Assignment #4

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 To perform k-means clustering, k-medoids, or agglomerative clustering, is it necessary to know the coordinates of the elements being clustered, or is it sufficient to have only their dissimilarity matrix? Explain your answer in detail.

Out of the three mentioned clustering methods, only k-means requires the coordinates of the elements being clustered. This is because, at each iteration of the algorithm, k-means needs to find the mean of all data points 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 select actual data points as cluster centers and use only pairwise dissimilarities for cluster assignments, while agglomerative clustering builds a hierarchy by progressively merging clusters based solely on dissimilarities between existing points. For k-medoids and agglomerative clustering, coordinates are more of a proxy for calculating the dissimilarity scores. Therefore, having direct access to the dissimilarity scores renders the coordinates useless.

2.2 The hierarchical clustering algorithm has already identified one cluster, and the dissimilarity matrix below represents the pairwise dissimilarities among the elements...

2.2.1 Complete linkage

First, we will start by finding the two most similar clusters. From the given dissimilarity matrix, we can conclude that the two clusters with the minimum dissimilarity score are cluster C and D. These two clusters would merge and form a new cluster $[C, D]$. Additionally, while updating the similarity matrix, we will choose the maximum dissimilarity between clusters as the new value for the dissimilarity. The resulting similarity matrix is given in the table below:

	{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

2.2.3

The final dendogram is shown in Figure1. First, clusters $\{A\}$ and $\{B\}$ are merged. Next, clusters $\{C\}$ and $\{D\}$. Next clusters $\{A, B\}$ and $\{C, D\}$ are merged. And finally, clusters $\{A, B, C, D\}$ and $\{E\}$ are merged.

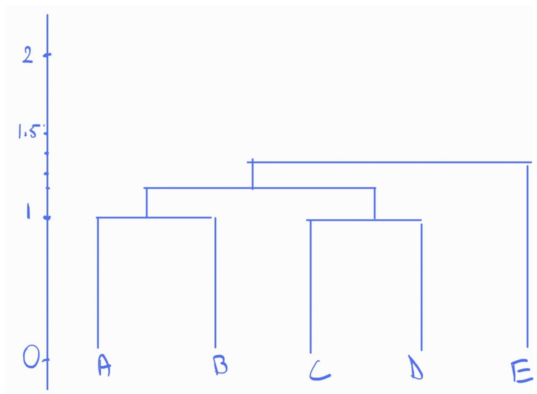

Figure 1: Resulting dendogram from the provided steps.