Chair of Decision Sciences and Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Business Analytics & Machine Learning
Homework sheet 8: Clustering

Prof. Dr. Martin Bichler, Prof. Dr. Jalal Etesami
Julius Durmann, Markus Ewert, Johannes Knörr, Yutong Chao

Exercise H8.1 *Hierarchical clustering*

You are given the following dataset:

i	x_i	y_i
0	0	0
1	1	5
2	4	2
3	-1	4
4	-2	-3
5	3	4
6	1	-5
7	0	1
8	-3	0
9	-1	-1

Table 1 Dataset

a) Complete the missing entries in the distance matrix below based on the L1 norm (Manhattan distance)

$$D_1 = \begin{bmatrix} 0 & \bigcirc & 6 & 5 & 5 & 7 & 6 & 1 & 3 & 2 \\ & 0 & 6 & 3 & 11 & 3 & \bigcirc & 5 & 9 & 8 \\ & & 0 & 7 & 11 & \bigcirc & 10 & 5 & 9 & 8 \\ & & & 0 & 8 & 4 & 11 & 4 & 6 & \bigcirc \\ & & & & 0 & 12 & 5 & 6 & 4 & 3 \\ & & & & & 0 & 11 & \bigcirc & 10 & 9 \\ & & & & & & 0 & 7 & 9 & 6 \\ & & & & & & & 0 & 4 & 3 \\ & & & & & & & & 0 & 3 \\ & & & & & & & & & 0 \end{bmatrix}$$

b) Complete the missing entries in the distance matrix below based on the L2 norm (Euclidean distance)

$$D_2 = \begin{bmatrix} 0 & 5.10 & 4.47 & 4.12 & 3.61 & 5 & 5.10 & 1 & 3 & \bigcirc \\ & 0 & 4.24 & 2.24 & \bigcirc & 2.24 & 10 & 4.12 & 6.40 & 6.32 \\ & & 0 & \bigcirc & 7.81 & 2.24 & 7.62 & 4.12 & 7.28 & 5.83 \\ & & & 0 & 7.07 & 4 & 9.22 & 3.16 & 4.47 & \bigcirc \\ & & & & 0 & 8.60 & 3.61 & 4.47 & 3.16 & 2.24 \\ & & & & & 0 & 9.22 & 4.24 & 7.21 & 6.40 \\ & & & & & & 0 & 6.08 & \bigcirc & 4.47 \\ & & & & & & & 0 & 3.16 & 2.24 \\ & & & & & & & & 0 & 2.24 \\ & & & & & & & & & 0 \end{bmatrix}$$

c) Perform bottom-up (agglomerative) hierarchical clustering using the Manhattan distance d_1 as distance measure. Use *complete-linkage clustering*, i.e. the distance between two sets of observations A, B is defined as $\max_{a \in A, b \in B} d(a, b)$. As tiebreaker rule, merge clusters in the order of their label i.

d) Draw the dendrogram. Which is a reasonable number of clusters in this example? Discuss your reasoning.

Exercise H8.2 *EM algorithm*

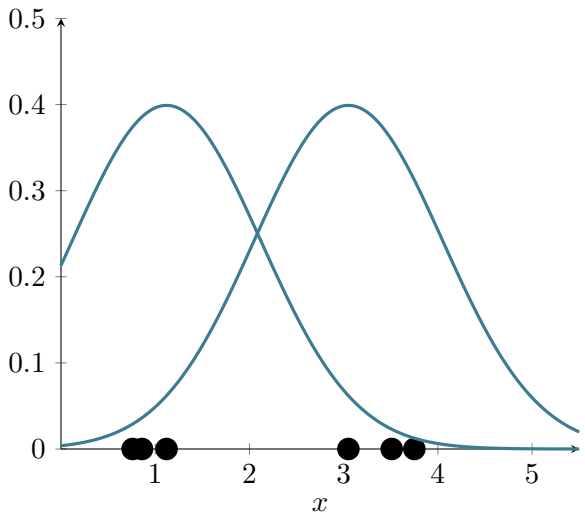
Given $k = 2$, perform EM algorithm with the following instances and initial distribution parameters:

Instance	Value
1	0.76
2	0.86
3	1.12
4	3.05
5	3.51
6	3.75

Table 2 Dataset

Parameter	Value
μ_A	1.12
σ_A	1.00
p_A	50%
μ_B	3.05
σ_B	1.00
p_B	50%

Table 3 Initial Distribution Parameters



Exercise H8.3 *Ensemble methods*

a) Name benefits that an ensemble model (ideally) has in comparison to a single model.

b) In terms of the training process, what is a major difference between bagging and boosting?