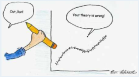
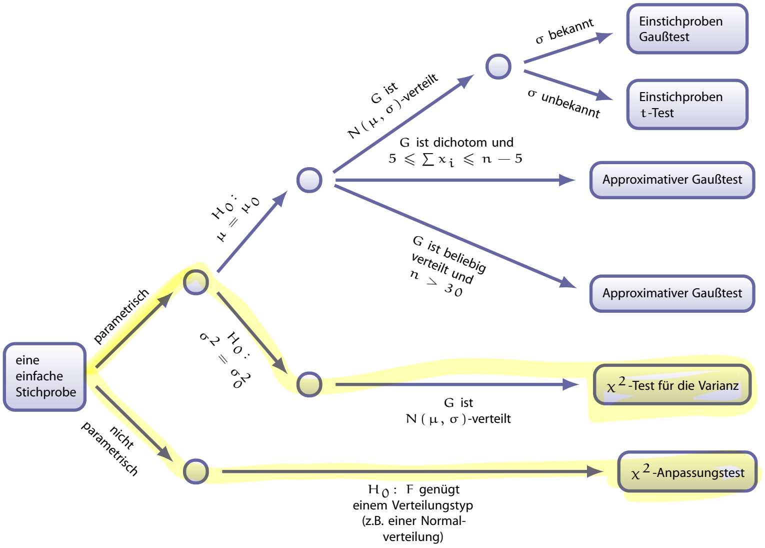
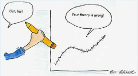


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Signifikanztests bei einer einfachen Stichprobe



(Umfangreichere Übersicht über alle möglichen Fälle siehe **bamberg2011** Seite 171ff.)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Chi-Quadrat-Test für die Varianz

- Gegeben: Einfache Stichprobe $X_1, \ldots, X_n \sim N(\mu; \sigma)$
- Hypothesenpaare:

a)	$H_0: \sigma^2 = \sigma_0^2$	$H_1: \sigma^2 \neq \sigma_0^2$
b)	$H_0: \sigma^2 = \sigma_0^2$ (oder $\sigma^2 \geqq \sigma_0^2$),	$H_1: \sigma^2 < \sigma_0^2$
c)	$H_0: \sigma^2 = \sigma_0^2$ (oder $\sigma^2 \leqq \sigma_0^2$),	$H_1: \sigma^2 > \sigma_0^2$

- **Vorgehensweise:**

Beispiel $\alpha = 0.1$

1. Festlegen des Signifikanzniveaus α.
2. Festlegen des Verwerfungsbereichs:

Fraktile der $\chi^2(n-1)$ Verteilung

$x_{\alpha/2} = x_{0,05} \approx$

$x_{1-\alpha/2} = x_{0,95} \approx$

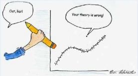
✓ $B = [0; x_{\alpha/2}) \cup (x_{1-\alpha/2}; \infty)$ im Fall a)

$B = [0; x_{\alpha})$ im Fall b)

$B = (x_{1-\alpha}; \infty)$ im Fall c)

3. Berechnung des Testfunktionswertes:

$$v = \frac{(n-1)s^2}{\sigma_0^2} = \frac{1}{\sigma_0^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$$



Beispiel: $G \sim N(\mu; \sigma)$

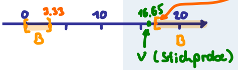
$$(x_1, \ldots, x_{10}) = (2100; 2130; 2150; 2170; 2210; 2070; 2230; 2150; 2230; 2200)$$

Prüfe $H_0: \ \sigma = 40, \ H_1: \ \sigma \neq 40$ zum Signifikanzniveau $\alpha = 0{,}1$

Freiheitsgrade: hier $10-1=9$

Lösung: χ^2-Test für die Varianz, Hypothese Fall a);
Voraussetzungen sind erfüllt

↓ α \ n →	1	2	3	4	5	6	7	8	9
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90
0.4	0.28	1.02	1.87	2.75	3.66	4.57	5.49	6.42	7.36
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34
0.6	0.71	1.83	2.95	4.04	5.13	6.21	7.28	8.35	9.41
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.80	11.03	12.24
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92



1. $\alpha = 0{,}1$
2. $\chi^2(9): \ x_{\frac{\alpha}{2}} = x_{0{,}05} = 3{,}33; \ x_{1-\frac{\alpha}{2}} = x_{0{,}95} = 16{,}92$
   (Tabelle der χ^2-Verteilung)

$$\Rightarrow B = [0; 3{,}33) \cup (16{,}92; \infty)$$

3. $\bar{x} = \frac{1}{10}(2100 + 2130 + \cdots + 2200) = 2164$

$$v = \frac{1}{40^2}\left[(2100 - 2164)^2 + \cdots + (2200 - 2164)^2\right] = 16{,}65$$

$\Rightarrow v \notin B \Rightarrow H_0$ nicht verwerfen

Zwei verbundene einfache Stichproben: Kontingenztest

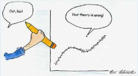
- Situation: In Grundgesamtheit G: Zwei verbundene einfache Stichproben, also Beobachtung zweier Merkmale X, Y
- Hypothese:

$$
\begin{aligned}
H_0 &: \text{ Die beiden Merkmale } X \text{ und } Y \text{ sind in } G \text{ unabhängig.} \\
H_1 &: \ X \text{ und } Y \text{ sind in } G \text{ abhängig.}
\end{aligned}
$$

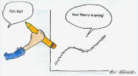
Vorgehensweise Kontingenztest:

1. Festlegen des Signifikanzniveaus α.
2. Unterteilung der x-Achse in $k \geqslant 2$ und die y-Achse in $l \geqslant 2$ disjunkte, aneinander angrenzende Intervalle $A_1, \ldots, A_k$ bzw. $B_1, \ldots, B_l$
3. Erstellen einer Kontingenztabelle mit Randhäufigkeiten:

$x\downarrow \ y\rightarrow$	B_1	B_2	$\cdots$	B_l	
A_1	h_{11}	h_{12}	$\cdots$	h_{1l}	$h_{1\bullet}$
A_2	h_{21}	h_{22}	$\cdots$	h_{2l}	$h_{2\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
A_k	h_{k1}	h_{k2}	$\cdots$	h_{kl}	$h_{k\bullet}$
	$h_{\bullet 1}$	$h_{\bullet 2}$	$\cdots$	$h_{\bullet l}$	n



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen



Vorgehensweise Kontingenztest (Fortsetzung):

Anzahl Zeilen

4. Mit dem Fraktilswert $x_{1-\alpha}$ der χ^2-Verteilung mit $(k-1)\cdot(l-1)$ Freiheitsgraden: Berechnung des Verwerfungsbereichs

Anzahl Spalten

$$B = (x_{1-\alpha}; \infty)$$

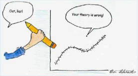
5. Zu jeder Kombination aus $i = 1, \ldots, k$ und $j = 1, \ldots, l$: Berechnung der Größe

$$\tilde{h}_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$$

6. Berechnung des Testfunktionswerts v:

$$v = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(\tilde{h}_{ij} - h_{ij}\right)^2}{\tilde{h}_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{h_{ij}^2}{\tilde{h}_{ij}} - n$$

7. Ablehnung von H_0 genau dann, wenn $v \in B$.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Zwei verbundene einfache Stichproben: Kontingenztest

Kontingenztest: Beispiel

- 400 Erstkandidaten einer praktischen Führerscheinprüfung schneiden abhängig von der besuchten Fahrschule folgendermaßen ab:

	Fahrschule		
	A	B	C
bestanden	130	88	62
durchgefallen	70	38	12

- Zum Signifikanzniveau von 5 % soll getestet werden, ob das Bestehen der Prüfung unabhängig von der besuchten Fahrschule ist.

Testdurchführung

1. Signifikanzniveau $\alpha = 5\,\%$
2. entfällt, da Skalenniveau nominal
3. Kontingenztabelle:

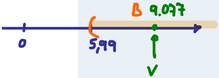
	A	B	C	$\sum$
best.	130	88	62	280
durchg.	70	38	12	120
$\sum$	200	126	74	400

4. Berechnung der $\tilde{h}_{ij}$:

	A	B	C
best.	140	88,2	51,8
durchg.	60	37,8	22,2

5. χ^2-Verteilung mit $(3-1)\cdot(2-1) = 2$ Freiheitsgraden: $x_{1-0,05} = x_{0,95} = 5{,}99$:

$$B = (5{,}99; \infty)$$



6. $$v = \frac{(130-140)^2}{140} + \ldots + \frac{(12-22{,}2)^2}{22{,}2} \approx 9{,}077$$

7. $v \in B$: Also wird H_0 abgelehnt, die Prüfungsergebnisse sind abhängig von der Fahrschule.

Statistik: Table of Contents

1 Statistik: Einführung

2 Deskriptive Statistik

3 Wahrscheinlichkeitstheorie

4 Induktive Statistik

5 Datenanalyse Einleitung

6 Repräsentation

7 Klassifikation

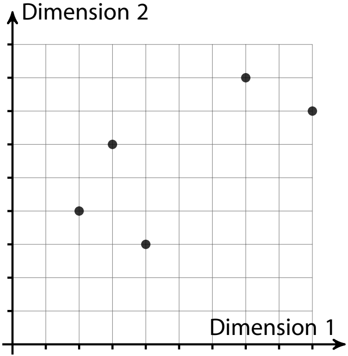
Bildquelle: Newton (2007)

6 Repräsentation
Einführung
Hauptkomponentenanalyse

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Ziele der Repräsentation

- Anordnung der Objekte in einem möglichst niedrig dimensionierten Raum,
- so dass die relative Lage der sich ergebenden Punkte (Objekte) die Ähnlichkeit der Objekte angemessen beschreibt.



- **Aufdeckung von Gruppierungen** leichter (Kontrolle einer Klassifikation)
- Durch **Interpretation der Achsen** evtl. Aufschluss über den Grund der Lage bestimmter Objekte

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Varianten der Repräsentation: Hauptkomponentenanalyse

- Berechnung der Repräsentation aus Varianz-Kovarianz-Strukturen der Datenmatrix

$$A = (a_{ik})_{n \times m} \xrightarrow{\text{Hauptkomponentenanalyse}} X = (x_{ik})_{n \times q}$$

z.B. 205 × 15 205 × 2

- Vorgehen: Datenmatrix wird direkt verarbeitet
- Problem: Lassen sich die m quantitativen Merkmale sinnvoll durch q (meist 2 oder 3) sogenannte Faktoren ersetzen?
- Voraussetzung: Datenmatrix darf nur metrische Merkmale enthalten

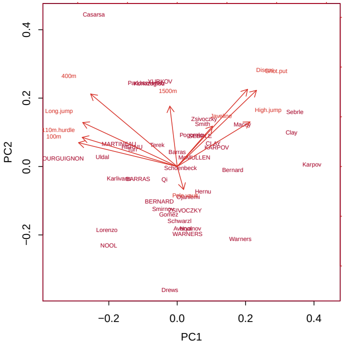
Typische Fragestellungen der **Hauptkomponentenanalyse** in der Marktforschung

- Läßt sich die Vielzahl der Eigenschaften, die die Käufer einer Marke als wichtig empfinden, auf wenige komplexe Faktoren reduzieren?
- Wie lassen sich darauf aufbauend die verschiedenen Marken anhand dieser Faktoren beschreiben?

Repräsentation: Beispiel Hauptkomponentenanalyse


```r
require(FactoMineR)
data(decathlon)
summary(decathlon)

##       100m         Long.jump        Shot.put
##  Min.   :10.4   Min.   :6.61   Min.   :12.7
##  1st Qu.:10.8   1st Qu.:7.03   1st Qu.:13.9
##  Median :11.0   Median :7.30   Median :14.6
##  Mean   :11.0   Mean   :7.26   Mean   :14.5
##  3rd Qu.:11.1   3rd Qu.:7.48   3rd Qu.:15.0
##  Max.   :11.6   Max.   :7.96   Max.   :16.4
##    High.jump        400m        110m.hurdle
##  Min.   :1.85   Min.   :46.8   Min.   :14.0
##  1st Qu.:1.92   1st Qu.:48.9   1st Qu.:14.2
##  Median :1.95   Median :49.4   Median :14.5
##  Mean   :1.98   Mean   :49.6   Mean   :14.6
##  3rd Qu.:2.04   3rd Qu.:50.3   3rd Qu.:15.0
##  Max.   :2.15   Max.   :53.2   Max.   :15.7
##      Discus       Pole.vault      Javeline
##  Min.   :37.9   Min.   :4.20   Min.   :50.3
##  1st Qu.:41.9   1st Qu.:4.50   1st Qu.:55.3
##  Median :44.4   Median :4.80   Median :58.4
##  Mean   :44.3   Mean   :4.76   Mean   :58.3
##  3rd Qu.:46.1   3rd Qu.:4.92   3rd Qu.:60.9
##  Max.   :51.6   Max.   :5.40   Max.   :70.5
##      1500m          Rank          Points
##  Min.   :262   Min.   : 1.0   Min.   :7313
##  1st Qu.:271   1st Qu.: 6.0   1st Qu.:7802
##  Median :278   Median :11.0   Median :8021
##  Mean   :279   Mean   :12.1   Mean   :8005
##  3rd Qu.:285   3rd Qu.:18.0   3rd Qu.:8122
##  Max.   :317   Max.   :28.0   Max.   :8893
##    Competition
##  Decastar:13
##  OlympicG:28
##
##
##
##
```

```r
decathlon$Long.jump = 10 - decathlon$Long.jump
decathlon.pca = prcomp(decathlon[,1:10], scale=TRUE)
round(decathlon.pca$sdev, 3)

##  [1] 1.809 1.318 1.185 1.028 0.828 0.774 0.672 0.630
##  [9] 0.463 0.427

biplot(decathlon.pca,cex=0.5)
```



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Varianten der Repräsentation: Mehrdimensionale Skalierung

- Berechnung der Repräsentation aus der Distanzmatrix

$$A = (a_{ik})_{n \times m} \xrightarrow{\text{Distanz}} D = (d_{ij})_{n \times n} \xrightarrow{\text{MDS}} X = (x_{il})_{n \times q}$$

