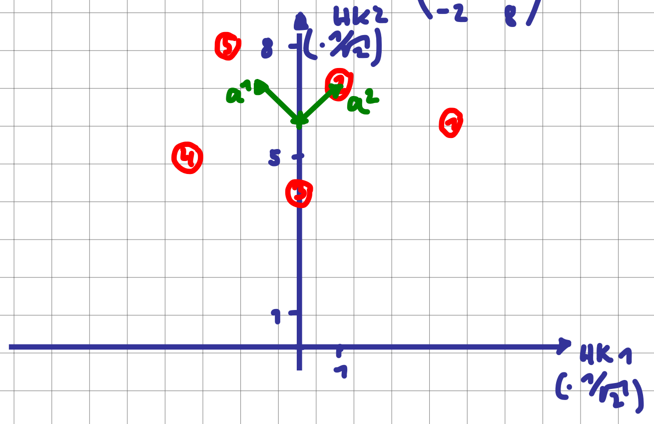
e) Betten Sie die ursprünglichen Merkmalsvektoren in den Faktorwerteplot ein.

a) $S = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$

b) $\Rightarrow \lambda_1 = 3 \,,\ \lambda_2 = 1 \qquad EV_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix},\ EV_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

c) $\frac{3}{3+1} = 0.75$

d) $\frac{1}{\sqrt{2}} \cdot A \cdot \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 4 & 6 \\ 1 & 7 \\ 0 & 4 \\ -3 & 5 \\ -2 & 8 \end{pmatrix}$



zu b) EW zu $S = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$

$\det(S - \lambda E) = (2-\lambda)^2 - ((-1)(-1))$
$= 4 - 4\lambda + \lambda^2 - 1 = \lambda^2 - 4\lambda + 3 = 0$
$\Leftrightarrow \lambda_{1/2} = \begin{cases} 1 \\ 3 \end{cases}$

$\lambda_{1/2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$

$\lambda_{1/2} = \frac{1}{2a}\left(-b \pm \sqrt{b^2 - 4ac}\right)$

$p = \frac{b}{a} \,,\quad q = \frac{c}{a}$

$ax^2 + bx + c = 0$
$\Leftrightarrow \quad x^2 + \frac{b}{a}x + \frac{c}{a} = 0$

Aufgabe 97

Gegeben ist die folgende Datenmatrix mit 4 Objekten, zwei metrischen sowie einem nominalem Merkmal mit den Ausprägungen „a“ bzw. „b“:

$$A = \begin{pmatrix} -2 & -1 & \mathrm{a} \\ 0 & 1 & \mathrm{a} \\ 2 & 3 & \mathrm{b} \\ 3 & 5 & \mathrm{b} \end{pmatrix}$$

Es soll ein lineares Diskrimimnanzkriterium nach dem Fisherschen Ansatz ermittelt werden:

a) Bestimmen Sie dazu die Kovarianzmatrix V der Innerklassenvarianz sowie die Zwischenklassenkovarianzmatrix Z.

b) Berechnen Sie $M = V^{-1}Z$.

c) Bestimmen Sie die für die Klassifikation der beiden Klassen nötigen Eigenwerte und Eigenvektoren von M.

d) Berechnen Sie die Klassenmittelwerte und den Globalmittelwert bzgl. der ersten Diskriminanzfunktion.

e) Bestimmen Sie die Trennebene bzgl. der ersten Diskriminanzfunktion und zeichnen Sie diese zusammen mit den Originaldaten in ein Koordinatensystem mit den beiden metrischen Merkmalen als Achsen ein.

f) Ordnen Sie das neue Objekt $(1, 0)$ mit Hilfe des Diskriminanzkriteriums einer der beiden Klassen „a“ oder „b“ zu.

