

Aufgabe 18:

Für die folgende Datenmatrix

Person	1	2	3	4	5
Alter	23	46	51	60	34
Einkommen (in Tsd)	25	55	100	84	40

wurde zum Teil bereits die Matrix der Distanzen zwischen den Personen berechnet:

$$D = \begin{bmatrix} 0 & 37.8 & 80.06 & 69.64 & 18.6 \\ & 0 & \dots & 32.20 & 19.21 \\ & & 0 & 18.36 & 62.36 \\ & & & 0 & 51.11 \\ & & & & 0 \end{bmatrix}.$$

- (a) Welches Distanzmaß wurde bei der Berechnung verwendet? Begründen Sie Ihre Antwort anhand eines der schon berechneten Elemente. Berechnen Sie danach die noch fehlenden Elemente in D .
- (b) Bestimmen Sie ausgehend von fünf einelementigen Klassen der Anfangspartition

$$C^0 = \{C_1^0, \dots, C_5^0\} = \{\{1\}, \dots, \{5\}\}$$

eine Hierarchie nach folgenden Verfahren. Zeichnen und interpretieren Sie anschließend die zugehörigen Dendrogramme.

- Single Linkage
 - Complete Linkage
- (c) Nennen Sie zwei andere übliche Linkage-Verfahren und erklären Sie kurz, was die beiden Verfahren unterscheidet.
- (d) Wann ist es sinnvoll, vor der Berechnung der euklidischen Distanzmatrix die Variablen zu skalieren (vgl. Aufgabe 19)?

Selbststudium:**Aufgabe 19:**

Die folgende Tabelle enthält für fünf Länder die Lebenserwartung von Frauen und die Anzahl der Kinder pro Frau.

Objekt	Lebenserwartung	Kinder pro Frau
Deutschland	81.1	1.29
Indien	64.9	2.97
Indonesien	79.6	2.02
Israel	81.0	2.70
Japan	85.0	1.33

Die empirische Kovarianzmatrix der beiden Variablen lautet

$$\mathbf{S} = \begin{bmatrix} 60.32 & -4.39 \\ -4.39 & 0.59 \end{bmatrix}.$$

Im Folgenden ist das Ziel, für diese 5 Länder ein hierarchisches Clustering mit der euklidischen Distanz durchzuführen.

- (a) Berechnen Sie die standardisierten Werte.
- (b) Zeigen Sie, dass sich folgende euklidische Distanzmatrix für die standardisierten Variablen ergibt:

$$\mathbf{D} = \begin{bmatrix} 0 & 3.0 & 1.0 & 1.8 & 0.5 \\ & 0 & 2.3 & 2.1 & 3.4 \\ & & 0 & 0.9 & 1.1 \\ & & & 0 & 1.8 \\ & & & & 0 \end{bmatrix}$$

- (c) Führen Sie eine Clusteranalyse mit Complete Linkage durch.
- (d) Zeichnen und interpretieren Sie das zugehörige Dendrogramm.