- Meistens: $q \in \{1,2,3\}$
- Datenmatrix wird nicht direkt verarbeitet; Umweg über Distanzmatrix
- Keine Anforderung an das Skalenniveau der Merkmale

Typische Fragestellungen der MDS in der BWL:

- Welche Produkte einer Gruppe sind sich ähnlich/unähnlich?
- Inwieweit entspricht das eigene Produkt den Idealvorstellungen der Konsumenten?
- Welches Image besitzt eine bestimmte Marke?
- Hat sich die Einstellung der Konsumenten zu einer Marke in den letzten Jahren verändert

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Faktorenanalyse: Grundlegende Idee

- Welche Aussagen lassen sich über die Struktur des Zusammenhangs von Merkmalsvariablen treffen?
- Idee: Darstellung der Struktur in den Daten durch paarweise Korrelationen
- Bei m gemessenen Merkmalen gibt es $\frac{m\cdot(m-1)}{2}$ Korrelationen
- m groß: Analyse des gesamten Variablenkomplexes dann oft schwierig.
- Einschränkung: HKA ist bivariate Analyse; betrachtet also nur jeweils den Zusammenhang je zweier Variablen; Beziehungen zwischen mehr als zwei Merkmalen werden nicht berücksichtigt

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Hauptaufgaben der Faktorenanalyse

- Extraktion von hypothetischen Merkmalen, den sog. Faktoren, aus den Merkmalen und ihren Beziehungen
- Also: Korrelierende Merkmale werden zu Merkmalspaketen zusammengefasst (Faktorenextraktion)
- Damit Reduktion der Komplexität in den Ausgangsdaten: Anstelle von m Variablen wenige Faktoren
- Dabei: Möglichst wenig Informationsverlust
- Durch Datenkompression: Evtl. leichtere Interpretation und Übersicht
- Außerdem: Faktoren sind konstruktionsbedingt unkorreliert und damit qualifiziert als Ausgangsbasis für korrelationssensible Verfahren (z.B. Regression)

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Hauptkomponentenanalyse

- Bevorzugte Variante der Faktorenanalyse bei metrischen Datenmatrizen: Hauptkomponentenanalyse.
- Dabei unterstellt: Lineare Beziehung zwischen den ursprünglich erhobenen und den neu zu bestimmenden hypothetischen Größen (Faktoren).
- Damit: Merkmalsvektoren a^k (Spaltenvektor) lassen sich als Linearkombination von q Faktoren $x^1, \ldots, x^q$ (Spaltenvektoren) darstellen.
- Im Falle $q = m$ immer möglich:

$$a^k = \sum_{i=1}^{m} f_{ki} x^i \quad (k = 1, \ldots, m)$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Faktoren und Ladungsvektoren

- Matriziell: $A = X \cdot F^T$ bzw. mit Koeffizienten:

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1m} \\
\vdots & & \vdots \\
a_{i1} & \cdots & a_{im} \\
\vdots & & \vdots \\
a_{n1} & \cdots & a_{nm}
\end{pmatrix}
=
\begin{pmatrix}
x_{11} & \cdots & x_{1q} \\
\vdots & \vdots & \vdots \\
\vdots & x_{ik} & \vdots \\
\vdots & \vdots & \vdots \\
x_{n1} & \cdots & x_{nq}
\end{pmatrix}
\cdot
\begin{pmatrix}
f_{11} & \cdots & f_{k1} & \cdots & f_{m1} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
f_{1q} & \cdots & f_{kq} & \cdots & f_{mq}
\end{pmatrix}
$$

- Mit:

x_{ip}	als	Faktorwert	f_{kp}	als	Faktorladung
x^p	als	Faktor	f^k	als	Ladungsvektoren
X	als	Faktorwertematrix	F	als	Faktorladungsmatrix

- Damit: Ein Objekt i, dargestellt durch die Zeile i der Datenmatrix A wird nun mit Hilfe der **Zeile i der Faktorwertematrix X** repräsentiert

Hauptkomponentenanalyse: Informationsgehaltt

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

- Problem: Ist diese Darstellung ohne großen Informationsverlust möglich?
- Dazu nötig: Maß für den Informationsgehalt einer Daten- bzw. Faktorwertematrix
- Also: Kovarianzmatrizen $S = \text{Cov}(A)$ und $C = \text{Cov}(X)$ als Basis für Informationsmaß
- Übliche Annahme: Informationsgehalt von Merkmalen/Faktoren ist umso größer, je größer die Varianz dieser Merkmale/Faktoren ist.
- Damit: Maß für Gesamtvariabilität aller Merkmale/Faktoren über Summe der Varianzen aller betrachteten Merkmale/Faktoren
- $\text{Spur}(S)$ bzw. $\text{Spur}(C)$ ermöglicht so einen Vergleich des Informationsgehaltes der beiden Matrizen.
  (Summe der Varianzen entspricht Spur der Kovarianzmatrix)

$$A = \begin{array}{cc} \text{Alter} & \text{Semester} \\ \left(\begin{array}{c} 22 \\ 25 \\ 21 \\ 28 \\ 24 \end{array}\right. & \left.\begin{array}{c} 5 \\ 10 \\ 4 \\ 13 \\ 8 \end{array}\right) \end{array} \Rightarrow \text{Varianz-Kovarianz-matrix} \quad S = \begin{pmatrix} \overset{\sigma_x^2}{6} & \overset{r \cdot \sigma_x \cdot \sigma_y}{8} \\ 8 & \underset{\sigma_y^2}{10{,}8} \end{pmatrix}$$

Var. … Merkmal

Va

Varianz (Alter): $\frac{1}{n} \sum \left(a_i^{Alter} - \bar{a}^{Alter}\right)^2$

Semester: $\frac{1}{n} \sum \left(a_i^{Semester} - \bar{a}_i^{Semester}\right)^2$

$$S = Cov(A) = \begin{pmatrix} Var(Alter) & Cov(Alter, Semester) \\ Cov(Alter, Sem.) & Var(Semester) \end{pmatrix}$$

$$Cov(Alter, Sem.) = \frac{1}{n} \sum \left(a_i^{Al.} - \bar{a}^{Al.}\right)\left(a_i^{Sem.} - \bar{a}^{Sem.}\right)$$

TR: $r \cdot \sigma_x \cdot \sigma_y$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Berechnung des Informationsgehaltes

- Für Datenmatrix A: Kovarianzmatrix $S = (s_{kl})_{m \times m}$:

$$s_{kl} = \frac{1}{n} \sum_{i=1}^{n} (a_{ik} - \bar{a}_{\cdot k})(a_{il} - \bar{a}_{\cdot l}) \qquad \text{mit} \quad \bar{a}_{\cdot k} = \frac{1}{n} \sum_{i=1}^{n} a_{ik}$$

- Maß für die Gesamtinformation von A:

$$\text{Spur } S = \sum_{k=1}^{m} s_{kk} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} (a_{ik} - \bar{a}_{\cdot k})^2$$

- Analog für Faktorwertematrix X: Kovarianzmatrix $C = (c_{kl})_{q \times q}$ und Spur(C):

$$\text{Spur}(C) = \sum_{k=1}^{q} c_{kk} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} (x_{ik} - \bar{x}_{\cdot k})^2$$

Beispiel Informationsgehalt

- Gegeben: Datenmatrix $A = \begin{array}{cc} \text{Alter} & \text{Semester} \\ \left(\begin{matrix} 22 \\ 25 \\ 21 \\ 28 \\ 24 \end{matrix}\right. & \left.\begin{matrix} 5 \\ 10 \\ 4 \\ 13 \\ 8 \end{matrix}\right) \end{array}$
- A soll mit Hilfe zweier unterschiedlicher Faktorkombinationen repräsentiert werden
- Faktorkombination 1 bzw. 2 ist dabei gegeben durch die Faktorladungsmatrizen F' bzw. F'' gemäß

$$F' = \begin{pmatrix} 0{,}6 & 0{,}8 \\ 0{,}8 & -0{,}6 \end{pmatrix} \quad \text{bzw.} \quad F'' = \begin{pmatrix} 0{,}9 & 0{,}2 \\ 0{,}8 & -0{,}5 \end{pmatrix}$$

- **Frage:** Wie gut sind die beiden daraus resultierenden Repräsentationen?

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel - Berechnung Kombination 1

Berechnet man für die Matrix A die Kovarianzmatrix S und die Spur S, so ergeben sich folgende Werte:

Var 1. Merkmal Cov(a^1, a^2)

$$S = \begin{pmatrix} 6 & 8 \\ 8 & 10{,}8 \end{pmatrix} \quad \Rightarrow \text{Spur}(S) = 6 + 10{,}8 = 16{,}8$$

Var 2. Merkmal

Mit Hilfe der Faktorladungsmatrix F' und $X = A \cdot (F'^T)^{-1}$ ergibt sich für X' folgendes:

$$X' = \begin{pmatrix} 17{,}2 & 14{,}6 \\ 23 & 14 \\ 15{,}8 & 14{,}4 \\ 27{,}2 & 14{,}6 \\ 20{,}8 & 14{,}4 \end{pmatrix} \qquad \text{sowie } C = \begin{pmatrix} 16{,}752 & -0{,}064 \\ -0{,}064 & 0{,}048 \end{pmatrix}$$

$$\Rightarrow \text{Spur}(C) = 16{,}752 + 0{,}048 = 16{,}8$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel - Interpretation Kombination 1

Mit Hilfe der neuen Repräsentation kann die Objektmenge **ohne Informationsverlust** dargestellt werden.
(Spur S = Spur $C = 16{,}8$)

Die neue Darstellung hat darüber hinaus den Vorteil, dass die entstandenen Faktoren (nahezu) **unkorreliert** sind, während die ursprünglichen Merkmale eine hohe Korrelation aufweisen. Es gilt nämlich:

$\rightarrow$ Korrelation $(a^1, a^2) = 0{,}9938$

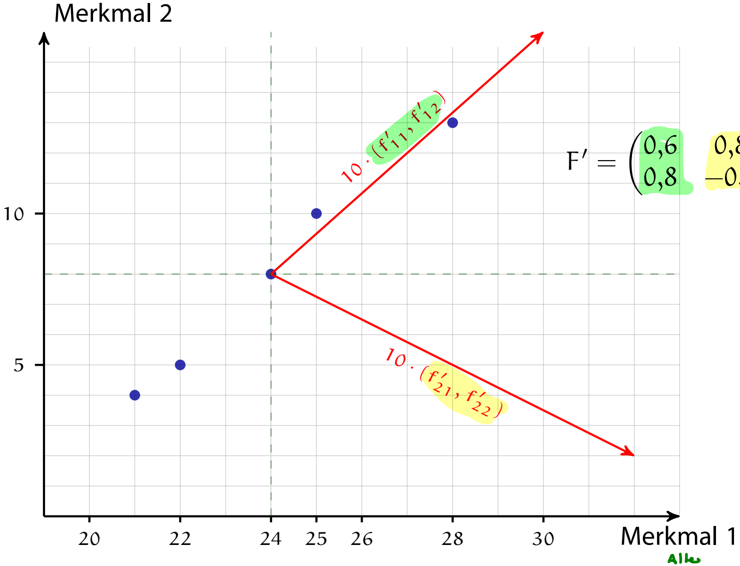
$\rightarrow$ Korrelation $(x^1, x^2) = -0{,}0714$

Beschränkt man sich auf eine Darstellung der Objekte mit Hilfe des ersten Faktors, so können immerhin noch $\frac{c_{11}}{c_{11}+c_{22}} \approx 99{,}71\%$ der Informationen dargestellt werden.

Bsp.: $\frac{16{,}752}{16.752 + 0.048}$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

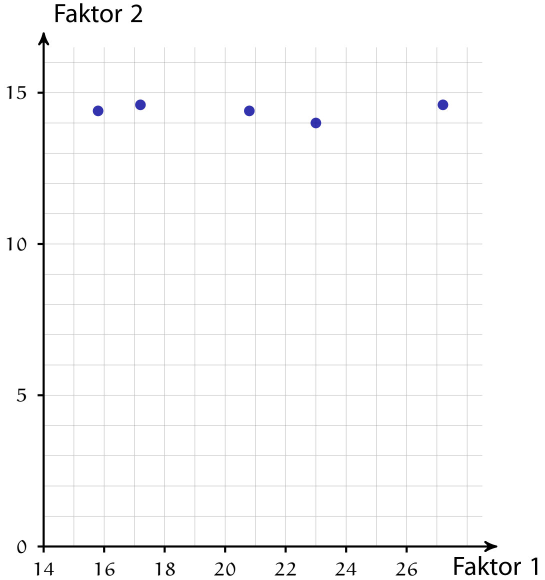
Beispiel: Graphik der Datenmatrix mit Faktorladungen von F'



$$F' = \begin{pmatrix} 0{,}6 & 0{,}8 \\ 0{,}8 & -0{,}6 \end{pmatrix}$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel: Graphik der Datenmatrix mit Faktorladungen von $\mathbb{F}'$



Beispiel - Berechnung Kombination 2

$$F'' = \begin{pmatrix} 0{,}9 & 0{,}2 \\ 0{,}8 & -0{,}5 \end{pmatrix}$$

- Wie oben bereits erwähnt ergibt sich für die Matrix A die Kovarianzmatrix S und die $\text{Spur}(S)$ wie folgt:

$$S = \begin{pmatrix} 6 & 8 \\ 8 & 10{,}8 \end{pmatrix} \quad \Rightarrow \quad \text{Spur}(S) = 6 + 10{,}8 = 16{,}8$$

- Mit Hilfe der Faktorladungsmatrix F'' und $X = A \cdot (F^T)^{-1}$ ergibt sich für X'' folgendes:

$$X'' = \begin{pmatrix} 19{,}67 & 21{,}47 \\ 23{,}77 & 18{,}03 \\ 18{,}52 & 21{,}63 \\ 27{,}21 & 17{,}54 \\ 22{,}29 & 19{,}67 \end{pmatrix} \qquad \text{sowie } C = \begin{pmatrix} 9{,}49 & -5{,}01 \\ -5{,}01 & 2{,}87 \end{pmatrix}$$

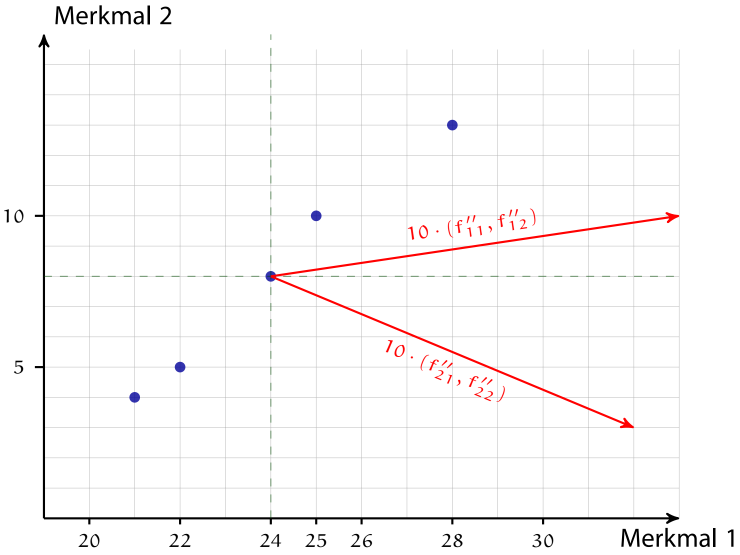
$$\Rightarrow \text{Spur}(C) = 9{,}49 + 2{,}87 = 12{,}36$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel - Interpretation Kombination 2

- Mit Hilfe dieser zweiten Repräsentation kann die Objektmenge **nicht ohne Informationsverlust** dargestellt werden. (Spur $S = 16{,}8$ und Spur$(C) = 12{,}36$)
- Die neue Darstellung hat darüber hinaus den Nachteil, dass die entstandenen Faktoren fast genauso **hoch** (absolut) **korreliert** sind wie die ursprünglichen Merkmale. Es gilt nämlich:
  - Korrelation $(a^1, a^2) = -0{,}9938$
  - Korrelation $(x^1, x^2) = -0{,}9597$
- Somit stellt sich also die Frage, wie die Repräsentation und damit die Faktorladungsmatrix F optimal gewählt werden soll.

