

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2009
Blatt 11

Aufgabe 11.1 (5 Punkte)

Betrachten Sie erneut den Schweizer-Banknoten-Datensatz. Frühere Aufgaben zeigten, dass es Unterschiede zwischen den beiden Gruppen (echte und falsche Scheine) gibt.

- a) Führen Sie Hauptkomponentenanalysen für den gesamten Datensatz durch und zwar sowohl
 - auf der Basis von Kovarianzen als auch
 - auf der Basis von Korrelationen.
- b) Wieviele Hauptkomponenten würden Sie wählen, um eine Dimensionsreduktion durchzuführen?
Für welche der beiden Hauptkomponentenanalysen aus Teil a) sollten die Loadings überhaupt interpretiert werden? Interpretieren Sie die Loadings der ersten Hauptkomponente.
- c) Erstellen Sie für beide Hauptkomponentenanalysen aus Teil a) einen Biplot (in R mit der Funktion `biplot.princomp` möglich). Welche Scores-Struktur liegt vor? Vergleichen und interpretieren Sie die Biplots.

Aufgabe 11.2 (5 Punkte)

- a) Zeigen Sie, dass sich stationäre AR(1) - Prozesse langfristig (d. h. für große t) um $\frac{\beta_1}{1-\beta_2}$ einschwingen.

Die Datei `aussenhandel.txt` enthält Viermonatsdaten (Trimesterdaten) für die Einnahmen y_t der BRD aus dem Reiseverkehr mit dem Ausland (gemessen in Mio. DM).

- b) Stellen Sie die Zeitreihe graphisch dar. Liegt ein Trend vor? Wenn ja, um welche Art von Trend handelt es sich?
- c) Schätzen Sie den Trend der Zeitreihe durch eine lineare Trendfunktion nach der Methode der Kleinsten Quadrate. Zeichnen Sie die geschätzte Trendgerade in das Schaubild der Originalzeitreihe mit ein. Hierzu sind in R die Funktionen `lm` und `abline` nützlich.

- d) Führen Sie eine Trendbereinigung mit einfachen gleitenden Durchschnitten der Längen 3 und 7 durch und zeichnen Sie die geglätteten Reihen ebenfalls in das Schaubild ein. Für die Berechnung gleitender Durchschnitte in R stehen mehrere Funktionen in verschiedenen Paketen zur Verfügung, z. B. die Funktion `runmean` im Paket `caTools`.