

Checkliste Lernziele:

- Wie führt man hierarchische Klassifikationen mit R durch?
- Wie kann man Dendrogramme in R erstellen?
- Wie lassen sich Minimalbäume berechnen und darstellen?
- Wie lassen sich optimale ultrametrische Modelle in R schätzen?

Hinweis: Für einige der Aufgaben werden die Zusatzpakete `vegan` und `clue` benötigt, welche heruntergeladen und installiert werden müssen. Zudem wurden manche der nötigen Arbeitsschritte bereits in vorhergehenden Sitzungen gelöst. Benutzen Sie die entsprechenden Skripte, die Sie hierfür erstellt haben. Hier finden Sie Angaben dazu, welche Parameter der `method`-Option des `hclust`-Befehls mit welchem SAHN-Algorithmus verbunden sind:

method-Parameter	SAHN-Algorithmus
<code>single</code>	Single Link(age) Methode
<code>complete</code>	Complete Link(age) Methode
<code>mcquitty</code>	WPGMA (Weighted Average) Methode
<code>median</code>	WPGMC (Weighted Centroid) Methode
<code>average</code>	UPGMA (Group Average) Methode
<code>centroid</code>	UPGMC (Unweighted Centroid) Methode
<code>ward</code>	Wards Minimum Variance Methode

Aufgaben:

1. Laden Sie den Datensatz `bs1.dat` und replizieren Sie Tabelle 2.3-3.
2. Verwenden Sie den Befehl `hclust`, um eine hierarchische Klassifikation mittels der Single Link Methode zu berechnen. Wie kann man Box 8.1-2 erzeugen?

3. Verwenden Sie die Befehle `plot`, `cutree` und `rect.clust`, um ein Dendrogramm zu erstellen und eine 3-Cluster-Lösung in dieses einzuzichnen. Wie unterscheidet sich die resultierende Grafik von Abbildung 8.1-1? Interpretieren Sie die Ergebnisse.
4. Benutzen Sie alle in Box 8.1-1 vorgestellten Verfahren, um eine hierarchische Klassifikation der Daten vorzunehmen. Vergleichen Sie die Dendrogramme der verwendeten SAHN-Algorithmen. Zeichnen Sie in jedes Dendrogramm eine 3-Cluster-Lösung ein und diskutieren Sie die Ergebnisse.
5. Laden Sie den Datensatz `klaus1.dat`. Führen Sie eine hierarchische Klassifikation mittels Wards Methode durch. Erstellen und interpretieren Sie das dazugehörige Dendrogramm. Lassen Sie sich von R sowohl eine 2-Cluster-Lösung als auch eine 4-Cluster-Lösung erstellen und vergleichen Sie diese.
6. Führen Sie Aufgabe 1 erneut aus. Erstellen Sie mit den Befehlen des `vegan`-Pakets Abbildung 8.2-1 und Tabelle 8.2-1. *Tipp:* Die Abstände d_{ij}^v werden auch als kophenetische Distanzen (*cophenetic distances*) bezeichnet.
7. Verwenden Sie den Befehl `cl.ultrametric` um Tabelle 8.2-2 nachzubilden. Wenden Sie den Befehl `plot` sowohl auf die hierarchische Klassifikation, als auch auf das mit dem Befehl `cl.ultrametric` erzeugte Ergebnis an. Welche Unterschiede zeigen sich?
8. Verwenden Sie den Befehl `ls.fit.ultrametric` um Tabelle 8.2-4 und Abbildung 8.2-2 zu replizieren. Welche Probleme zeigen sich hierbei? Wie kann man diese eventuell lösen?