a) $V = \frac{1}{4}\left(2V^a + 2V^b\right) = \frac{1}{4}\left(2\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + 2\begin{pmatrix}1/4 & 1/2\\ 1/2 & 1\end{pmatrix}\right) = \begin{pmatrix}0.625 & 0.75\\ 0.75 & 1\end{pmatrix}$

$\bar{x}^a = (-1 \;\; 0)$, $\bar{x}^b = (2.5 \;\; 4)$

$\bar{x} = (0.75 \;\; 2)$

$Z = \begin{pmatrix}3.0625 & 3.5\\ 3.5 & 4\end{pmatrix}$

b) $V^{-1} = \begin{pmatrix}16 & -12\\ -12 & 10\end{pmatrix}$

$\Rightarrow M = V^{-1}Z = \begin{pmatrix}7 & 8\\ -1.75 & -2\end{pmatrix}$

c) $\lambda_1 = 5$, $\lambda_2 = 0$

d.h. nur 1. EV ist wichtig:

$(M - \lambda_1 E)g = 0 \iff \begin{pmatrix}2 & 8\\ -1.75 & -7\end{pmatrix}\begin{pmatrix}g_1\\ g_2\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$

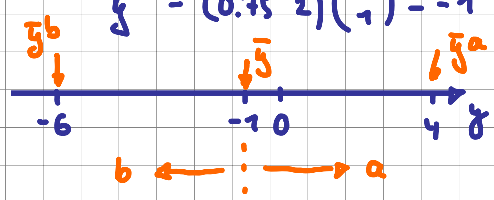
$2g_1 + 8g_2 = 0 \iff g_1 = -4g_2$

z.B. $\left.\begin{matrix} g_1 = -4 \\ g_2 = 1 \end{matrix}\right\}$ $g^1 = \begin{pmatrix}-4\\ 1\end{pmatrix}$

d) $\bar{y}^a = (-1 \;\; 0)\begin{pmatrix}-4\\ 1\end{pmatrix} = 4$

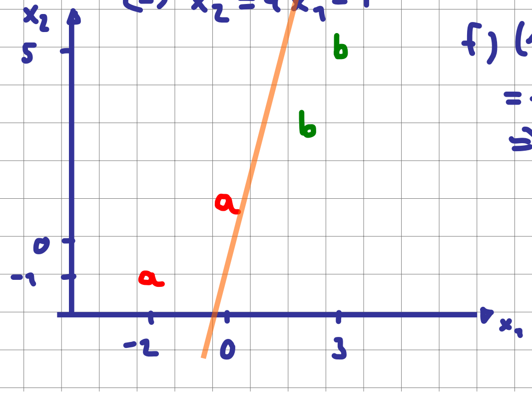
$\bar{y}^b = (2.5 \;\; 4)\begin{pmatrix}-4\\ 1\end{pmatrix} = -6$