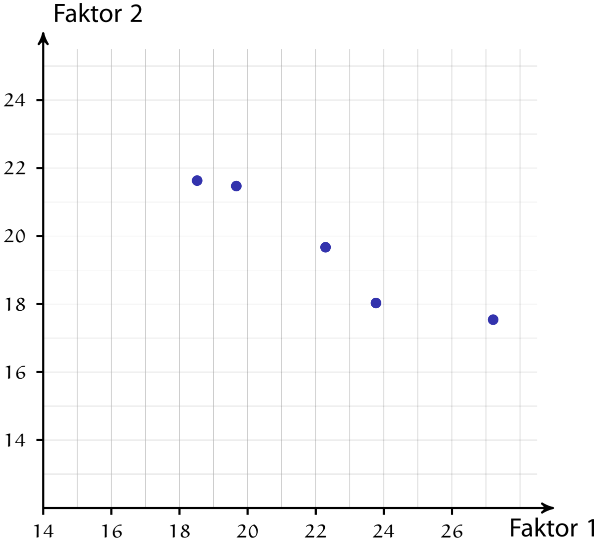
Beispiel: Graphik der Datenmatrix mit Faktorladungen von $\mathbb{F}'$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel: Graphik der Datenmatrix mit Faktorladungen von F''



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Satz 1 der Hauptkomponentenanalyse

Sei $A = X \cdot F^T$ und F orthogonal (d.h. $F^T \cdot F = F \cdot F^T = E$), dann gilt:

1. Spur S = Spur C
2. $C = F^T \cdot S \cdot F$

Bemerkungen:

- Spur S = Spur C $\leftrightarrow$ Information bleibt erhalten
- F orthogonal $\Rightarrow$ $A = X \cdot F^T \Leftrightarrow A \cdot F = X \cdot F^T \cdot F = X$
- $c_{kk} = {f^k}^T \cdot S \cdot f^k$, d.h. der durch den Faktor x^k erklärte Anteil der Varianz hängt nur von f^k ab.

F orthogonal $\Leftrightarrow$ $F^T = F^{-1}$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Hauptkomponentenanalyse: Problem der Faktorenwahl

- Ziel der HKA: Merkmalsreduktion, d.h., man möchte mit wenigen, unkorrelierten Faktoren auskommen und trotzdem einen Großteil der Information darstellen.
- Die durch die Faktoren erklärten Varianzanteile sollen mit wachsendem Index abnehmen, d.h., Faktor x^1 soll den größtmöglichen Varianzanteil erklären, Faktor x^2 den zweitgrößten Anteil ...
- Also: $c_{11} \geq c_{22} \geq c_{33} \geq \ldots \geq c_{qq}$ für $q \leq m$
- Optimierungsproblem:

  1. $\max c_{11} = \max {f^1}^T \cdot S \cdot f^1$ mit ${f^1}^T \cdot f^1 = 1$
  2. $\max c_{22} = \max {f^2}^T \cdot S \cdot f^2$ mit ${f^2}^T \cdot f^2 = 1$ und ${f^2}^T \cdot f^1 = 0$
  3. $\ldots$

$$C = \begin{pmatrix} 16{,}752 & -0{,}064 \\ -0{,}064 & 0{,}048 \end{pmatrix}$$

$$F' = \begin{pmatrix} 0{,}6 & 0{,}8 \\ 0{,}8 & -0{,}6 \end{pmatrix}$$

- Oder Allgemein: Optimierungsproblem $(*)$:

$$\max c_{kk} = \max {f^k}^T \cdot S \cdot f^k \text{ mit } {f^k}^T \cdot f^k = 1$$
$$\text{und } {f^k}^T \cdot f^l = 0 \text{ für } l = 1, \ldots, k-1$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Satz 2 der Hauptkomponentenanalyse

Die Lösung der Optimierungsprobleme $(*)$ impliziert folgendes Eigenwertproblem von S:

$$(S - \lambda \cdot E) \cdot f = 0$$

mit

- $c_{11}, c_{22}, c_{33}, \ldots, c_{mm} \geq 0$ sind Eigenwerte von S
- $f^1, f^2, \ldots, f^q$ mit ${f^k}^T \cdot f^k = 1, {f^k}^T \cdot f^l = 0 (l \neq k)$ sind Eigenvektoren von S
- Für die Matrix C gilt: $C = \begin{pmatrix} c_{11} & & 0 \\ & \ddots & \\ 0 & & c_{qq} \end{pmatrix}$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Konsequenzen aus Satz 2 der HKA

$A = X \cdot F^T \leftrightarrow X = A \cdot F$
$\rightarrow$ Faktoren $x^1 = A \cdot f^1, \ldots, x^m = A \cdot f^m$

C ist Diagonalmatrix
$\rightarrow$ Faktoren sind **paarweise unkorreliert**.

Numerierung der Faktoren $x^1, \ldots, x^m$ derart, dass
$\lambda_1 = c_{11} \geq \lambda_2 = c_{22} \geq \ldots \geq \lambda_m = c_{mm} \geq 0$

$\rightarrow x^1$ erklärt mit $\lambda_1 / \sum \lambda_k$ den größten Anteil der Varianz
$\rightarrow x^2$ erklärt mit $\lambda_2 / \sum \lambda_k$ den zweitgrößten Anteil ...

Bewertung des Informationsverlustes einer Merkmalsreduktion auf $q < m$ Faktoren

$$b(q) = 1 - \frac{\lambda_1 + \ldots + \lambda_q}{\lambda_1 + \ldots \lambda_m} = \left[1 - \frac{c_{11} + \ldots + c_{qq}}{c_{11} + \ldots + c_{mm}}\right] \in [0,1] \text{ und } b(m) = 0$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Vorgehensweise der HKA

Datenmatrix A, maximaler Informationsverlust b_0
Berechne Kovarianzmatrix S
Löse Eigenwertproblem $(S - \lambda \cdot E) \cdot f = 0$ $\lambda_1 = c_{11} \geq \lambda_2 = c_{22} \geq \ldots \geq \lambda_m = c_{mm} \geq 0$ $(f^1, f^2, \ldots, f^m) = F$
Bestimme q minimal, so dass $b(q) = 1 - \frac{\lambda_1 + \ldots + \lambda_q}{\lambda_1 + \ldots + \lambda_m} \leq b_0$
Ladungsmatrix $F_q = (f^1, f^2, \ldots, f^q)$ Faktorwertematrix $X_q = A \cdot F_q = (x^1, x^2, \ldots, x^q)$

Beispiel: 3 Merkmale, 4 Objekte

$$A = \begin{array}{c} \\ \\ \end{array}\begin{array}{ccc} a_1 & a_2 & a_3 \\ \downarrow & \downarrow & \downarrow \end{array}$$

$$A = \begin{pmatrix} 6 & 0 & 0 \\ 8 & 8 & 4 \\ 0 & 4 & 0 \\ 2 & 8 & 8 \end{pmatrix} \Rightarrow S = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 11 & 9 \\ 0 & 9 & 11 \end{pmatrix}$$

$Cov(a_1,a_1) = Var[a_1]$, $\to Cov(a_1,a_2)$, $\to Cov(a_1,a_3)$

Eigenwerte: $\det(S - \lambda E) = \det\begin{pmatrix} 10-\lambda & 0 & 0 \\ 0 & 11-\lambda & 9 \\ 0 & 9 & 11-\lambda \end{pmatrix}\begin{array}{cc} 10-\lambda & 0 \\ 0 & 11-\lambda \\ 0 & 9 \end{array}$

$$= (10-\lambda)(11-\lambda)^2 - 9^2 \cdot (10-\lambda) = 0$$

$$\Leftrightarrow (10-\lambda)\cdot[121 - 22\lambda + \lambda^2 - 81] = 0$$

$$\Leftrightarrow (10-\lambda)\cdot[\lambda^2 - 22\lambda + 40] = 0$$

$$\Leftrightarrow (10-\lambda)(\lambda-2)(\lambda-20) = 0$$

$$\Rightarrow \lambda_1 = 20, \quad \lambda_2 = 10, \quad \lambda_3 = 2$$

wähle max. 10% Informationsverlust

probiere 1 Hauptkomponente:

$\Rightarrow$ Info. gehalt $\frac{20}{20+10+2} = 0{,}625$

(Verlust: $1 - 0{,}625 = 0{,}375 \to$ zu viel)

2 Hauptkomponenten: $\frac{20+10}{20+10+2} = \frac{30}{32} = 0{,}9375$

(Verlust: $1 - 0{,}9375 = 0{,}0625$ ✓)

Also berechne 2 H.K. (Eigenvektoren zu EW. λ_1, λ_2)

Eigenvektor zu $\lambda_1 = 20$:

$$(S - \lambda_1 \cdot E)\cdot x = 0$$

$$\Leftrightarrow \begin{pmatrix} 10-20 & 0 & 0 \\ 0 & 11-20 & 9 \\ 0 & 9 & 11-20 \end{pmatrix}\cdot\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} -10 & 0 & 0 \\ 0 & -9 & 9 \\ 0 & 9 & -9 \end{pmatrix}\cdot\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

$$\Leftrightarrow x_1 = 0, \quad -x_2 + x_3 = 0 \Leftrightarrow x_2 = x_3$$

$\Rightarrow$ EV: $\begin{pmatrix} 0 \\ a \\ a \end{pmatrix}$ wähle a so, dass EV Länge 1 hat

$\Rightarrow 0^2 + a^2 + a^2 = 1$

$\Leftarrow a = 1/\sqrt{2}$

EV zu $\lambda_2 = 10$:

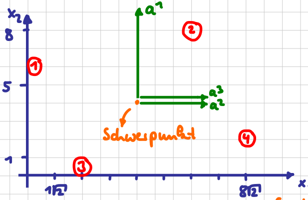
$$(S - \lambda_2 E)\cdot x = \begin{pmatrix} 10-10 & 0 & 0 \\ 0 & 11-10 & 9 \\ 0 & 9 & 11-10 \end{pmatrix}\cdot x = 0$$

$$x_2 + 9x_3 = 0, \quad 9x_2 + x_3 = 0 \Rightarrow x_2 = x_3 = 0$$

$$\Rightarrow EV = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Repräsentationsmatrix X

$$X = A \cdot \underset{\substack{\text{Matrix} \\ \text{der Eigenvektoren}}}{F} = \underset{4\times 3}{\begin{pmatrix} 6 & 0 & 0 \\ 8 & 8 & 4 \\ 0 & 4 & 0 \\ 2 & 8 & 8 \end{pmatrix}} \cdot \underset{3\times 2}{\begin{pmatrix} 0 & 1 \\ 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \end{pmatrix}} = \begin{pmatrix} 0 & 6 \\ 6\sqrt{2} & 8 \\ 2\sqrt{2} & 0 \\ 8\sqrt{2} & 2 \end{pmatrix}$$



(erste Haupt-komponente)

Einbetten der Originalmerkmale:

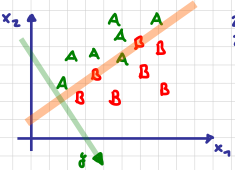
$$a^1: \quad (1 \;\; 0 \;\; 0)\cdot F = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$a^2: \quad (0 \;\; 1 \;\; 0)\cdot F = \begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}$$

$$a^3: \quad (0 \;\; 0 \;\; 1)\cdot F = \begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}$$

für Zeichnung! mal 5: $\begin{pmatrix} 0 \\ 5 \end{pmatrix}$, $\begin{pmatrix} 3.5 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 3.5 \\ 0 \end{pmatrix}$

Lineare Diskriminanzanalyse



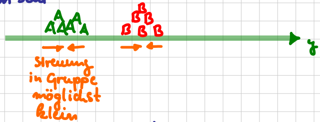
2 Klassen, 2 unabh. Merkmale

Gesucht: Projektion von x_1 und x_2 auf Gerade, so dass Klassen A und B optimal getrennt werden

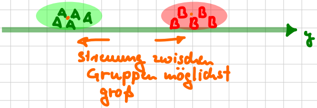
$$y = x_1 \cdot g_1 + x_2 \cdot g_2$$

Kriterien:

