

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2010

Blatt 1

Aufgabe 1.1 (3 Punkte)

In einer Werkstatt werden Schalter zusammengebaut. 40% aller Schalter montiert die Hilfskraft H, die restlichen ihr Kollege K. In der Regel arbeiten 90% der von H zusammengebauten Schalter einwandfrei. Die von K montierten Schalter funktionieren in 95% der Fälle.

- Mit welcher Wahrscheinlichkeit ist ein in diesem Werk produzierter Schalter in Ordnung?
- Ein der Produktion zufällig entnommener Schalter erweist sich als defekt. Mit welcher Wahrscheinlichkeit hat H ihn zusammengebaut?

Hinweis: Der Satz von Bayes könnte nützlich sein!

Aufgabe 1.2 (5 Punkte)

In der Vorlesung haben Sie bisher die eindimensionale Normalverteilung kennengelernt. Meist haben wir es aber mit p -dimensionalen Zufallsvariablen (mit $p \geq 2$) zu tun. Mit $\mathbf{x} = (x_1, \dots, x_p)' \in \mathbb{R}^p$ lautet die Dichtefunktion der mehrdimensionalen oder multivariaten Normalverteilung

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \det(\boldsymbol{\Sigma})^{1/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right).$$

Dabei ist $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)' \in \mathbb{R}^p$ der Erwartungswertvektor und $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ die Kovarianzmatrix.

- Ziehen Sie 500 Beobachtungen aus der zweidimensionalen Normalverteilung mit Erwartungswertvektor $\boldsymbol{\mu} = (0, 0)'$ und Kovarianzmatrix

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Hierbei ist die Funktion `rmvnorm` aus dem Paket `mvtnorm` hilfreich.

Erstellen Sie einen Scatterplot (`plot`) und zeichnen Sie die Konturlinien der Normalverteilungsdichte ein. Hierzu sind die Funktionen `dmvnorm`, `expand.grid` und `contour` nützlich.

- b) Betrachten Sie nur die zur ersten Variable X_1 gehörigen Beobachtungen. Erstellen Sie ein Histogramm (R-Funktion `hist` mit Option `freq = FALSE`) und zeichnen Sie die Dichte der univariaten Normalverteilung mit Erwartungswert 0 und Varianz 2 ein. Nützliche Funktionen sind `dnorm` und `curve`.

Berechnen Sie außerdem die Summe der Beobachtungen zu den beiden Variablen X_1 und X_2 . Plotten Sie das Histogramm und zeichnen Sie die Dichte der univariaten Normalverteilung mit Erwartungswert 0 und Varianz 4 ein.

- c) Ändern Sie die Kovarianzmatrix in

$$\Sigma_2 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 2 \end{pmatrix}$$

und erzeugen Sie dieselben Plots wie in Teil a). Was ändert sich?

Aufgabe 1.3 (4 Punkte)

Lesen Sie die Daten aus der Datei `aufgabe_1_3.txt` in R ein (`read.table`).

- a) Schauen Sie sich mit Hilfe der Funktion `summary` einige Kennzahlen zu den Daten an und zeichnen Sie die Scatterplot-Matrix (`pairs`). Berechnen Sie Kovarianz- und Korrelationsmatrizen und interpretieren Sie die Einträge.
- b) Standardisieren Sie den Datensatz und berechnen Sie für die 4 Variablen die 0.05, 0.1, 0.15, ..., 0.9, 0.95-Quantile (`quantile`). Vergleichen Sie die berechneten Quantile mit den theoretischen Quantilen der eindimensionalen Normalverteilung mit Erwartungswert 0 und Varianz 1 (`qnorm`). Kann man annehmen, dass die einzelnen Variablen normalverteilt sind?