
Assignment #3

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari

7072310

raad00002@stud.uni-saarland.de

Dhimitrios Duka

7059153

dhdu00001@stud.uni-saarland.de

2 Problem 2 (Hierarchical clustering and dissimilarity)

2.1

Out of the three mentioned clustering methods, only k -means requires the coordinates of the elements being clustered. This is due to the fact that at each iteration of the algorithm, k -means needs to find the mean of all datapoints in a given cluster C_i . In contrast, both k -medoids and agglomerative clustering can work with just the dissimilarity matrix, as they don't create new points, such as the mean of the cluster, or perform coordinate-based calculations. k -medoids selects actual data points as cluster centers and uses only pairwise distances for cluster assignments, while agglomerative clustering builds a hierarchy by progressively merging clusters based solely on distances between existing points. For k -medoids and agglomerative clustering, coordinates are more of a proxy to calculating the dissimilarity scores. Therefore, having direct access to the dissimilarity scores renders the coordinates useless.

2.2

2.2.1 Complete linkage

First, we will start by finding the two most similar clusters. These clusters would be the ones who would end up merging. From the given dissimilarity matrix, we can conclude that the two clusters with the minimum dissimilarity are cluster C and D . So, these two clusters would merge and form a new cluster $[C, D]$. Keeping in mind how the complete linkage method works, we write down the updated similarity matrix:

	{A,B}	{C,D}	E
{A,B}	0	1.8	2.3
{C,D}	1.8	0	1.7
E	2.3	1.7	0

2.2.2 Single linkage

For the single linkage method, we will follow a similar approach to the Complete linkage method. First, we will merge the two most similar clusters, and then we will update the similarity matrix. As in the previous exercise, we will merge clusters C and D . However, while updating the similarity matrix, we will choose the minimum distance between clusters instead of the maximum. The resulting similarity matrix is given in the table below:

	{A,B}	{C,D}	E
{A,B}	0	1.2	2.3
{C,D}	1.2	0	1.4
E	2.3	1.4	0

28 **2.2.3**

The final dendrogram is shown in Figure1

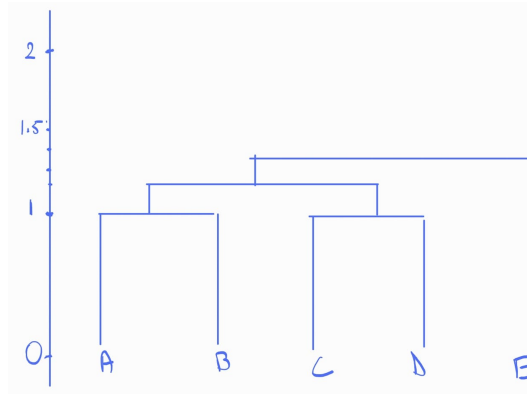


Figure 1: Resulting dendrogram from the provided steps.

29