① Innerklassenvarianzen^{w} sollen möglichst klein sein



② Zwischenklassenvarianz^{b} soll möglichst groß sein



$\Rightarrow$ Diskriminanzkriterium von Fisher

$$\lambda(g) = \frac{b}{w} \rightarrow \max$$

$$b = g^T Z g \quad , \quad w = g^T V g$$

Z: Zwischen-Klassen Kov. matrix; V: Inner Klassen-Kov. matrix

$$\Rightarrow \;\; \dots \;\; \Rightarrow \;\; \boxed{\underbrace{V^{-1} \cdot Z}_{\text{Matrix}} \cdot g = \underbrace{\lambda}_{\text{Zahl}} \cdot g}$$

$\Rightarrow$
- Eigenwerte von $V^{-1}Z$ maximieren $\lambda(g)$
- Eigenvektoren sind die Gewichte g

Beispiel:

4	1	A
2	4	A
2	3	A
3	6	A
4	4	A
9	10	B
6	8	B
9	5	B
8	7	B
10	8	B

Innerklassenkov.matrix V:

TR

$$\mathrm{Cov}(A) = \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{pmatrix} = V^A$$

$$\mathrm{Cov}(B) = \begin{pmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{pmatrix} = V^B$$

$$\Rightarrow V = \frac{1}{10}\left(5\cdot V^A + 5\cdot V^B\right) = \begin{pmatrix} 1.32 & -0.22 \\ -0.22 & 2.64 \end{pmatrix}$$

Zwischenklassen kov. matrix Z

Mittelwerte:
- A: $\bar{x}^A = (3{,}0 \quad 3{,}6)$
- B: $\bar{x}^B = (8{,}4 \quad 7{,}6)$
- gesamt: $\bar{x} = (5.7 \quad 5.6)$

für TR:

		Freq.
3.0	3.6	5
8.4	7.6	5

Cov. = $r\cdot x_s\cdot y_s$
$r\cdot s_x\cdot s_y$

$$\Rightarrow Z = \begin{pmatrix} 7.29 & 5.4 \\ 5.4 & 4.0 \end{pmatrix}$$

Exkurs: Invertieren einer 2x2 Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\Rightarrow V^{-1} \approx \begin{pmatrix} 0.768 & 0.064 \\ 0.064 & 0.384 \end{pmatrix}$$

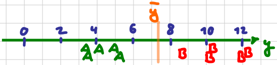
$$\Rightarrow M = V^{-1} Z = \begin{pmatrix} 5.946 & 4.405 \\ 2.541 & 1.882 \end{pmatrix}$$

$\Rightarrow$ Eigenwerte: $\lambda_1 = 7{,}82$, $\lambda_2 = 0$

Eigenvektoren $\begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$

Berechne $\hat{Y} = X\cdot g$

$$\hat{Y} = X\cdot\begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix} = \left.\begin{matrix} 4.07 \\ 3.41 \\ 3.02 \\ 5.12 \\ 5.25 \end{matrix}\right\} A \qquad \left.\begin{matrix} 12.2 \\ 8.66 \\ 10.24 \\ 10.11 \\ 12.34 \end{matrix}\right\} B$$



$\bar{y} = \bar{x}\cdot\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 7.44$

$\bar{y}_A = \bar{x}_A\cdot\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 4.17$

$\bar{y}_B = \dots = 10.71$

neues Objekt soll zugeordnet werden: z. B. $x = (7 \quad 3)$

$\Rightarrow y = (7 \quad 3)\begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix} \approx 7.62$
$> 7.44 \rightarrow$ B

Beispiel: 3 Merkmale, 4 Objekte

$$A = \begin{array}{c} \begin{array}{ccc} a^1 & a^2 & a^3 \end{array} \\ \begin{pmatrix} 6 & 0 & 0 \\ 8 & 8 & 4 \\ 0 & 4 & 0 \\ 2 & 8 & 8 \end{pmatrix} \end{array} \Rightarrow S = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 11 & 9 \\ 0 & 9 & 11 \end{pmatrix}$$

$Cov(a^1, a^1) = Var[a^1]$

$Cov(a^1, a^2) = r \cdot \sigma_{a^1} \cdot \sigma_{a^2}$

$Cov(a^1, a^3)$

Eigenwerte:

$$\det(S - \lambda E) = \det \begin{pmatrix} 10-\lambda & 0 & 0 \\ 0 & 11-\lambda & 9 \\ 0 & 9 & 11-\lambda \end{pmatrix} = 0$$

$$\Leftrightarrow (10-\lambda)(11-\lambda)^2 - 9 \cdot 9 \cdot (10-\lambda) = 0$$

$$\Leftrightarrow (10-\lambda)[121 - 22\lambda + \lambda^2 - 81] = 0$$

$$\Leftrightarrow (10-\lambda) \cdot [\lambda^2 - 22\lambda + 40] = 0$$

$$\Leftrightarrow (10-\lambda)(\lambda-2)(\lambda-20) = 0$$

$$\Rightarrow \lambda_1 = 20, \quad \lambda_2 = 10, \quad \lambda_3 = 2$$

maximal 10% Infoverlust: nur 1 Hauptkomp.:

Gesamt inform.: $\frac{20}{20+10+2} = \frac{20}{32} = 0{,}625$

(Verlust: $1 - 0{,}625 = 0.375$

2 Hauptk.: $\frac{20+10}{32} = \frac{30}{32} = 0{,}9375$

(Verlust: $1 - 0.9375 = 0.0625$ ✓)

Also: Berechne Eigenvektoren zu λ_1, λ_2:

EV zu λ_1:

$$(S - \lambda_1 E) \cdot x = 0$$

$$\Leftrightarrow \begin{pmatrix} 10-20 & 0 & 0 \\ 0 & 11-20 & 9 \\ 0 & 9 & 11-20 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

$$\Leftrightarrow \begin{pmatrix} -10 & 0 & 0 \\ 0 & -9 & 9 \\ 0 & 9 & -9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \Leftrightarrow \left.\begin{matrix} x_1 = 0 \\ x_2 = x_3 \end{matrix}\right\}$$

$$\Rightarrow \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \xrightarrow[\text{normiere mit } \sqrt{2} \; (\sqrt{0^2+1^2+1^2})]{} \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = f^1$$

EV zu $\lambda_2 = 10$:

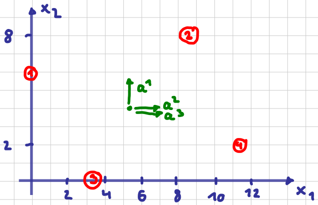
$$\begin{pmatrix} 10-10 & 0 & 0 \\ 0 & 11-10 & 9 \\ 0 & 9 & 11-10 \end{pmatrix} x = 0$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & 9 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \quad \left\}\begin{matrix} x_1 \text{ beliebig} \\ x_2 + 9x_3 = 0 \\ 9x_2 + x_3 = 0 \end{matrix}\right.$$

$$\Rightarrow x_1 = 1, \quad x_2 = x_3 = 0$$

$$\Rightarrow f^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$X = A \cdot F = \begin{pmatrix} 6 & 0 & 0 \\ 8 & 8 & 4 \\ 0 & 4 & 0 \\ 2 & 8 & 8 \end{pmatrix} \cdot \begin{array}{c} \begin{array}{cc} f^1 & f^2 \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \end{pmatrix} \end{array} = \begin{pmatrix} 0 & 6 \\ 6\sqrt{2} & 8 \\ 2\sqrt{2} & 0 \\ 8\sqrt{2} & 2 \end{pmatrix}$$



$$\begin{pmatrix} 0 & 6 \\ 6\sqrt{2} & 8 \\ 2\sqrt{2} & 0 \\ 8\sqrt{2} & 2 \end{pmatrix} \begin{array}{l} 1 \\ 2 \\ \\ \\ \end{array}$$

Originalmerkmale einbetten: $a^1: \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \cdot F = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$a^2: \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \cdot F = \begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}$

$a^3: \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \cdot F = \begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel 1 - Hauptkomponentenanalyse

Für die Produkte P_1, P_2, P_3 und P_4 konnten bezüglich der Merkmale M_1, M_2 und M_3 folgende Werte erhoben werden:

$$A = \begin{pmatrix} 6 & 0 & 0 \\ 8 & 8 & 4 \\ 0 & 4 & 0 \\ 2 & 8 & 8 \end{pmatrix}$$

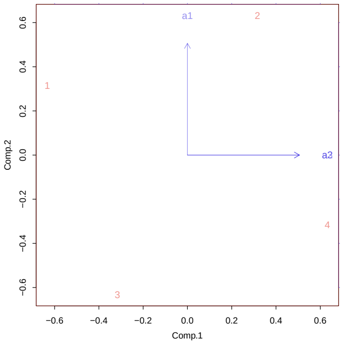
- Berechnen Sie die Kovarianzmatrix der Merkmale und interpretieren Sie diese.
- Lösen Sie das zugehörige Eigenwertproblem und interpretieren Sie die Ergebnisse.
- Bestimmen Sie die Faktorwertematrix und stellen Sie diese zweidimensional dar.
- Betten Sie die ursprünglichen Merkmalsvektoren in den Faktorwerteplot ein.

Ergebnisse Beispiel

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

```
Farben = c("#dd100070", "#1000dd70'
# Beispiel 1 HKA
A <- data.frame(
  a1=c(6,8,0,2),
  a2=c(0,8,4,8),
  a3=c(0,4,0,8))
A

##   a1 a2 a3
## 1  6  0  0
## 2  8  8  4
## 3  0  4  0
## 4  2  8  8

S <- cov(A)*3/4
S

##    a1 a2 a3
## a1 10  0  0
## a2  0 11  9
## a3  0  9 11

F <- eigen(S)$vectors
round(F,3)

##       [,1] [,2]   [,3]
## [1,] 0.000    1  0.000
## [2,] 0.707    0 -0.707
## [3,] 0.707    0  0.707
```

```
# Beispiel 1 HKA
L <- eigen(S)$values
L

## [1] 20 10  2

A.pca <- princomp(A)
summary(A.pca)

## Importance of components:
##                         Comp.1 Comp.2 Comp.3
## Standard deviation      4.472 3.1623 1.4142
## Proportion of Variance  0.625 0.3125 0.0625
## Cumulative Proportion   0.625 0.9375 1.0000

biplot(A.pca, col=Farben)
```



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel 2 - Hauptkomponentenanalyse

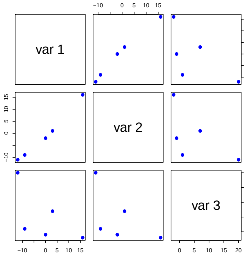
Gegeben ist die folgende Datenmatrix:

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

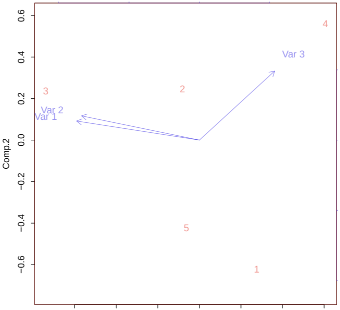
- Berechnen Sie die Kovarianzmatrix der Merkmale und interpretieren Sie diese.
- Lösen Sie das zugehörige Eigenwertproblem und interpretieren Sie die Ergebnisse.
- Bestimmen Sie die Faktorwertematrix und stellen Sie diese zweidimensional dar.
- Betten Sie die ursprünglichen Merkmalsvektoren in den Faktorwerteplot ein.

Ergebnisse Beispiel

```
# Beispiel 2 HKA
S=cov(A)*(n-1)/n
pairs(A,pch=19,col="blue")
F <- eigen(S)$vectors
round(F,3)

##        [,1]  [,2]   [,3]
## [1,] -0.660 0.252  0.708
## [2,] -0.633 0.320 -0.705
## [3,]  0.405 0.913  0.052
```




```
# Beispiel 1 HKA
L <- eigen(S)$values
L

## [1] 218.6298  36.2523   0.5579

A.pca <- princomp(A)
summary(A.pca)

## Importance of components:
##                          Comp.1 Comp.2   Comp.3
## Standard deviation      14.7861 6.0210 0.746909
## Proportion of Variance   0.8559 0.1419 0.002184
## Cumulative Proportion    0.8559 0.9978 1.000000

biplot(A.pca, col=Farben)
```



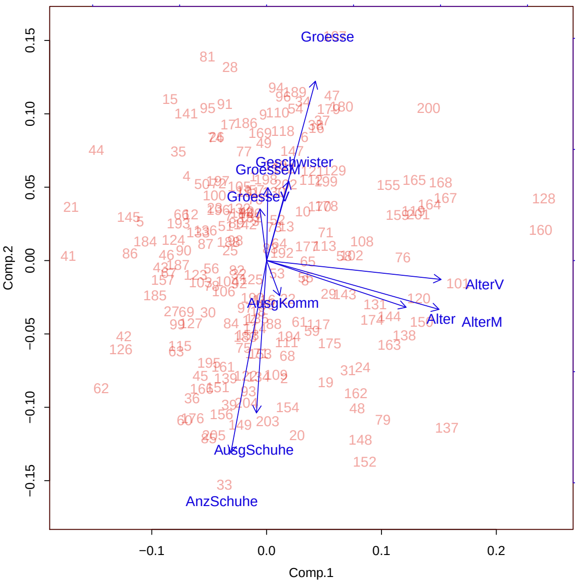
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Beispiel Umfrage

```
# Beispiel 2 HKA
A = MyData[,c(1,2,4:8,10:12)]
str(A)

## 'data.frame': 205 obs. of  10 variables:
##  $ Alter      : int  21 20 19 20 20 24 20 27 23 21 ...
##  $ Groesse    : int  173 164 172 168 169 185 170 165 184 178 ...
##  $ AlterV     : int  54 57 49 45 43 54 49 53 52 55 ...
##  $ AlterM     : int  51 57 58 49 42 52 53 53 48 55 ...
##  $ GroesseV   : int  187 172 193 185 182 179 182 175 182 180 ...
##  $ GroesseM   : int  170 166 162 164 164 163 172 165 175 168 ...
##  $ Geschwister: int  1 0 3 3 5 2 2 1 2 1 ...
##  $ AusgKomm   : num  156 450 240 35.8 450 250 100 300 450 1300 ...
##  $ AnzSchuhe  : int  17 22 15 15 22 8 20 10 3 7 ...
##  $ AusgSchuhe : int  50 500 400 100 450 90 250 200 300 200 ...

MD.pca = princomp(A, cor=TRUE)
summary(MD.pca, digits=3)

## Importance of components:
##                          Comp.1 Comp.2 Comp.3 Comp.4
## Standard deviation     1.4558 1.3194 1.2384 1.0462
## Proportion of Variance 0.2119 0.1741 0.1534 0.1095
## Cumulative Proportion  0.2119 0.3860 0.5394 0.6488
##                          Comp.5   Comp.6   Comp.7   Comp.8
## Standard deviation     0.96679 0.91429 0.77335 0.71009
## Proportion of Variance 0.09347 0.08359 0.05981 0.05042
## Cumulative Proportion  0.74231 0.82590 0.88571 0.93613
##                          Comp.9 Comp.10
## Standard deviation     0.64469 0.47232
## Proportion of Variance 0.04156 0.02231
## Cumulative Proportion  0.97769 1.00000
```


