



Exercise Sheet 8

Intelligent Systems

December 18, 2019

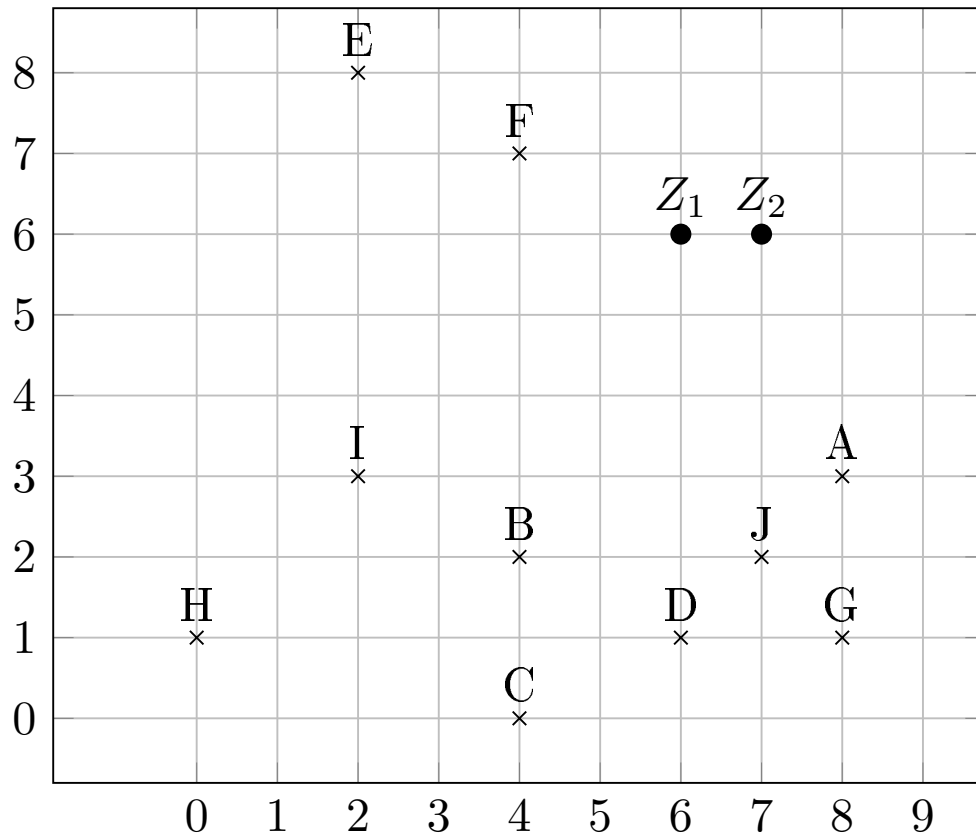
Clustering

Exercise 1 - Single, Complete und Average Linkage

- A. How does Single Linkage Clustering work? Visualise the procedure using the data set of Figure 1 with $C = 3$ and Figure 2 with $C = 2$. As distance measure choose the Manhattan distance.
- B. What are Pros and Cons of Single Linkage in comparison to complete Linkage regarding the treatment of outliers and the tendency of producing chains?
- C. What is the difference between Single Linkage, Complete Linkage, and Average Linkage? How would Complete or Average Linkage cluster the data points in Figure 1 with $C = 2$ and the Manhattan distance?

Exercise 2 - c-Means Clustering

- A. Proceed 4 iterations of the c-Means clustering on the points given in Figure 3. Note the Euclidean distances into the table. Remark: You can use a ruler to measure the Euclidean distances.

Figure 3: Points $A - J$ and initial center distribution Z_1, Z_2 .

Iteration		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>
1	$Z_1 = (\quad , \quad)$										
	$Z_2 = (\quad , \quad)$										
2	$Z_1 = (\quad , \quad)$										
	$Z_2 = (\quad , \quad)$										
3	$Z_1 = (\quad , \quad)$										
	$Z_2 = (\quad , \quad)$										
4	$Z_1 = (\quad , \quad)$										
	$Z_2 = (\quad , \quad)$										

Exercise 3 - c-Means

- Visualise and explain with the help of Figure 4 how the *c*-Means algorithm works.
- What are Pros and Cons of the *c*-Means algorithm?
- Which steps can be applied to optimise the clustering results?

