
Praktikumsaufgaben 5

WiMa-Praktikum

Aufgabe 1 Berechnen Sie für den Datensatz `divorceUS.dat` die gebräuchlichen Ähnlichkeitskoeffizienten mit der Prozedur `DISTANCE`. Dieser Datensatz enthält gerichtlich anerkannte Scheidungsgründe aus zwanzig US-Bundesstaaten aus dem Jahr 1982. Eine 1 bzw. 0 bedeutet, dieser Scheidungsgrund ist in dem am Anfang der Zeile aufgeführten Bundesstaat gerichtlich anerkannt bzw. nicht anerkannt. Zu den Gründen zählen Inkompatibilität, Gewalt, Verlassen worden sein, Mangel an Unterstützung, Alkohol, Verbrechen, Impotenz, Geisteskrankheit und Trennung.

Aufgabe 2 Der Wirtschaftsdatensatz `economy.dat` (Institut der deutschen Wirtschaft (1993), Nr. 147-158) enthält folgende Daten von 20 bestimmten Industrieländern: Die Investitionsquote (`invest`), die Inflationsrate (`inflatn`), das Bruttosozialprodukt (`gnp`), die Steuerquote (`tax`), Zunahme des Bruttoinlandsprodukts gegenüber dem Vorjahr (`nukes`), die Arbeitslosenquote (`unempld`), Arbeitskosten je Stunde in DM (`labcost`), die Bevölkerung in Millionen (`populatn`) und die Anzahl der Streiktage je 1000 (`strike`).

Führen Sie eine Clusteranalyse mit den bekannten hierarchischen Verfahren durch und vergleichen Sie diese anhand von Dendrogrammen und φ -Funktionen. Hierfür können die Prozeduren `CLUSTER`, `TREE` und `GLOT` verwendet werden. Dabei wird in `PROC GLOT` der in `PROC TREE` berechnete Datensatz nach Anzahl der Cluster sortiert übergeben. Bei `PROC CLUSTER` unterdrückt die Option `NONORM` im `PROC` Statement die voreingestellte Normierung der Fusionsdistanzen.

Folgender Code kann hilfreich sein:

```
DATA economy ;
INFILE 'Pfad/economy.dat' ;
INPUT country $ 1-15 invest inflatn gnp tax nukes unempld labcost
        populatn strike ;
RUN;

PROC CLUSTER DATA = ... METHOD = ... OUTTREE = ... NONORM;
...
RUN;

PROC TREE GRAPHICS DATA= ... horizontal;
RUN;

PROC ...
```

Aufgabe 3 Erstellen Sie zu den Daten in `agriculture.dat` einen farbkodierten Plot der Punkte nach den einzelnen Clustern mittels der Prozedure `GLOT` und berechnen Sie die Mahalanobis-Distanz. Die Kovarianzmatrix kann mit der Prozedur `CORR` berechnet werden.

Aufgabe 4 Zeigen Sie, dass der quadrierte euklidische Abstand $\| \cdot \|_2^2$ keine Metrik ist.