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

```
# Beispiel 1 HKA
Farben = c("#dd100060", "#1000ddff")
biplot(MD.pca, col=Farben, xlim=c(-0.172,0.25), ylim=c(-0.17,0.16))
```



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Hauptkomponentenanalyse: Interpretation

Interpretationshilfen

- **Aussage über den Zusammenhang** r_{kp} zwischen den ursprünglichen Merkmalen a^k und den Hauptkomponenten (Faktoren) x^p z.B. wie folgt:

$$r_{kp} = \frac{\frac{1}{n}\sum_{i=1}^{n}(a_{ik} - \bar{a}_{\cdot k})(x_{ip} - \bar{x}_{\cdot p})}{\sqrt{s_{kk}}\sqrt{c_{pp}}}, \qquad \begin{array}{l} k = 1, \ldots, m \\ p = 1, \ldots, q \end{array}$$

- Korrelationskoeffizient r_{kp}: Zusätzlicher Anhaltspunkt bei der Interpretation der Analyseergebnisse

Kommunalitäten

- Mit **Korrelationskoeffizienten** r_{kp}: Anteil der auf die ersten q Faktoren $(p = 1, ..., q)$ übertragene Information des k-ten Merkmals a^k berechenbar
- Dazu: Kommunalität k_k mit

$$k_k = \sum_{p=1}^{q} r_{kp}^2 \quad \in \quad [0,1] \;\text{ für }\; k = 1, \ldots, m$$

- k_k gibt an, wieviel Prozent der Informationen des (standardisierten) Merkmals a^k noch in den Faktoren $x^1 \ldots x^q$ enthalten sind.

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Wahl der Repräsentationsdimension

Allgemeingültige Kriterien zur Bestimmung einer problem-adäquaten **Repräsentationsdimension** q sind nicht bekannt. Man kann sich aber an den folgenden Regeln orientieren:

- Man wählt ein q aus, bei dem man bei weiterer Reduktion der Faktorenzahl einen verhältnismäßig hohen zusätzlichen Informationsverlust hätte (Ellenbogenkriterium).
- Man wählt ein minimales $q \geq 1$ mit $b(q) < b_{\text{max}}$, d.h. ein q mit maximal zu akzeptierendem Informationsverlust , (mit z.B.: $b_0 = 10\%$ oder $b_0 = 25\%$).
- Man wählt ein maximales $q \leq m$ mit $\lambda_q \geq (\lambda_1 + \ldots + \lambda_m)/m$, d.h. man gibt vor, dass jeder Faktor mindestens die durchschnittliche Merkmalsvarianz erklären soll (Kaiser-Kriterium).

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Schlußbemerkungen zur Hauptkomponentenanalyse

- Die HKA hat einige Vorteile:
- Statt Untersuchungsobjekte durch kaum übersehbare und hochkorrelierte Merkmalsbatterien zu beschreiben, gelangt man zu **wenigen und wichtigen** und weitgehend orthogonalen **Dimensionen** des Merkmalsraumes,
- in dem Untersuchungsobjekte anhand ihrer Faktorwerte positioniert sind

Fehlinterpretationsmöglichkeiten I

- Die HKA projiziert mit Hilfe einer linearer Abbildung
  - die m-dimensionalen Untersuchungsobjekte (unter Verwendung der Matrix X_q) und
  - die m Merkmale a^k (unter Verwendung der Matrix F_q)

  in einen gemeinsamen q-dimensionalen Teilraum.
- Deswegen: Sowohl die Objekte als auch die Merkmalsvektoren dürfen nur **relativ zueinander interpretiert** werden.
- Absolute Aussagen über die Lage der Objekte hinsichtlich der einzelnen Merkmalsvektoren gelten nur approximativ.

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Fehlinterpretationsmöglichkeiten II

- Wesentlich: Zahl und Art der Merkmale, die in die FA eingehen. Aus den entsprechenden Gleichungen wird deutlich, daß auf einem Faktor jene Merkmale **hoch laden**, die auch **hoch korreliert** sind. Wenn ein Faktor durch sehr viele gleichartige Merkmale vertreten ist, so laden diese auch hoch auf ihm.
- Die **Extraktion** des ersten Faktors nach dem Kriterium des **höchsten Varianzbeitrages** wird dann als bedeutendsten Faktor jenen präsentieren, der durch viele Merkmale vertreten ist.
- Damit zeigt sich aber, daß die FA für sich allein genommen nicht in der Lage ist, die Wichtigkeit von Merkmalsbereichen zu ermitteln ($\longrightarrow$ **Conjointanalyse**)
- Bei Wahl von Merkmalen aus unterschiedlichen Bereichen (z.B. soziodemographische und psychographische Merkmale), so kommt es vor, dass die unterschiedlichen Merkmale auf ein und denselben Faktor laden. Interpretation des Faktors dann wenig sinnvoll bzw. schwierig.
- Gängig: Faktoren nach dem am höchsten ladenden Merkmalen benennen (Leitvariablenkonzept). Führt unter Umständen zu einer Vernachlässigung relevanter Informationen für die Beschreibung der Untersuchungsobjekte.

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
Einführung
Hauptkomponentenanalyse

Kovarianz- oder Korrelationsmatrix?

- Die Hauptkomponentenanalyse kann statt mittels der Kovarianzmatrix von A auch auf Basis der Korrelationsmatrix von A durchgeführt werden.
- Aber: die Eigenwerte und Eigenvektoren der Kovarianzmatrix können nicht in die der Korrelationsmatrix überführt werden und geben deshalb unterschiedliche Informationen.
- Vorteile der Korrelationsmatrix:
  - Die Ergebnisse zweier Analysen können direkt miteinander verglichen werden
  - Hauptkomponentenanalyse basierend auf Kovarianzmatrizen ist sehr sensitiv bzgl. der Einheiten der Merkmale
- Vorteil der Kovarianzmatrix:
  - Hauptkomponentenanalyse basierend auf Kovarianzmatrizen ist sehr sensitiv bzgl. der Einheiten der Merkmale (Manchmal möchte man diesen Umstand benutzen)

Statistik: Table of Contents

1 Statistik: Einführung

2 Deskriptive Statistik

3 Wahrscheinlichkeitstheorie

4 Induktive Statistik

5 Datenanalyse Einleitung

6 Repräsentation

7 Klassifikation

Bildquelle: Newton (2007)

7 Klassifikation

- Problemstellung
- Beispielproblem
- Lineare Diskriminanzanalyse
- Klassifikationsbäume
- Support-Vector-Machines

Kreditausfälle

- Mai 2009: Kreditkartenkonzern American Express Co.: Kreditausfallraten im April 2009 erstmals auf mehr als 10 Prozent
- Herbst 2013: Commerzbank: Kreditausfallrate bei Mittelstandskrediten stark gestiegen
- Typisches Problem: Wie kann bekannte Information über Kreditantragsteller benutzt werden, um Bonität zu beurteilen?

Problemstellung in der Klassifikation

- Ausganssituation: Daten von Objekten, bei denen jeweils die Zugehörigkeit zu einer (von mehreren möglichen) Gruppen bekannt ist
- Problem: Bei einem neuen Objekt fehlt Information zur Gruppenzugehörigkeit
- Aufgabe: Mittels der bekannten Daten ein Modell bestimmen, dass die Gruppenzugehörigkeit des neuen Objektes aufgrund der vorhandenen Informationen ermöglicht

Genauer:

- Abhängiges Merkmal Y (Gruppenzugehörigkeit) soll mit Hilfe der unabhängigen Merkmale $X_1, X_2, \ldots, X_m$ erklärt (= klassifiziert, identifiziert) werden.
- Dazu nötig: Bestimmung einer Funktion $f(x_{i1}, ..x_{im})$, die Objekte i ihrer „richtigen“ Klasse k_i zuordnet.
- Unabhängige Merkmale können sowohl metrisch als auch nominal interpretierbar sein.

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

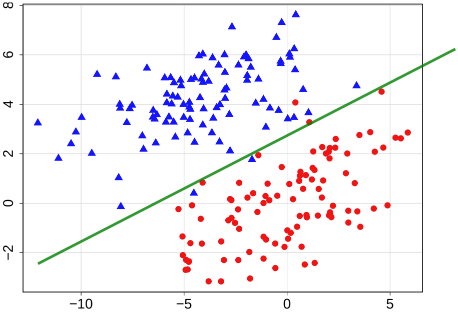
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Klassifikationsmodelle: Varianten

Lineare Trennfunktion

- Manche Methoden erzeugen Trennmodelle aus linearen Funktionen:
- Dann Aufgabe: Ermittlung von „Gewichten“ $\alpha_1, \alpha_2, \ldots, \alpha_m$ und Merkmale $X_1, X_2, \ldots, X_m$ mit

$$Y = \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \ldots + \alpha_m \cdot X_m$$



Lineare Trennung möglich

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Nichtlineare Modelle

- Für manche Situationen hilfreich: Nichtlineare Trennmodelle:
- Dann Modell allgemeiner: Eine Funktion f und Merkmale $X_1, X_2, \ldots, X_m$ mit

$$Y = f(X_1, X_2, \ldots, X_m)$$

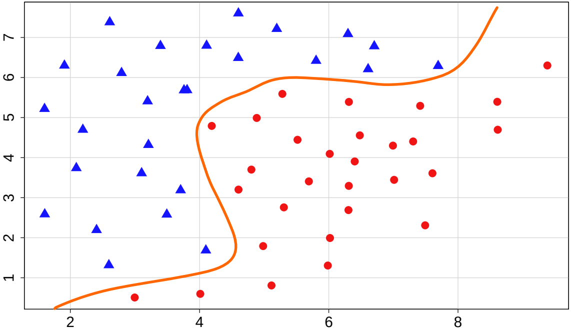
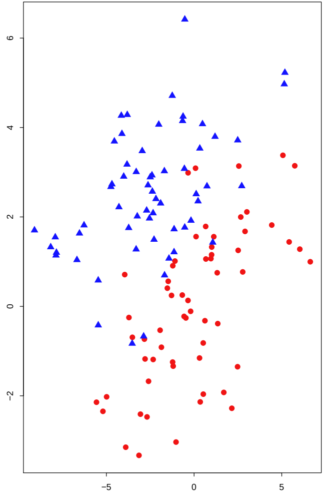


Bild: Lineare Trennung beschreibt Struktur nicht ideal

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Qualität des Modells

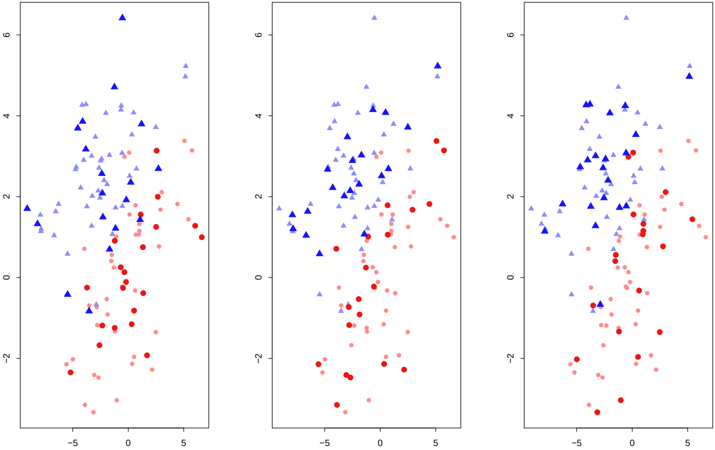
- Maßzahlen für die Güte eines Modells: Anteil der korrekt/falsch klassifizierten Objekte
- Aber: Durch viele Modellvarianten und Parametrisierungen Gefahr der Anpassung des Modells an die Stichprobe und nicht an die Struktur in den Daten (Overfitting)
- Sinnvoll deswegen: Dreiteilung der Daten nach Berry und Linoff (2004) in Trainingsdaten, Testdaten und Evaluierungsdaten ▸ Mehr



Dreiteilung der Daten

Zurück

- Beispieldaten in 2 Klassen,
- pro Klasse jeweils 60 Objekte und zwei metrische Merkmale

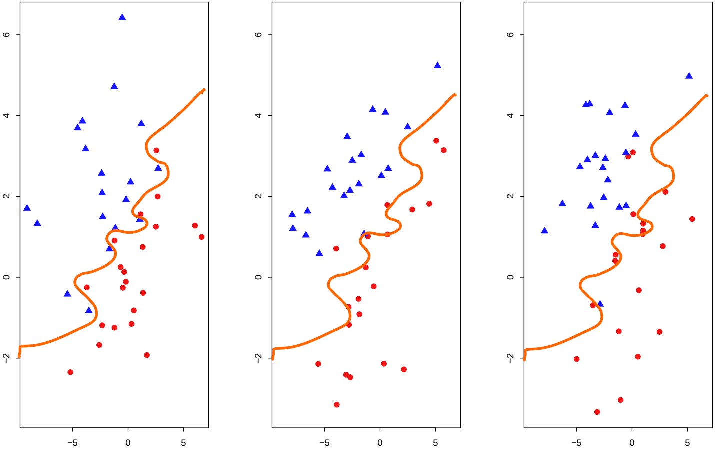


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Dreiteilung der Daten

Zurück

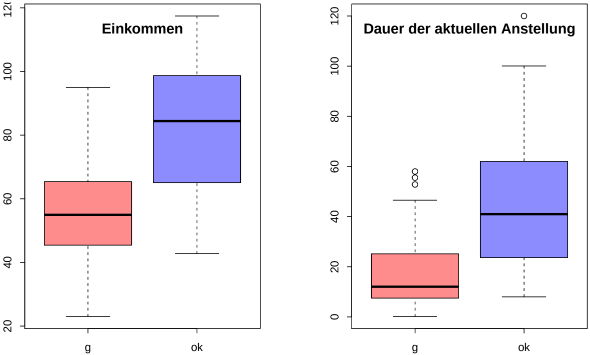
- Beispieldaten in 2 Klassen,
- pro Klasse jeweils 60 Objekte und zwei metrische Merkmale