$\bar{y} = (0.75 \;\; 2)\begin{pmatrix}-4\\ 1\end{pmatrix} = -1$



e) $y = (x_1 \;\; x_2)\begin{pmatrix}-4\\ 1\end{pmatrix} \iff -1 = -4x_1 + x_2$

$\iff x_2 = 4x_1 - 1$

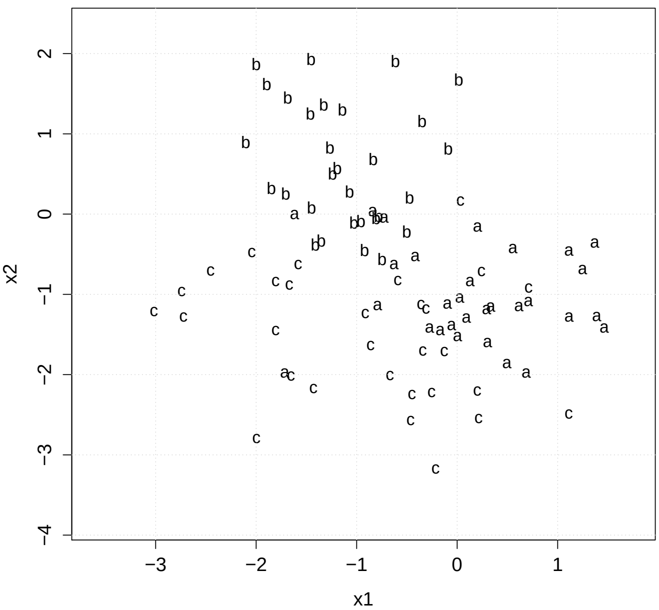


f) $(1 \;\; 0)\begin{pmatrix}-4\\ 1\end{pmatrix} = -4 < -1$

$\Rightarrow$ Gruppe b

Aufgabe 98

In der nebenstehenden Grafik sind 90 Objekte bzgl. der Ausprägungen zweier metrischer Merkmale sowie ihre Zugehörigkeit zu den Klassen a, b, c dargestellt. Diese Daten sollen als Grundlage eines linearen Diskriminanzmodells dienen, um neue Objekte einer der drei Klassen zuzuordnen. Für die Klassenmittelwerte bzw. die zwei nötigen Gewichtungsvektoren g^1, g^2 ergibt sich



Klasse	$\overline{x}_1$	$\overline{x}_2$
a	$0,160$	$-1,018$
b	$-1,123$	$0,626$
c	$-0,933$	$-1,529$

insg. $\quad -0.632 \quad -0.640$

$$\text{bzw.}\quad g^1 = \begin{pmatrix} -0,032 \\ 1,374 \end{pmatrix}, \quad g^2 = \begin{pmatrix} -1,208 \\ -0,320 \end{pmatrix}.$$

R.A. Fisher hat zur Identifikation neuer Objekte mittels dieser Gewichtungsvektoren folgende Entscheidungsregel vorgeschlagen:

> *„Berechne für ein neues $x^T = (x_1, x_2)$ mit $y^1 = x \cdot g^1$ und $y^2 = x \cdot g^2$ pro Klasse $K \in \{a, b, c\}$ den Ausdruck $d(K) = \left(y^1 - \overline{y}_1^K\right)^2 + \left(y^2 - \overline{y}_2^K\right)^2$ und ordne x dann der Klasse K mit dem kleinsten $d(K)$ zu.“*

Dabei bezeichnet $\overline{y}_1^K = \overline{x}^K g^1$ bzw. $\overline{y}_2^K = \overline{x}^K g^2$ den Mittelwert der beiden Diskriminanzvariablen für die Klasse K.

Ordnen Sie das Objekt $x^T = (0, -1)$ gemäß dieser Regel einer der drei Klassen zu.

$y^1 = (0\ \ {-1}) \begin{pmatrix} -0,032 \\ 1,374 \end{pmatrix} = -1.374 \quad , \quad y^2 = (0\ \ {-1}) \begin{pmatrix} -1,208 \\ -0,320 \end{pmatrix} = +0.320$

Klassenmitten $\quad \overline{y}_1^a = (0,160\ \ {-1,018}) \begin{pmatrix} -0,032 \\ 1,374 \end{pmatrix} \approx -1.404$

$\overline{y}_2^a = (0.160\ \ {-1.018}) \begin{pmatrix} -1,208 \\ -0,320 \end{pmatrix} \approx 0.132$

$\overline{y}_1^b = 0.896 \ , \ \overline{y}_2^b = 1.156 \quad ; \quad \overline{y}_1^c = -2.071 \ , \ \overline{y}_2^c = 1.616$

Global: $\left(\overline{y}_1 = \overline{x} \cdot g^1 = -0.859 \quad , \quad \overline{y}_2 = \overline{x} \cdot g^2 = 0.968\right)$

k	$(y^1 - \overline{y}_1^k)^2 + (y^2 - \overline{y}_2^k)^2$
a	0.036
b	5.85
c	2.17

→ Zuordnung von (0,-1) zu a