

2. Suppose that we have four observations, for which we compute a dissimilarity matrix, given by

$$\begin{bmatrix} & 0.3 & 0.4 & 0.7 \\ 0.3 & & 0.5 & 0.8 \\ 0.4 & 0.5 & & 0.45 \\ 0.7 & 0.8 & 0.45 & \end{bmatrix}.$$

For instance, the dissimilarity between the first and second observations is 0.3, and the dissimilarity between the second and fourth observations is 0.8.

- On the basis of this dissimilarity matrix, sketch the dendrogram that results from hierarchically clustering these four observations using complete linkage. Be sure to indicate on the plot the height at which each fusion occurs, as well as the observations corresponding to each leaf in the dendrogram.
- Repeat (a), this time using single linkage clustering.
- Suppose that we cut the dendrogram obtained in (a) such that two clusters result. Which observations are in each cluster?
- Suppose that we cut the dendrogram obtained in (b) such that two clusters result. Which observations are in each cluster?
- It is mentioned in the chapter that at each fusion in the dendrogram, the position of the two clusters being fused can be swapped without changing the meaning of the dendrogram. Draw a dendrogram that is equivalent to the dendrogram in (a), for which two or more of the leaves are repositioned, but for which the meaning of the dendrogram is the same.

Answer

a)

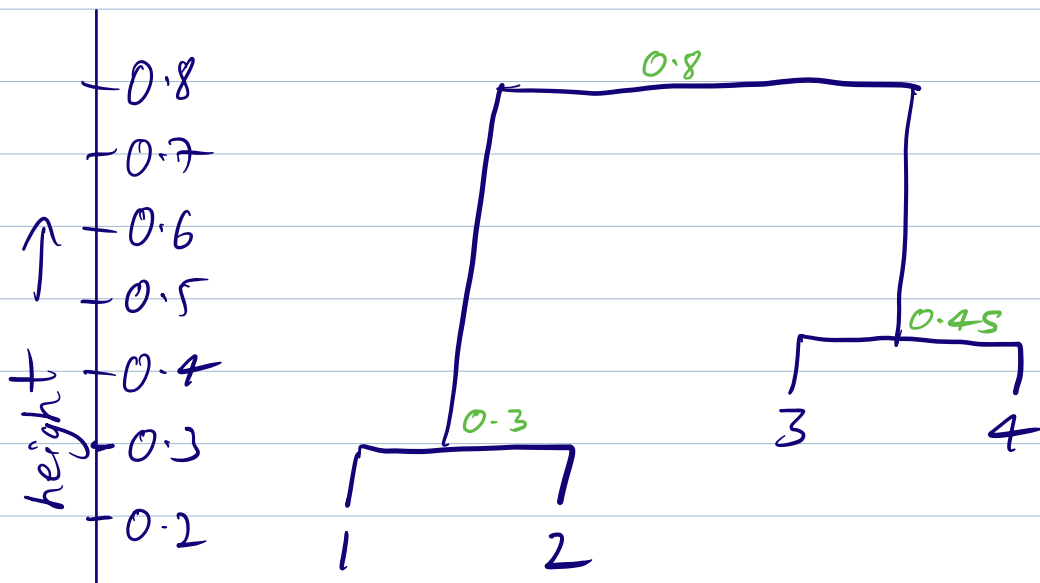
	1	2	3	4
1	0	<u>0.3</u>	0.4	0.7
2	<u>0.3</u>	0	0.5	0.8
3	0.4	0.5	0	0.45
4	0.7	0.8	0.45	0

1,2

$$\begin{array}{c} 1,2 \\ 3 \\ 4 \end{array} \begin{array}{c} 1,2 \quad 3 \quad 4 \\ \left[\begin{array}{ccc} 0 & 0.5 & 0.8 \\ 0.5 & 0 & \underline{0.45} \\ 0.8 & \underline{0.45} & 0 \end{array} \right] \end{array} \quad \begin{array}{c} 1,2 \\ 3,4 \end{array}$$

$$\begin{array}{c} 1,2 \\ 3,4 \end{array} \begin{array}{c} 1,2 \quad 3,4 \\ \left[\begin{array}{cc} 0 & \underline{0.8} \\ \underline{0.8} & 0 \end{array} \right] \end{array} \quad 1,2,3,4$$

$$1,2,3,4 \left[\begin{array}{c} 1,2,3,4 \\ 0 \end{array} \right]$$



b)