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Beispiel: Kreditscoring

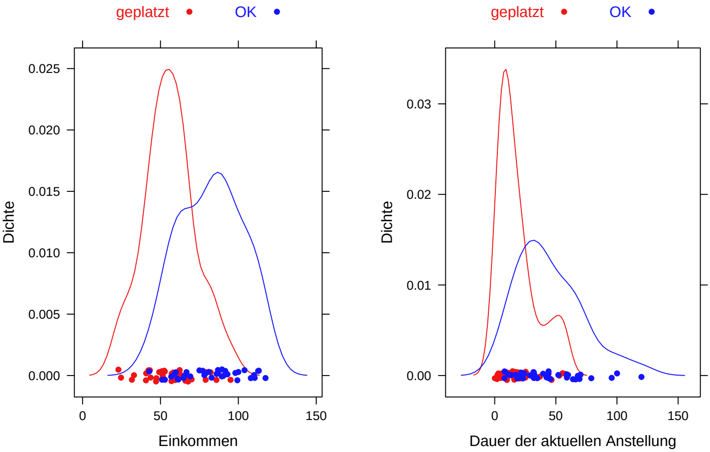
- Daten von 72 vergebenen Krediten, 2 metrische Merkmale, 1 nominales Merkmal
- Geschichtete Lernstichprobe mit jeweils 36 guten ("ok") und 36 schlechten ("g"eplatztten) Krediten
- Ziel: Abschätzen der Bonität von potentiellen Kreditnehmern
- Metrische Merkmale: Einkommen des Kreditnehmers, Dauer des Beschäftigungsverhältnisses bei Kreditantrag (in Monaten)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

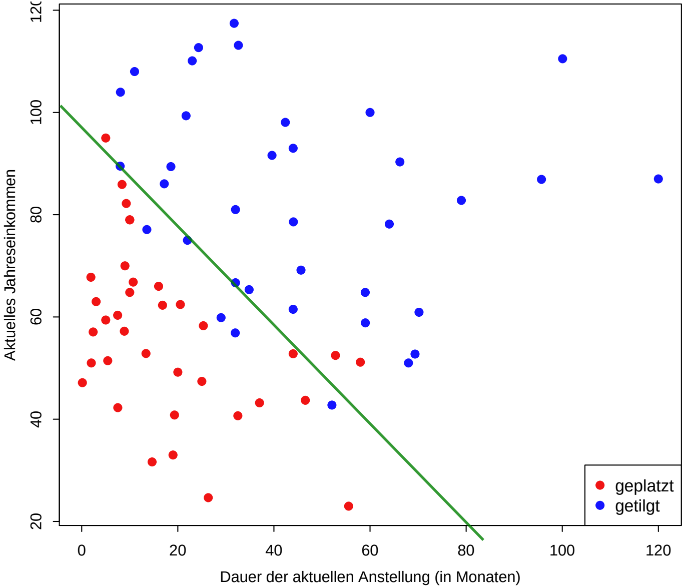
Inspektion der Daten

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines



- Dichte-Plots nach Merkmal und Bonität
- Anscheinend: Trennung bezüglich eines Merkmals nicht sinnvoll

Inspektion der Daten: Streuplot



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

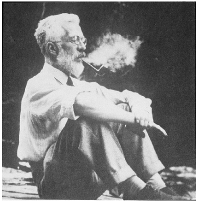
Zuordnung von Skalenniveaus zu Verfahren

unabhängige Variablen X_i / abhängige Variable Y	quantitativ	nominal
quantitativ	multiple Regression	Varianzanalyse
quantitativ	Kovarianzanalyse	
ordinal		Conjointanalyse
nominal	Diskriminanz-analyse	
binär	Logistische Regression	

- Y heißt auch **endogene Variable**
- $X_1, \ldots, X_m$: **exogene Variablen**
- Zu beachten: **Kausale Abhängigkeit** muss vorliegen

Lineare Diskriminanzanalyse

- Lineare Diskriminanzanalyse: Klassisches lineares Verfahren nach Fisher (1936)
- Keine Verteilungsannahmen nötig, gleich mächtige Klassengrößen aber sinnvoll
- Idee: Lineare Reduktion aller Merkmale auf eine (oder mehrere) Response-Variablen

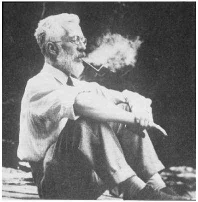


Ronald Aylmer Fisher
(1890 – 1962)

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Lineare Diskriminanzanalyse

- Lineare Diskriminanzanalyse: Klassisches lineares Verfahren nach Fisher (1936)
- Keine Verteilungsannahmen nötig, gleich mächtige Klassengrößen aber sinnvoll
- Idee: Lineare Reduktion aller Merkmale auf eine (oder mehrere) Response-Variablen
- Bestimmung linearer Gewichte und einer Diskriminanzfunktion



Ronald Aylmer Fisher
(1890 – 1962)

$$Y = g_1 \cdot X_1 + g_2 \cdot X_2 + \ldots + g_m \cdot X_m$$

- Wähle g_i so, dass Streuung zwischen Klassenmitten bzgl. Y groß und innerhalb der Klassen klein wird (Optimale Werte: Eigenwertproblem)
- Geometrisch: Projektion der Daten auf Ursprungsgerade in Richtung von α

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Eigenschaften

- Gegeben: m **metrische Merkmale** $X_1, \ldots, X_m$ sowie ein **nominales Merkmal** Y.
- Y entsteht häufig aufgrund einer Zerlegung der Objektmenge in disjunkte Klassen $K_1, \ldots, K_s$.
- Gesucht: Lineare Funktion, d.h. **eine Gewichtung** der Merkmale $X_1, \ldots, X_m$, so dass die Variable Y bestmöglich approximiert wird $(1 < s < n)$:

$$Y \cong g_1 X_1 + \ldots + g_m X_m$$

$$\text{bzw.} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \cong \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix} = X \cdot g$$

Diskriminanzkriterium

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Wann ist Gewichtung gut gewählt?

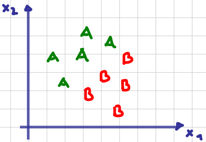
- Gewichtung der unabhängigen Variablen so,
- dass die Unterschiede zwischen den prognostizierten Werten **einer Klasse möglichst klein**,
- die **verschiedener Klassen aber möglichst groß** sind, d.h.

$$\begin{array}{lclcl} i,j \in K_k & \rightarrow & (\hat{y}_i - \hat{y}_j)^2 & \longrightarrow & \min \\ i \in K_k \;\text{ und }\; j \in K_l \;\text{ mit }\; k \neq l & \rightarrow & (\hat{y}_i - \hat{y}_j)^2 & \longrightarrow & \max \end{array}$$

- Deswegen: Varianzen **innerhalb** der Klassen sollen **möglichst klein** und die Varianz **zwischen** den Klassen **möglichst groß**.
- Kriterium sichert, dass Gewichtung gefunden wird, bei der die Mittelwerte der einzelnen Klassen - und damit auch die Klassen selbst - sehr verschieden sind und gleichzeitig auch die Homogenität innerhalb der Klassen sehr groß ist. Überschneidungen bleiben trotzdem möglich.

Lineare Diskriminanzanalyse

Einfaches Beispiel: 2 Klassen, 2 unabh. Merkmale



Datenmatrix

$$\left(\begin{array}{cc|c} x_{11} & x_{12} & a_{13} \\ x_{21} & x_{22} & a_{23} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & a_{n3} \end{array}\right)$$

unabh. Merkmale $\in \mathbb{R}$ | abh. Merkm. $\in \{A, B\}$

- Gesucht: Projektion von x^1 und x^2 auf Gerade, so dass Klassen A, B optimal getrennt werden

$$y_i = x_{i1} g_1 + x_{i2} g_2 \qquad i = 1, \dots, n$$

- Kriterien: ① Innerklassenvarianzen sollen möglichst klein sein

Varianz innerhalb Klasse soll klein sein

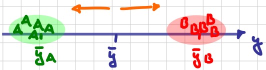
Klasse A: $w_A = \underbrace{\frac{1}{|K_A|}}_{\text{Anzahl Obj. in Klasse A}} \sum_{i \in A} (y_i - \bar{y}_{K_A})^2$

analog Klasse B: w_B

gesamte Innerklassenvarianz w:

$$w = \frac{1}{n} \left(|K_A| \cdot w_A + |K_B| \cdot w_B \right)$$

② Zwischenklassenvarianz soll möglichst groß sein



$$b = \frac{1}{n} \left(|K_A| (\bar{y}_A - \bar{y})^2 + |K_B| (\bar{y}_B - \bar{y})^2 \right)$$

Betrachte Kovarianzmatrizen der unabh. Merkmale

① Innerklassen Kovarianzmatrix

$$V^A_{kl} = \frac{1}{|K_A|} \sum_{i \in K_A} (x_{ik} - \bar{x}_{K_A k})(x_{il} - \bar{x}_{K_A l})$$

k, l sind Indizes der x_i, also z.B.
1,1 , 1,2
2,1 , 2,2

Mittelwert im Merkmal k in der Klasse A

analog V^B_{kl}

Gesamt: $\boxed{V_{kl} = \frac{1}{n} \left(|K_A| \cdot V^A_{kl} + |K_B| \cdot V^B_{kl} \right)}$

② Zwischenklassenkovarianzmatrix

$$z_{kl} = \frac{1}{n} \left[|K_A| \cdot (\bar{x}_{Ak} - \bar{x}_k)(\bar{x}_{Al} - \bar{x}_l) + |K_B| \cdot (\bar{x}_{Bk} - \bar{x}_k)(\bar{x}_{Bl} - \bar{x}_l) \right]$$

mit $y = X \cdot g$ folgt für Innerklassenvarianz

$$\Leftrightarrow \quad w = g^T V g$$

analog Zwischenklassenvarianz:

$$b = g^T Z g$$

Fishers Diskriminanzkriterium:

$$\lambda(g) = \frac{b}{w} = \frac{g^T Z g}{g^T V g} \longrightarrow \max$$

.... $\Rightarrow$ $\boxed{\underbrace{V^{-1}Z}_{\text{Matrix}} \cdot \underbrace{g}_{\text{Vektor}} = \overset{\text{Zahl}}{\lambda} \cdot g}$

$\Rightarrow$ - Eigenwerte von $V^{-1}Z$ maximieren $\lambda(g)$

- Projektionsgewichte sind Eigenvektoren

Beispiel:

$$\left.\begin{matrix} 4 & 1 \\ 2 & 4 \\ 2 & 3 \\ 3 & 6 \\ 4 & 4 \end{matrix}\right\} A \qquad \left.\begin{matrix} 9 & 10 \\ 6 & 8 \\ 9 & 5 \\ 8 & 7 \\ 10 & 8 \end{matrix}\right\} B$$

Innerklassenkov. matrix V:

$$\mathrm{Cov}(X^A) = \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{pmatrix} = V^A$$

$$\mathrm{Cov}(X^B) = \begin{pmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{pmatrix} = V^B$$

$$\Rightarrow V = \frac{1}{10}\left(5 \cdot V^A + 5 \cdot V^B\right) = \begin{pmatrix} 1.32 & -0.22 \\ -0.22 & 2.64 \end{pmatrix}$$

Zwischenklassenkov. matrix Z:

$\bar{x}^A = (3{,}0 \quad 3{,}6)$; $\bar{x}^B = (8.4 \quad 7.6)$; $\bar{x} = (5.7 \quad 5.6)$

Für TR:

$$\left.\begin{matrix} \left.\begin{matrix} 3.0 & 3.6 \\ 3.0 & 3.6 \\ \vdots & \vdots \\ 3.0 & 3.6 \end{matrix}\right\} \text{5 mal} \\ \left.\begin{matrix} 8.4 & 7.6 \\ \vdots & \vdots \\ 8.4 & 7.6 \end{matrix}\right\} \text{5 mal} \end{matrix}\right\} \begin{matrix} \mathrm{Cov} = r \cdot s_x \cdot s_y \\ \mathrm{Var} = s_x^2 \end{matrix}$$

$$\Rightarrow Z = \begin{pmatrix} 7.29 & 5.4 \\ 5.4 & 4.0 \end{pmatrix}$$

Exkurs: Invertierung einer 2×2-Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

damit $V^{-1} \approx \begin{pmatrix} 0.768 & 0.064 \\ 0.064 & 0.384 \end{pmatrix}$

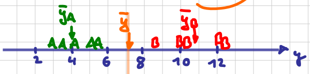
$$\Rightarrow M = V^{-1}Z = \begin{pmatrix} 5.946 & 4.405 \\ 2.541 & 1.882 \end{pmatrix}$$

$\Rightarrow$ Eigenwerte: $\lambda_1 = 7.82$, $\lambda_2 = 0$

Eigenvektor zu λ_1: $\begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$

Berechne $\hat{Y} = X \cdot g$

$$\hat{Y} = X \cdot \begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix} = \left.\begin{matrix} 4.07 \\ 3.41 \\ 3.02 \\ 5.12 \\ 5.25 \end{matrix}\right\} A \qquad \left.\begin{matrix} 12.2 \\ 8.66 \\ 10.24 \\ 10.11 \\ 12.34 \end{matrix}\right\} B$$



$\bar{y} = \bar{x} \cdot \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 7.44$

$\bar{y}_A = \bar{x}_A \cdot \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 4.17$

