Prof. Dr. E. Spodarev
D. Meschenmoser

Wintersemester 2009 / 2010
28. Januar 2010

Statistik II – Übungsblatt 13

Abgabe: 04. Februar 2010, vor den Übungen

Aufgabe 1

Auf der Homepage der Vorlesung ist der Datensatz `economy.txt` verfügbar, der die Werte für die Arbeitslosenquote, die Zunahme des Bruttoinlandsprodukts, die Inflationsrate, die Investitionsquote, die Steuerquote, die Bevölkerung in Mio, die Arbeitskosten je Stunde, die Anzahl der Streiktage je 1000 Beschäftigte sowie die Anzahl der in Betrieb befindlichen Atomkraftwerke für 20 Länder enthält.

a) Führe eine Hauptkomponentenanalyse mit den Daten durch und interpretiere das Ergebnis. (2)

b) Führe eine Hauptkomponentenanalyse mit den gewichteten Daten $X_i/\sqrt{\operatorname{Var} X_i}$, $i = 1, \ldots, 9$, durch. Wie erklärt sich der Unterschied? (3)

c) Führe eine Hauptkomponentenanalyse mit den gewichteten Daten durch, wobei die Nebenbedingung $|\alpha_k| = 1$ ersetzt wird durch $|\alpha_k| = \sqrt{\lambda_k}$, $k = 1, \ldots, 9$. (3)

d) Bei der Frage, wieviele Hauptkomponenten verwendet werden sollen, kann ein sog. Scree-Test (scree=Geröllhalde) helfen. Hierbei wird ein Scatterplot erzeugt, bei dem die Eigenwerte der empirischen Kovarianzmatrix der Daten gegen deren Indizes abgetragen werden, wobei gilt $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. Verbindet man die Punkte durch Streckenzüge, entsteht häufig eine Form, die an den Fuß eines Berges erinnert. Diejenigen Eigenwerte, die in etwa horizontal liegen (in der Geröllhalde), führen dazu, die zugehörigen Hauptkomponenten aus den weiteren Überlegungen auszuschließen. Als Wert für die Anzahl der Hauptkomponenten wählt man diejenige Nummer, deren zugehöriger Eigenwert der letzte am Berg vor der Geröllhalde ist. Mathematisch formuliert: Betrachte nur die Hauptkomponenten, deren Eigenwerte größer als das arithmetische Mittel sämtlicher Eigenwerte sind. Bestimme mit Hilfe des Scree-Tests die Anzahl der Hauptkomponenten für a), b) und c). (6)

Aufgabe 2 (5)

Betrachte die ersten beiden Spalten der Daten `pca-example.txt` von Blatt 12, Aufgabe 3. Nimm an, dass die Werte Realisierungen eines zweidimensionalen, normalverteilten Zufallsvektors mit Erwartungswertvektor $\mu = (0, 0)^\top$ und Kovarianzmatrix $\Sigma = \begin{pmatrix} 3 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ sind. Plotte die Konturellipsen der entsprechenden Normalverteilungs-Dichte für $c = 1, 2, 3$ gemeinsam mit den Eigenvektoren α_1 und α_2.