Tut 11: More Hierarchical Clustering

May/June 2022

Hierarchical Clustering

1. (May 2020 Final Q3(b)) Given an appropriate example to explain why the Minkowski distance

$$M(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{p} |x_i - y_i|^r \right)^{\frac{1}{r}}, \quad \boldsymbol{x},\ \boldsymbol{y} \in \mathbb{R}^p$$

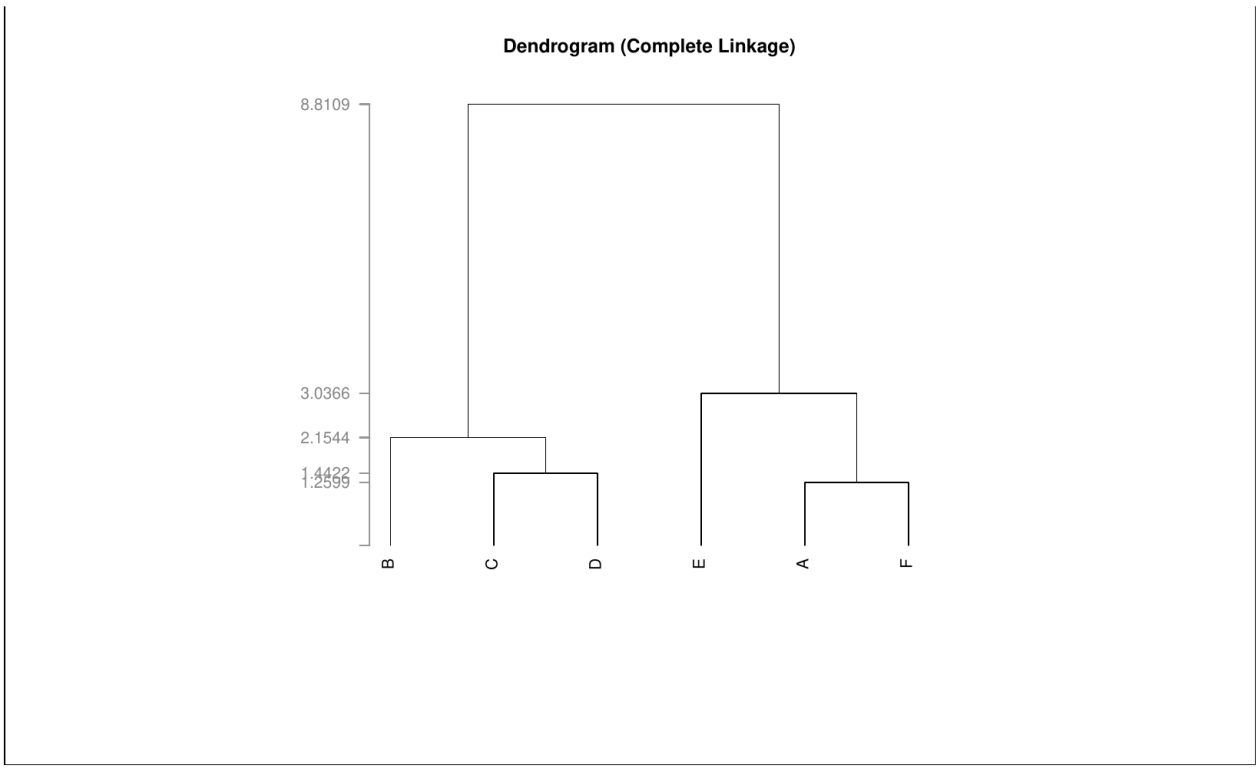
will no longer be a distance function when $r = \frac{1}{2}$. (2 marks)

2. (May 2020 Final Q3(c)) Group the observations in Table 3.1 using hierarchical clustering and the **Minkowski distance** with $r = 3$ (refer to part (b) for the definition of Minkowski distance) and **complete linkage** and draw the dendrogram formed by the hierarchical clustering.

Table 3.1: Unlabelled data.

Obs	x_1	x_2	x_3
A	1	3	2
B	5	7	9
C	6	9	8
D	7	8	9
E	2	3	5
F	1	4	3

(4 marks)

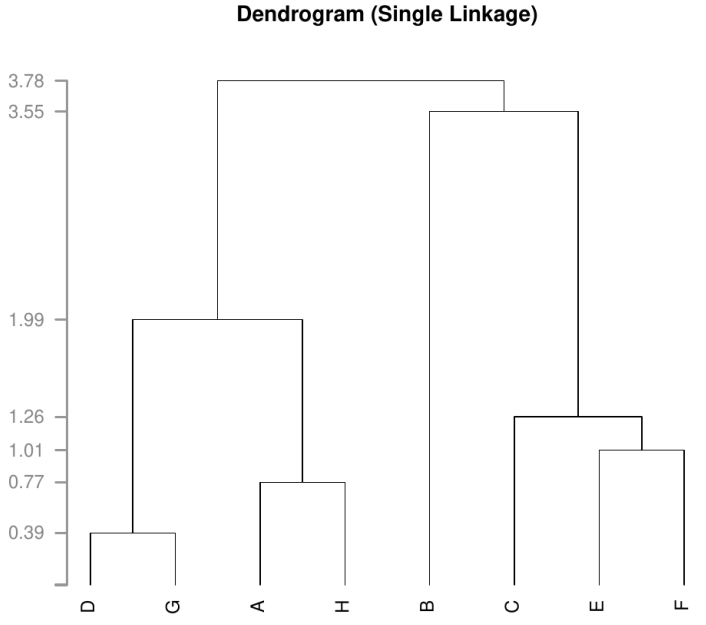


3. (Jan 2021 Final Q4(a). Hand calculation is possible but Excel/R is recommended) Group the observations in Table 4.1 using hierarchical clustering and the **Manhattan distance** and **single linkage** and draw the dendrogram formed by the hierarchical clustering.

Table 4.1: Unlabelled data.

Obs	x_1	x_2
A	-2.68	-2.02
B	3.06	-0.83
C	1.91	1.57
D	-1.06	-0.88
E	0.49	2.42
F	0.83	1.75
G	-0.71	-0.84
H	-2.01	-1.92

(5 marks)



4. (Jan 2022 Final Q5(b)) Given the three-dimensional points in Table 5.2,

Table 5.2: Three-dimensional points.

Label	x_1	x_2	x_3
P_1	3.3	4.4	2.5
P_2	2.4	3.1	2.1
P_3	0.1	1.9	1.1
P_4	0.3	2.4	1.5
P_5	-0.6	1.1	1.1
P_6	-2.9	-0.1	0.1
P_7	4.3	6.4	5.5
P_8	3.4	5.1	5.1
P_9	1.1	3.9	4.1

Use the k-means clustering method with **Manhattan distance** to cluster the given points into $k = 3$ clusters by using P_5, P_4, P_7 as the initial clusters, find the **stable cluster centres**. (8 marks)