$\bar{y}_B = \bar{x}_B \cdot \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 10.71$

neues Objekt soll zugeordnet werden

z.B. $x = (7 \quad 3)$

$\Rightarrow y = (7 \quad 3) \begin{pmatrix} 0.9196 \\ 0.3930 \end{pmatrix}$

$= 7{,}6162 > 7.44 \rightarrow B$

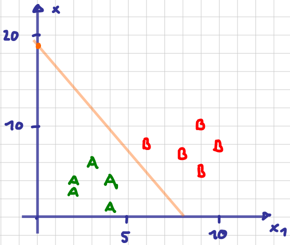
Trennebene (hier: -gerade)

dazu: suche (x_1, x_2), so dass $(x_1, x_2)g = \bar{y}$

$\Leftrightarrow \quad 0.9196 \cdot x_1 + 0.393\, x_2 = 7.44$

$\Leftrightarrow \quad x_2 = \frac{7.44}{0.393} - \frac{0.9196}{0.393} x_1$

$\Leftrightarrow \quad x_2 = 18.9 - 2.3\, x_1$



$$\left.\begin{matrix} 4 & 1 \\ 2 & 4 \\ 2 & 3 \\ 3 & 6 \\ 4 & 4 \end{matrix}\right\} A$$

$$\left.\begin{matrix} 9 & 10 \\ 6 & 8 \\ 9 & 5 \\ 8 & 7 \\ 10 & 8 \end{matrix}\right\} B$$

Definition Innergruppen- und Zwischengruppenvarianz

- **Innergruppenvarianz** der Klasse K

$$w_K = \frac{1}{|K|} \sum_{i \in K} (\hat{y}_i - \bar{y}_K)^2 \quad \text{mit} \quad \bar{y}_K = \frac{1}{|K|} \sum_{i \in K} \hat{y}_i$$

- **Gesamt-Innergruppenvarianz** (gewichtete Summe der w_k)

$$w = \frac{1}{n} \sum_K |K| \cdot w_K$$

- **Zwischengruppenvarianz**

$$b = \frac{1}{n} \sum_K |K| (\bar{y}_K - \bar{y})^2 \quad \text{mit} \quad \bar{y} = \frac{1}{n} \sum_i \hat{y}_i$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Diskriminanzanalyse: Kovarianzmatrizen

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Definition entsprechender **Kovarianzmatrizen** für unabhängige Variablen $X_1, \ldots, X_m$ gemäß $X = (x_{ik})_{n \times m}$)

- **Innergruppenkovarianzmatrix** $V^K = (v^K_{kl})_{m \times m}$

$$v^K_{kl} = \frac{1}{|K|} \sum_{i \in K} (x_{ik} - \bar{x}_{Kk})(x_{il} - \bar{x}_{Kl}) \quad \text{mit} \quad \bar{x}_{Kk} = \frac{1}{|K|} \sum_{i \in K} x_{ik}$$

- **Gesamt-Innergruppenkovarianzmatrix**

$$V = \frac{1}{n} \sum_K |K| \cdot V^K$$

- **Zwischengruppenkovarianzmatrix** $Z = (z_{kl})_{m \times m}$

$$z_{kl} = \frac{1}{n} \sum_K |K| (\bar{x}_{Kk} - \bar{x}_{.k})(\bar{x}_{Kl} - \bar{x}_{.l}) \quad \text{mit} \quad \bar{x}_{.k} = \frac{1}{n} \sum_i x_{ik}$$

Definitionen

- Gegeben: Datenmatrix $X = (x_{ik})_{n \times m}$ der m unabhängigen Variablen,
- und es gelte $y = X \cdot g$.
- Ferner gegeben: Kovarianzmatrizen $V_K (k = 1, \ldots, s)$, V, Z und $S = \mathrm{Cov}(X)$ gegeben.

$$\begin{array}{lclclcl} b & = & g^T Z g & \text{und} & c & = & g^T S g \\ w_K & = & g^T V_K g & \text{und} & w & = & g^T V g \end{array}$$

- Zu beachten dabei: Gesamtvarianz c von Y ist gemäß $c = w + b$ zu berechnen
- Daraus: Diskriminanzkriterium (D):

$$\lambda = \frac{b}{w} = \frac{g^T Z g}{g^T V g} \quad \longrightarrow \quad \max_g$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Satz

- V sei regulär und positiv definit.
- Dann impliziert die Lösung von (D) ein **Eigenwertproblem** der Matrix $V^{-1}Z$ gemäß

$$(V^{-1}Z + \lambda \cdot E)g = 0.$$

- Es existieren dann q Eigenwerte der Form

$$\lambda = \frac{b}{w} = \frac{g^T Z g}{g^T V g} > 0$$

- wobei $q = \operatorname{Rg} Z \leq \min\{m, s-1\}$ ist.
- Die zugehörigen Eigenvektoren $g^1, \ldots, g^q$ sind linear unabhängig.

Folgerungen aus dem Hauptsatz (1)

- Mit den Eigenwerte $\lambda_1, \ldots, \lambda_q$
- und den zugehörigen Eigenvektoren $g^1, \ldots, g^q$ folgt:

$$\hat{y}^k = Xg^k$$

- Gewichtungsvektor g^1 trennt die Klassen bestmöglich,
- Gewichtungsvektor g^2 trennt die Klassen am zweitbesten, usw.
- wobei die $\hat{y}^k$ paarweise unkorreliert sind.
- Terminologie ($(k = 1, \ldots, q)$):

$\hat{y}_{ik} = x_i^T g^k$	k-te Diskriminanzfunktion
$\hat{y}_{ik}$	Diskriminanzvariable
g^k	k-ter Diskriminanzkoeffizientenvektor

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Folgerungen aus dem Hauptsatz (2)

Man erhält folgende Mittelwerte:

- Klassenmittelwerte bzgl. der k-ten DA-Funktion

$$\bar{y}_{Kk} = \bar{x}_K^T g^k \quad \text{mit} \quad \bar{x}_K = (\bar{x}_{Kl})_{m \times 1}$$

- Globalmittelwerte bzgl. der k-ten DA-Funktion

$$\bar{y}_{\cdot k} = \bar{x}^T g^k \quad \text{mit} \quad \bar{x} = (\bar{x}_{\cdot l})_{m \times 1}$$

- Und: Trennebenen bzgl. der k-ten DA-Funktion:

$$\bar{y}_{\cdot k} = \bar{x}^T g^k = x^T g^k \qquad \text{(mit } x \text{ als Variablenvektor)}$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Folgerungen aus dem Hauptsatz (3)

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Identifikation neuer Objekte

- Durch Trennebenen: Neue, in der Analyse von X und Y noch nicht enthaltene Objekte x_j können durch Vergleich mit Globalmittelwerten den verschiedenen Klassen zugeordnet werden.
- Schema:

$$\text{Falls} \quad x_j^T g^k \begin{cases} \geq & \bar{y}_{\cdot k} \quad \text{Zuordnung zu } K \text{ mit } \bar{y}_{Kk} \geq \bar{y}_{\cdot k} \\ < & \bar{y}_{\cdot k} \quad \text{Zuordnung zu } K \text{ mit } \bar{y}_{Kk} < \bar{y}_{\cdot k} \end{cases}$$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Trefferquote

- Oft: Mit Identifikationsregeln für neue Objekte wird eine Identifikation der Objekte erstellt, die zur Generierung des Modells verwendet wurden.
- Durch einen Vergleich der prognostizierten Klassenzugehörigkeit und der wahren Klassenzugehörigkeit kann dann der Anteil der richtig klassifizierten Objekte bestimmt werden.
- Also: **Trefferquote** TQ gemäß

$$\text{TQ} = \frac{1}{n} \cdot \text{Anzahl korrekt klassifizierter Objekte}$$

- Deskriptives Maß TQ kann als Maß für die **Güte der Diskriminanzanalyse** verwendet werden.

Folgerungen aus dem Hauptsatz (5)

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Anzahl der Trennebenen

- Jede DA-Funktion trennt genau in zwei Gruppen.
- Konsequenz für den Mehrgruppen-Fall: Zur korrekten Trennung der s Klassen sind $s-1$ DA-Funktionen notwendig.
- Maß für die relative Wichtigkeit der einzelnen Diskriminanz-Funktionen: Eigenwertanteil gemäß

$$EA_k = \frac{\lambda_k}{\lambda_1 + \ldots + \lambda_q}$$

- EA_k: Anteil der durch die k-te DA-Funktion erklärten Streuung an der durch alle DA-Funktionen erklärten Gesamtstreuung

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Diskriminanzanalyse: Ablauf

Modellerstellung

1. Gegeben: Matrix X, Klassifikation $K_1, \ldots, K_s$
2. Berechne $V, V^{-1}, Z, \mathrm{Rg} Z, V^{-1} \cdot Z$
3. Löse Eigenwertproblem $(V^{-1}Z - \lambda \cdot E)g = 0$
4. Bestimme $\lambda_1 \geq \ldots \geq \lambda_q > 0$
5. Bestimme $G = (g^1, \ldots, g^q)$
6. Berechne $\hat{Y} = X \cdot G$ $(n \times q)$-Matrix der DA-Variablen
7. sowie Klassenmittelwerte $\bar{y}_K^T = \bar{x}_K^T \cdot G$
8. Globalmittelwerte $\bar{y}^T = \bar{x}^T \cdot G$
9. Trennebenen $\bar{y}^T = x^T \cdot G$ mit variablem x

Identifikation

Falls $x_j^T g^k \begin{cases} \geq & \bar{y}_{\cdot k} \quad \text{Zuordnung zu } K \text{ mit } \bar{y}_{Kk} \geq \bar{y}_{\cdot k} \\ < & \bar{y}_{\cdot k} \quad \text{Zuordnung zu } K \text{ mit } \bar{y}_{Kk} < \bar{y}_{\cdot k} \end{cases}$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Anwendung der LDA mit R

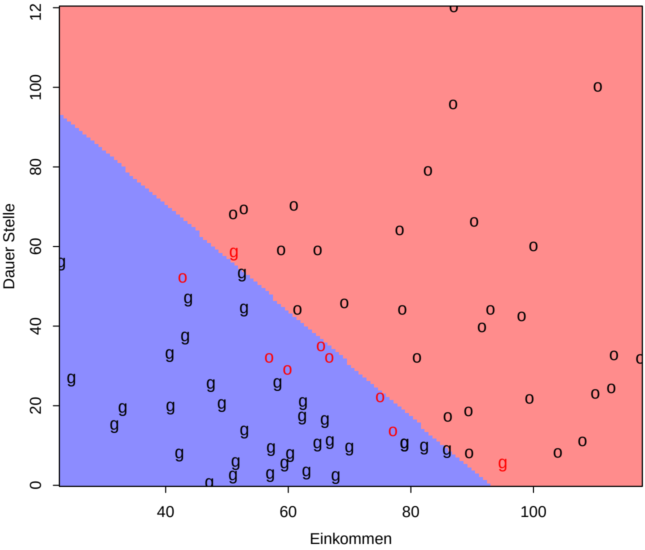
```r
set.seed(1504)
s = sample(1:72, 10)
bank.lda = lda(geplatzt ~ .,
               bankLong[-s,])
bank.lda

## Call:
## lda(geplatzt ~ ., data = bankLong[-s, ])
##
## Prior probabilities of groups:
##      g     ok
## 0.5323 0.4677
##
## Group means:
##    Einkommen DauerStelle
## g      56.91       17.23
## ok     85.86       42.06
##
## Coefficients of linear discriminants:
##                 LD1
## Einkommen   0.04879
## DauerStelle 0.03923
```

```r
test = bankLong[s,]
test$Prognose =
  predict(bank.lda,
          test[,c("Einkommen",
                  "DauerStelle")])$class
test

##    Einkommen DauerStelle geplatzt Prognose
## 11     51.00       68.00       ok       ok
## 35     52.74       69.36       ok       ok
## 3      64.80       59.00       ok       ok
## 12     81.00       32.00       ok       ok
## 62     51.15       57.99        g       ok
## 22     33.00       19.00        g        g
## 5      87.00      120.00       ok       ok
## 25     66.69       32.00       ok        g
## 68     52.85       13.37        g        g
## 26     77.10       13.56       ok        g
```

Klassifikation der Beispieldaten mit LDA

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Beispieldaten Bank



Klassifikationsbäume

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Grundprinzip Klassifikationsbäume

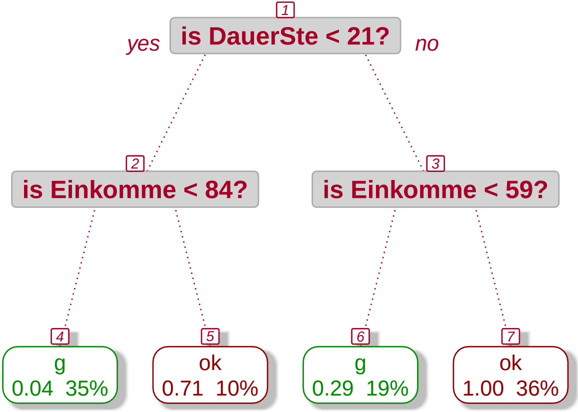
- Merkmalsraum soll pro Merkmal so aufgeteilt werden, dass Untermengen in sich möglichst homogen, untereinander möglichst heterogen sind
- Grundlegende Arbeit: Von Breiman et al. (1984)

Schrittweiser Aufbau eines Entscheidungsbaumes

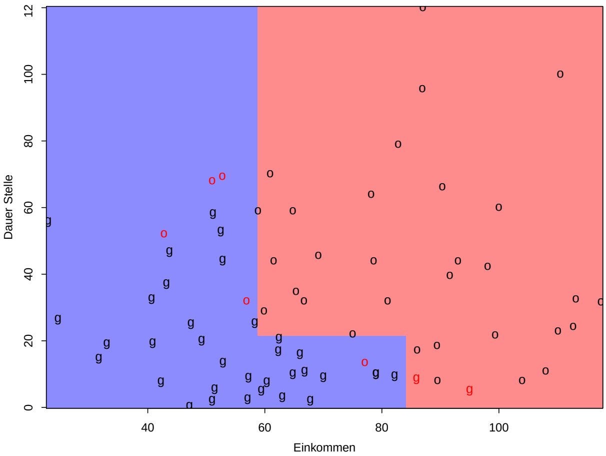
1. Suche Merkmal und Trennpunkt mit bester Unterscheidung zwischen Klassen
2. Teile Daten gemäß dieser Regel
3. Für jeden der beiden Bereiche wende **1.** analog an, bis Stopkriterium erfüllt

