

Aufgabenserie 7 zur Vorlesung "Computergestützte Datenanalyse"

1. Führen Sie für die Datensätze "tiermilch", "essen", "margarine4", "staedte", "mineralwasser", "monde", "waschmaschine", "essen2" und "autod" eine Clusteranalyse durch. Dabei ist sowohl ein hierarchisches Verfahren als auch der k-means-Algorithmus anzuwenden. Lassen Sie sich zu den Ergebnissen entsprechende Grafiken (Dendrogramm, Mittelwertplot) ausgeben. Wählen Sie beim hierarchischen Verfahren eine geeignete Anzahl von Clustern aus. Hängt dabei das Ergebnis von der Wahl der Distanz bzw. von der Wahl des Fusionsverfahrens ab? Alternativ können folgende Datensätze analysiert werden: "galapagos", "naehrstoffinh", "kinder", "loesungsmittel", "elastholz", "teigwaren", "geburttod", "toepferware", "degzns", "azidose", "wdaten", "wstruktur", "usuntern", "schlaf".