

Diskriminanzanalyse

Aufgabe 1:

Die Einteilung von Geschlechtern in drei Leistungsklassen und die a priori-Wahrscheinlichkeiten für die jeweilige Klassenzugehörigkeit seien durch die folgende Tabelle bestimmt.

	X=Frau	X=Mann	apriori-Wahrscheinlichkeit
Klasse1	0.8	0.2	0.1
Klasse2	0.4	0.6	0.8
Klasse3	0.1	0.9	0.1

- Bestimmen Sie die ML-Zuordnung für X=Frau und X=Mann. Welche Fehlerrate ϵ_{12} ergibt sich für die ML-Zuordnungsregel?
- Bestimmen Sie die Bayes-Zuordnung für X=Frau und X=Mann.
- Welcher Klasse würden Sie eine Frau unter Benutzung der Bayeszuordnungsregel mit Kosten zuordnen, wenn folgende Kostentabelle vorliegt?

c_{ij}	j=1	j=2	j=3
i=1	0	1	2
i=2	5	0	5
i=3	2	1	0

Aufgabe 2:

Der Datensatz `leukaemie.txt` enthält Daten zu $n = 72$ Versuchspersonen. Man hat die dichotome Variable y (1: gesunde Versuchsperson; 2: Leukämiepatient) sowie Ausprägungen von 10 verschiedenen Genen der Personen zur Verfügung. Im Folgenden sollen die Patienten anhand der Genexpressionen mit verschiedenen Verfahren klassifiziert werden (bzgl. y).

- Verschaffen Sie sich einen Überblick über die Daten. Plotten Sie die Variable y (1: gesunde Versuchsperson; 2: Leukämiepatient) in Abhängigkeit der Genexpressionen `M92287.at` und `M11722.at` in unterschiedlichen Farben.
- Führen Sie eine lineare Diskriminanzanalyse mit den Kovariablen aus a) durch.
Hinweis: Verwenden Sie die Funktion `lda()` aus der R-library `MASS` zur Berechnung einer linearen Diskriminanzanalyse.
- Verwenden Sie den Wertebereich der betrachteten Einflussgrößen um neue Daten zu generieren. Wenden Sie die in b) geschätzte Zuordnungsregel auf die neu definierten Daten an (Funktion `predict()`). Wie gut funktioniert die Klassifizierung? Visualisieren Sie zusätzlich die dadurch geschätzte Zuordnungsregel in Abhängigkeit der Einflussgrößen graphisch.
- Führen Sie Teilaufgabe b) und c) mit einer quadratischen Diskriminanzanalyse (Funktion `qda()`) durch. Vergleichen Sie die beiden Ergebnisse.