
Arbeitsblatt 11

A 1 Analysieren Sie den Datensatz `Ex11.csv` mit Hilfe eines k-Means-Algorithmus. Wählen Sie anhand geeigneter Gütekriterien ein k sinnvoll aus.

A 2 Der Datensatz `clust` (Moodle: `clust.RData`) besteht aus 126 Beobachtungen und 4 Merkmalen.

- Führen Sie ein Clustering mit der Funktion `kmeans` in R für $k = \{2, \dots, 7\}$ durch und erstellen Sie sich eine Visualisierung der Clusterungen.
- Berechnen Sie die Silhouetten und Silhouettenkoeffizienten für die Clusterungen aus a) und visualisieren Sie sich diese. (Im Package `cluster` Methoden `silhouette` und `plot.silhouette`)
- Für welche Clusterung würden Sie sich entscheiden? Begründen Sie Ihre Entscheidung.
- Das Ergebnis einer Clusterung mittels k-Means-Alg. hängt von den zufällig gewählten Mittelwerten beim Start ab. Überlegen Sie sich eine Möglichkeit um den Silhouettenkoeffizienten trotzdem sinnvoll einsetzen zu können und implementieren Sie diese mit Hilfe von R.