Signature Task - Clustering Challenge

In clustering, the labels (here: the author of the signature) in the training data can be removed. Each training sample (recording of a single signature) should be treated equally. The task is to identify clusters of same authors. To present your results make a short presentation with 4 slides describing the following steps:

- A. Preprocessing (Outlier detection/Normalisation/Segmentation)
- B. Extraction of your lower dimensional features
- C. Dimensionality reduction via PCA
- D. Clustering (DBSCAN, C-Means, NN)

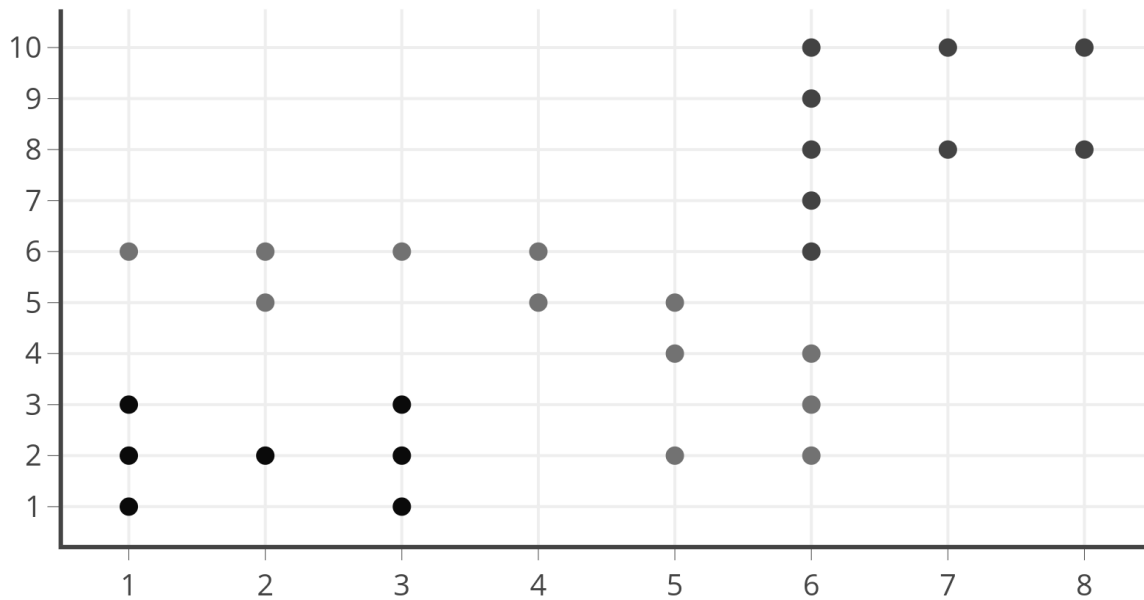


Figure 1: Single Linkage Clustering C=3

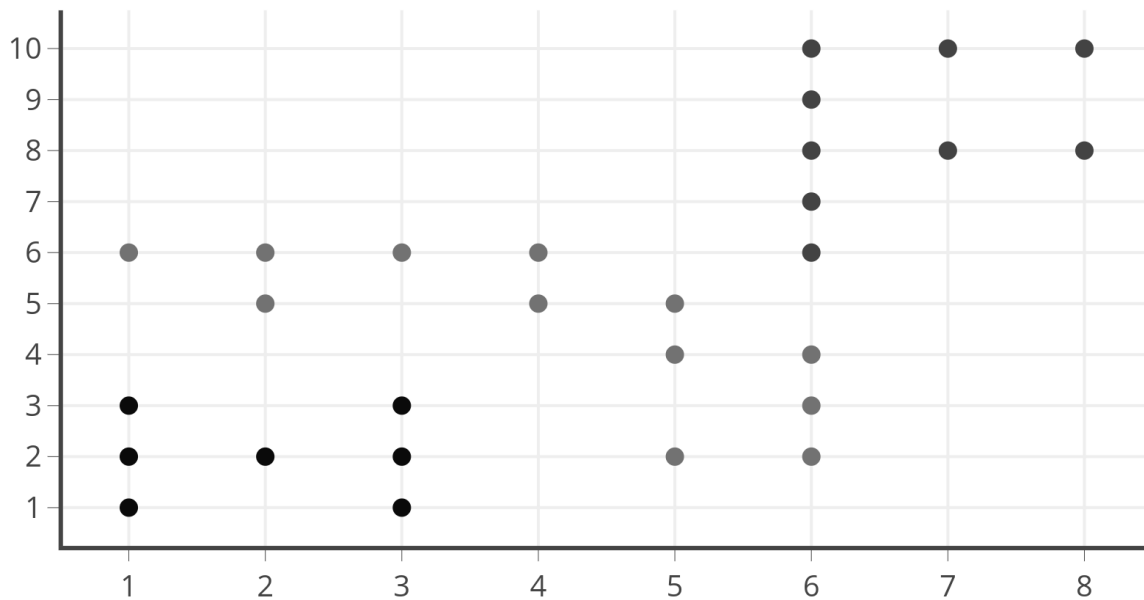


Figure 2: Single Linkage Clustering C=2

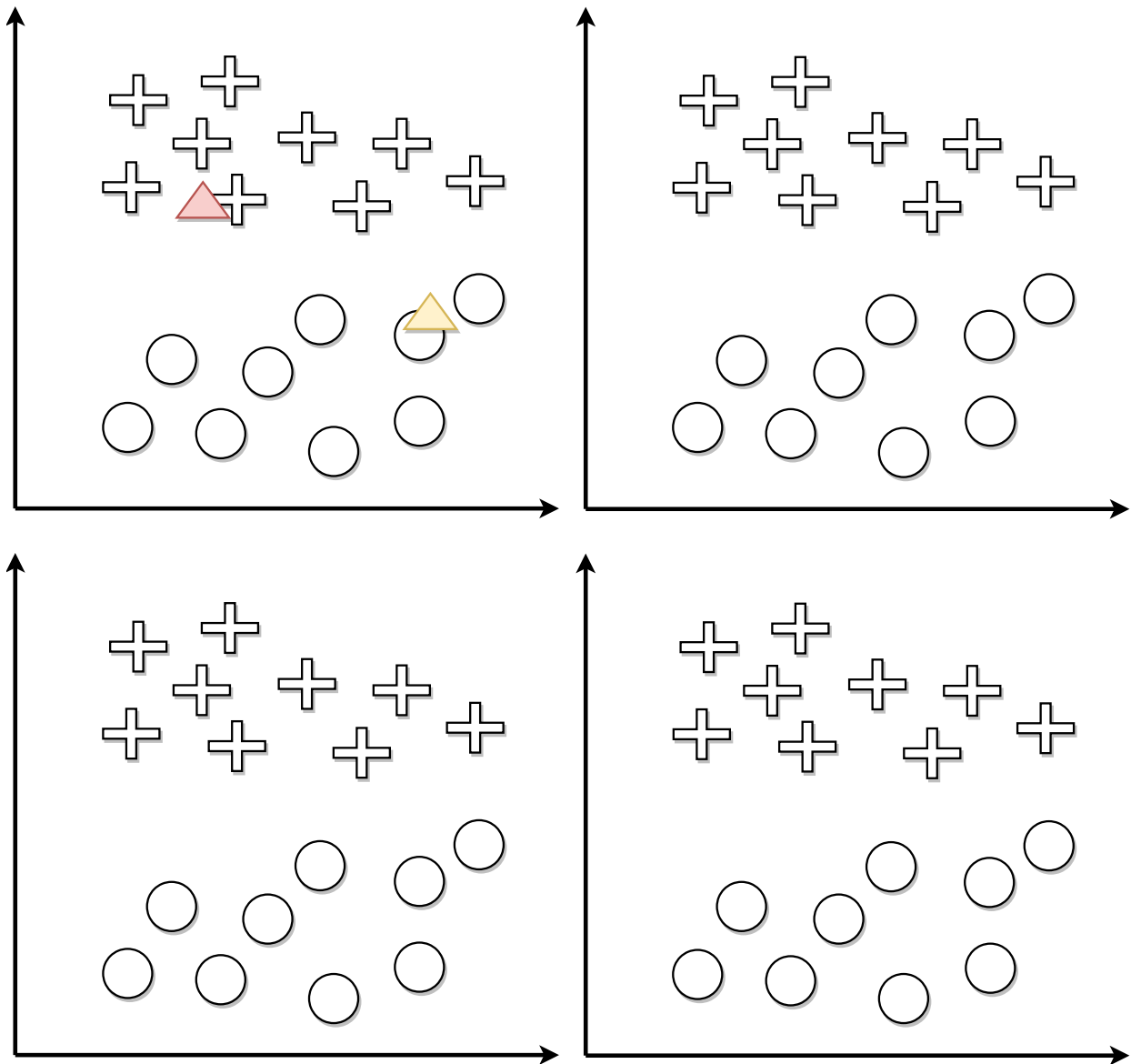


Figure 4: Clustering with c-means