Data Mining / Intelligente Datenanalyse WS 2017/2018

PD Dr.-Ing. habil. Christian Borgelt

2. Übungsblatt

Aufgabe 7 Hauptkomponentenanalyse

Gegeben sei der folgende zweidimensionale Datensatz:

x	0	1	1	2	3	3	4	5	5	6
y	0	1	2	3	2	3	4	4	6	5

Schon aus dieser Tabelle sieht man, daß die Merkmale x und y stark korreliert sind.

a) Berechnen Sie die Kovarianzmatrix der Daten!

b) Berechnen Sie den Korrelationskoeffizienten von x und y!

c) Führen Sie eine Hauptkomponentenanalyse durch (auf unnormierten Daten), d.h., bestimmen Sie die Richtungen der beiden Hauptachsen der Datenpunktwolke!

Aufgabe 8 Schätzfunktionen

Sei $X = (X_1, \ldots, X_n)$ der einer Zufallsstichprobe vom Umfang n zugrundeliegende Zufallsvektor. Die Zufallsvariablen X_i seien unabhängig und identisch verteilt mit der Exponentialverteilung $f_X(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x > 0$. Der Parameter θ dieser Verteilung sei zu schätzen. Der am häufigsten verwendete Schätzer für θ ist $W_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$. Wir betrachten hier jedoch den Schätzer $W_2 = n \cdot X_{\min} = n \min_{i=1}^{n} X_i$. Bestimmen Sie für diesen Schätzer die Wahrscheinlichkeitsdichte, also $f_{W_2}(w;\theta)$.
Hinweis: Erinnern Sie sich an den technischen Trick, den wir bereits in Aufgabe 1 benutzt haben, nämlich das Komplementereignis zu betrachten.

Aufgabe 9 Eigenschaften von Schätzfunktionen

Zeigen Sie: Die relative Häufigkeit r_A, mit der ein Ereignis A in einer gegebenen Stichprobe vom Umfang n auftritt, ist ein konsistenter und erwartungstreuer Schätzer für den Parameter $p = P(A)$ einer Binomialverteilung $b_X(x; p, n)$. (p ist die Wahrscheinlichkeit, mit der A bei einer einzelnen Durchführung des Zufallsexperimentes — eines Bernoulli-Experimentes — auftritt.)
Hinweis: Betrachten Sie das arithmetische Mittel von n unabhängigen Zufallsvariablen $Y_1, \ldots, Y_n$ über einem Bernoulli-Experiment mit

$$Y_i = \begin{cases} 1, & \text{falls das Ereignis } A \text{ im } i\text{-ten Versuch eintritt,} \\ 0, & \text{sonst.} \end{cases}$$

Aufgabe 10 Erwartungstreue von Schätzfunktionen

a) Es seien W_1 und W_2 zwei erwartungstreue Schätzer für den unbekannten Parameter θ. Wenn $W = aW_1 + bW_2$ ebenfalls ein erwartungstreuer Schätzer für θ sein soll, was muß dann für a und b gelten?

b) Zeigen Sie: Wenn $\mu = E(X)$ der zu schätzende Parameter ist, dann ist $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ immer ein erwartungstreuer Schätzer für μ.

Aufgabe 11 Effizienz von Schätzfunktionen

In der Vorlesung wurde als erwartungstreuer Schätzer für den Parameter θ einer Gleichverteilung auf dem Intervall $[0, \theta]$ der Schätzer $W_1 = \frac{n+1}{n} X_{\max} = \frac{n+1}{n} \max_{i=1}^{n} X_i$ verwendet. Alternativ kann man, mit einer ähnlichen Ableitung, auch den (ebenfalls erwartungstreuen) Schätzer $W_2 = (n+1)X_{\min} = (n+1)\min_{i=1}^{n} X_i$ benutzen. Welcher dieser beiden Schätzer ist effizienter?

Aufgabe 12 Maximum-Likelihood-Schätzung

Bestimmen Sie einen Maximum-Likelihood-Schätzer für den Parameter θ einer Gleichverteilung auf dem Intervall $[0, \theta]$! Zur Erinnerung: Die Zufallsvariablen des Stichprobenvektors haben die Wahrscheinlichkeitsdichte

$$f_X(x; \theta) = \begin{cases} 0, & \text{falls } x < 0, \\ \frac{1}{\theta}, & \text{falls } 0 \leq x \leq \theta, \\ 0, & \text{falls } x > \theta. \end{cases}$$

Ist der sich ergebende Schätzer konsistent und erwartungstreu?
(Hinweis: Beachten Sie bei der Maximierung der Likelihood-Funktion, daß bestimmte Nebenbedingungen einzuhalten sind. Vergleichen Sie das Ergebnis mit den in der Vorlesung behandelten Schätzern.)

Aufgabe 13 Maximum-Likelihood-Schätzung

Bestimmen Sie Maximum-Likelihood-Schätzer für die Parameter $\theta_1, \ldots, \theta_k$ einer Polynomialverteilung $f_{X_1,\ldots,X_k}(x_1, \ldots, x_k; \theta_1, \ldots, \theta_k, n)$! Zur Erinnerung: Die Polynomialverteilung ist definiert als

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k; \theta_1, \ldots, \theta_k, n) = \frac{n!}{\prod_{i=1}^{k} x_k!} \prod_{i=1}^{k} \theta_i^{x_i}.$$

(Hinweis: Ggf. ist es einfacher, wenn Sie zunächst versuchen, einen Maximum-Likelihood-Schätzer für den Parameter einer Binomialverteilung zu bestimmen, und dann das Ergebnis übertragen.)