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Entscheidungsbaum der Beispieldaten



Entscheidungsbäume: Klassifikation der Beispieldaten



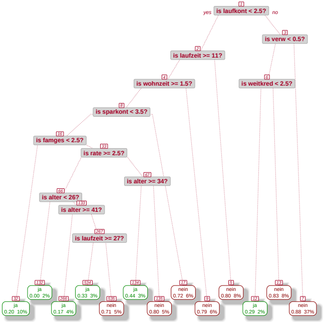
- Entscheidungsbäume sind nichtlineare Klassifikationsverfahren: Damit Gefahr des Overfitting

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

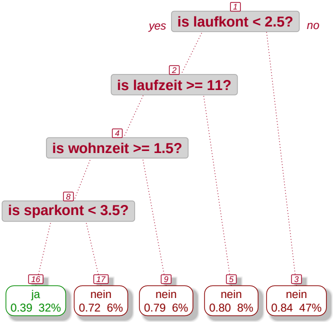
Zurückschneiden eines Baumes: Pruning

- Komplexität der Bäume a priori festlegen ist i.a. nicht optimal
- Idee: Bäume erst wachsen lassen, anschließend zurückstutzen (Pruning)

Entscheidungsbaum, unbeschnitten



Entscheidungsbaum, beschnitten



Zurückschneiden eines Baumes: Pruning

Zurück

- Komplexität der Bäume a priori festlegen ist nicht optimal
- Idee: Bäume erst wachsen lassen, anschließen zurückstutzen (Pruning)
- Bewertung des zurückgeschnittenen Baumes z.B. mittels Kosten-Komplexitäts-Maß

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|$$

- Dabei bedeutet:
  - $R(T)$: Fehlklassifikationsrate durch den Baum T
  - $|\tilde{T}|$: Anzahl der Knoten im Baum T
  - $\alpha > 0$: Faktor des Komplexitäts-Bestrafungs-Terms

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Vorgehen beim Beschneiden von Bäumen

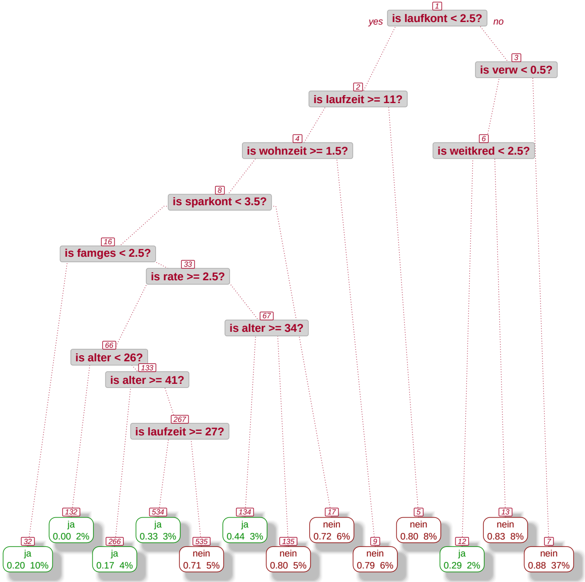
1. Beschneide Baum um die Äste, die die Fehlklassifikationsrate nicht verbessern
2. Bewerte bei allen Nicht-Endknoten A_t und den davon ausgehenden Ästen $T(A_t)$ mittels Kosten-Komplexitäts-Maß
   - den Knoten, wenn er ein Endknoten wäre $R_\alpha(A_t)$ und
   - den kompletten Ast ab diesem Knoten $R_\alpha\left(T(A_t)\right)$
3. Stutze Baum beim Knoten mit dem kleinsten Unterschied von

$$R_\alpha(A_t) - R_\alpha\left(T(A_t)\right)$$

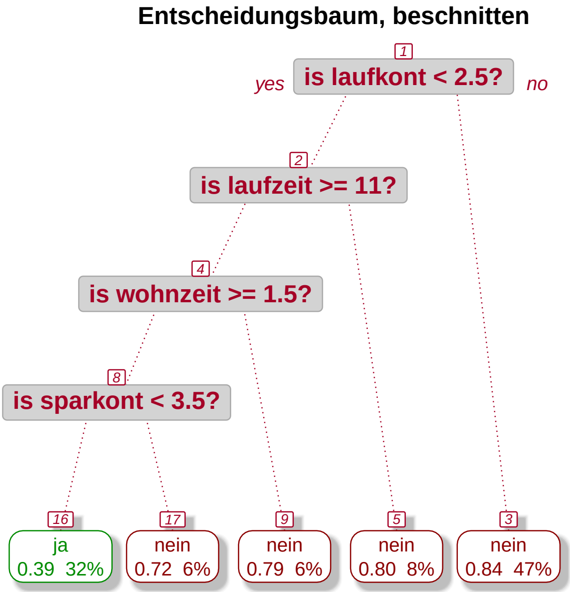
4. Gehe zu Schritt **2.** solange $R_\alpha(A_t) - R_\alpha\left(T(A_t)\right) > 0$

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
   - Problemstellung
   - Beispielproblem
   - Lineare Diskriminanzanalyse
   - Klassifikationsbäume
   - Support-Vector-Machines

Entscheidungsbaum, unbeschnitten



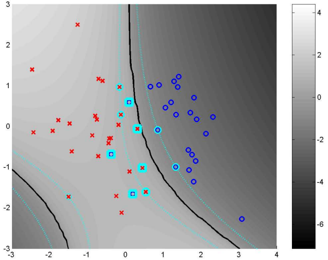
Beispiel Pruning mit $\alpha = 0.045$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

- Klassifikationsverfahren seit Mitte der 1990er Jahre (siehe Vapnik, 1999)
- Idee: Unterteilung der Daten in 2 Klassen durch Abbildung der Daten in höhere Dimension
- In höherer Dimension: Benutzung einer einfachen Funktion zur Abstandsbestimmung und Bestimmung einer "Trennungslinie"



- Dabei möglichst: Abstand der Klassen zur Trennungslinie maximal
- Äußerste Punkte der Klassen: Stützvektoren (Support Vector)
- Falsch klassifizerte Objekte erhöhen eine Kostenfunktion ▸ Mehr

Support-Vector-Machines

Zurück

- Kernel Methoden: Nutzen Informationen über das innere Produkt zwischen Objekten
- Viele Algorithmen lassen sich umformulieren, so dass sie nur noch das innere Produkt zwischen Daten benötigen
- Wenn der Kernel bekannt ist, muss man die Eigenschaften der Daten nicht spezifizieren
- Unabh. Daten im Raum: $x \in X$, Abhängige Daten in $y \in Y = \{-1, 1\}$; Zum Beispiel

$$f(x) = \langle w, x \rangle + b, \qquad h(x) = \operatorname{sign}(f(x))$$

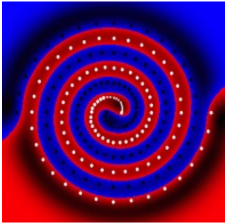
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

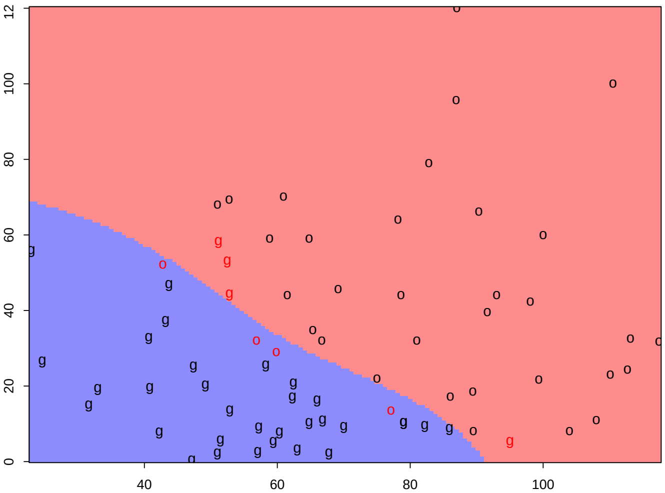
- d.h. Transformation in Feature-Raum muss nicht gemacht werden, es genügt, wenn die zugehörige Kernel-Funktion bekannt ist.
- Das heißt für einen Kernel

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

  muss weder die Dimension, noch die Transformation ϕ bekannt sein. Es genügt der Kernel
- Kernel-Trick: Bilde Daten in Repräsentationsraum ab, in dem sie (einigermaßen vernünftig) linear separierbar sind,
- Messe Kosten aber nur via Kernelfunktion...

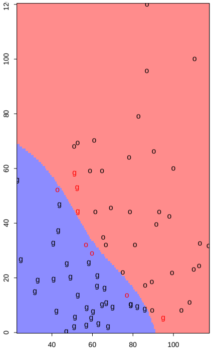


Support-Vector-Machines: Ergebnisse mit Beispieldaten



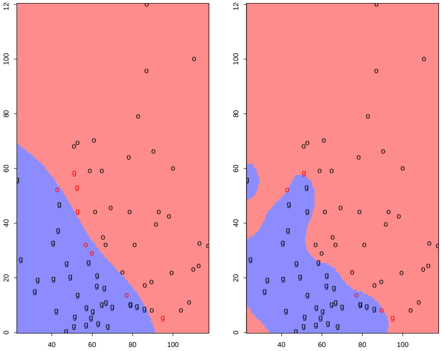
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

Support-Vector-Maschines und Overfitting



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

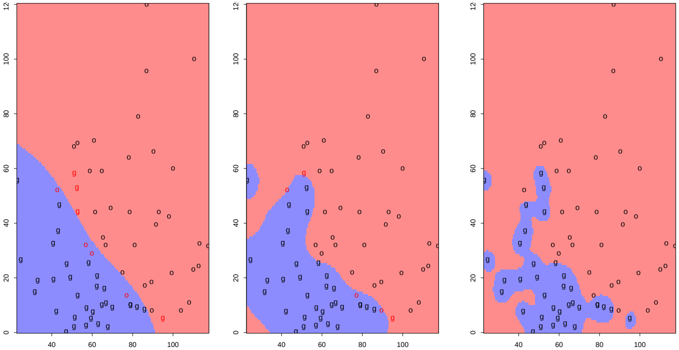
Support-Vector-Maschines und Overfitting



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines

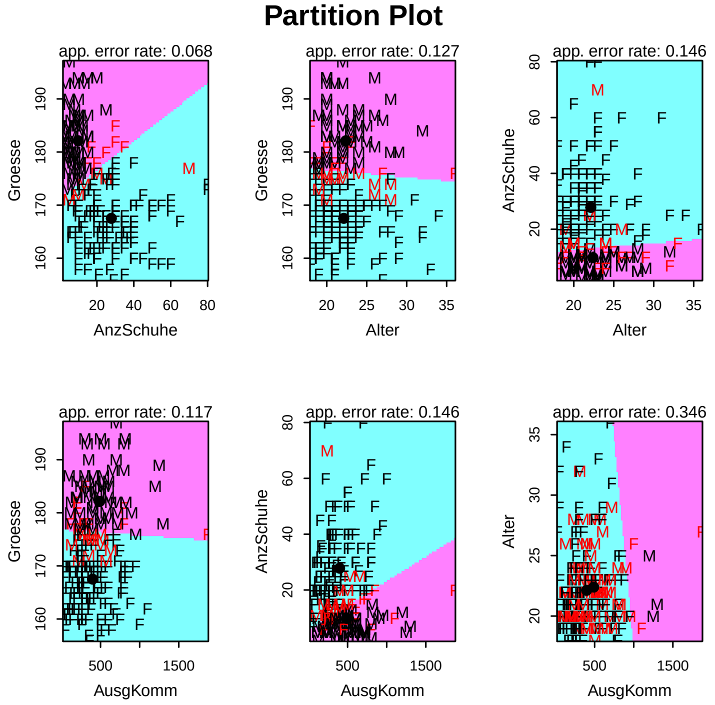
Support-Vector-Maschines und Overfitting

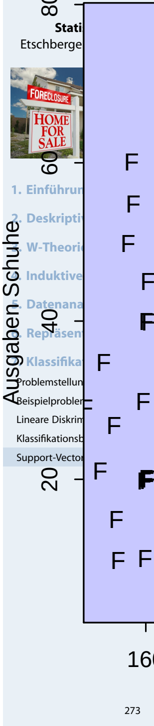


- Gefahr des Overfitting durch sukzessives Anpassen der Methodenparameter sehr groß
- Teil der Daten als Test- und ggf. Validierungsdaten unbedingt erforderlich

```
partimat(Geschlecht ~ Groesse + AnzSchuhe + Alter + AusgKomm, data=MyData, method="lda")

set.seed(1504)
M = as.data.frame(sapply(MyData[,c(1,2,11)], jitter))
M$Geschlecht = factor(MyData$Geschlecht, levels=c("Frau", "Mann"))

drawparti(M$Geschlecht, M$Groesse,
          M$AnzSchuhe, method="lda",
          prec=drawpartiprecision,
          ylab="Ausgaben Schuhe", xlab="Groesse",
          cex.mean=0, cex=2,
          imageplot=TRUE,
          image.colors=c(blue20, red20) )

M.lda = lda(Geschlecht ~ Groesse + AnzSchuhe, data=M)
M.predict = predict(M.lda, data=M)
errormatrix(M$Geschlecht, M.predict$class)

##         predicted
## true     Frau Mann -SUM-
##   Frau    127    7     7
##   Mann      7   64     7
##   -SUM-     7    7    14

# percentages
round(100 * errormatrix(M$Geschlecht, M.predict$class, relative=TRUE), 1)

##         predicted
## true     Frau Mann -SUM-
##   Frau   94.8  5.2   5.2
##   Mann    9.9 90.1   9.9
##   -SUM- 50.0 50.0   6.8
```


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
5. Datenanalyse
6. Repräsentation
7. Klassifikation
Problemstellung
Beispielproblem
Lineare Diskriminanzanalyse
Klassifikationsbäume
Support-Vector-Machines


Partition Plot
app. error rate: 0.068
app. error rate: 0.127
app. error rate: 0.146
app. error rate: 0.117
app. error rate: 0.146
app. error rate: 0.346
Groesse
AnzSchuhe
Alter
AusgKomm


FORECLOSURE
HOME FOR SALE
Ausgaben Schuhe
273