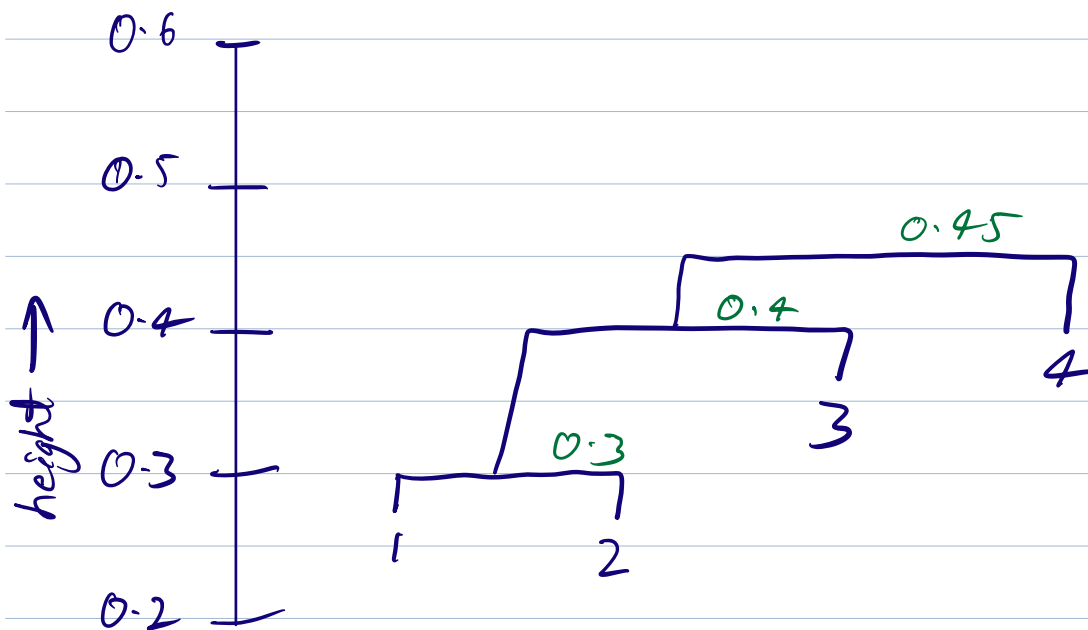
$$\begin{array}{c}
 1 \\
 2 \\
 3 \\
 4
 \end{array}
 \begin{bmatrix}
 & 1 & 2 & 3 & 4 \\
 1 & 0 & \underline{0.3} & 0.4 & 0.7 \\
 2 & \underline{0.3} & 0 & 0.5 & 0.8 \\
 3 & 0.4 & 0.5 & 0 & 0.45 \\
 4 & 0.7 & 0.8 & 0.45 & 0
 \end{bmatrix}
 \begin{array}{c}
 \\
 1, 2 \\
 \\
 \\
 \end{array}$$

$$\begin{array}{c}
 1, 2 \\
 3 \\
 4
 \end{array}
 \begin{bmatrix}
 & 1, 2 & 3 & 4 \\
 1, 2 & 0 & \underline{0.4} & 0.7 \\
 3 & \underline{0.4} & 0 & 0.45 \\
 4 & 0.7 & 0.45 & 0
 \end{bmatrix}
 \begin{array}{c}
 \\
 1, 2, 3 \\
 \\
 \\
 \end{array}$$

$$\begin{array}{c}
 1, 2, 3 \\
 4
 \end{array}
 \begin{bmatrix}
 & 1, 2, 3 & 4 \\
 1, 2, 3 & 0 & \underline{0.45} \\
 4 & \underline{0.45} & 0
 \end{bmatrix}
 \begin{array}{c}
 \\
 1, 2, 3, 4 \\
 \\
 \\
 \end{array}$$

$$1, 2, 3, 4 \begin{bmatrix} & 1, 2, 3, 4 \\ & 0 \end{bmatrix}$$

(Diagram on Next Page)



c) If we cut the dendrogram, we get 2 clusters as follows:

Cluster 1 \rightarrow 1, 2

Cluster 2 \rightarrow 3, 4

d) If we cut the dendrogram, we get 2 clusters as follows:

Cluster 1 \rightarrow 1, 2, 3

Cluster 2 \rightarrow 4

(Next Page)

e)

