Q1. State the difference among Single Linkage, Complete Linkage Average Linkage, and Centroid Linkage Hierarchical Clustering Algorithms and give an example of each.

Ans. Hierarchical clustering treats each data point as a singleton cluster, and then successively merges clusters until all points have been merged into a single remaining cluster. A hierarchical clustering is often represented as a [dendrogram](https://nlp.stanford.edu/IR-book/dendrogram.html) (from Manning et al. 1999).

In complete-link (or complete linkage) hierarchical clustering, we merge in each step the two clusters whose merger has the smallest diameter (or: the two clusters with the smallest **maximum** pairwise distance).

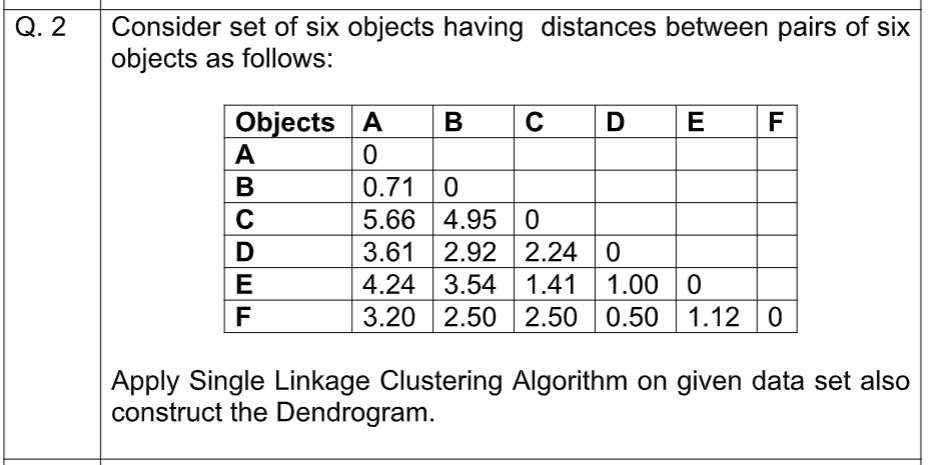
In single-link (or single linkage) hierarchical clustering, we merge in each step the two clusters whose two closest members have the smallest distance (or: the two clusters with the smallest **minimum** pairwise distance).

Complete-link clustering can also be described using the concept of clique. Let dn be the diameter of the cluster created in step n of complete-link clustering. Define graph G(n) as the graph that links all data points with a distance of at most dn. Then the clusters after step n are the cliques of G(n). This motivates the term complete-link clustering.

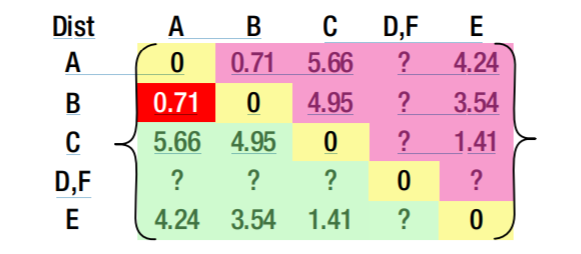
Single-link clustering can also be described in graph theoretical terms. If dn is the distance of the two clusters merged in step n, and G(n) is the graph that links all data points with a distance of at most dn, then the clusters after step n are the connected components of G(n). A single-link clustering also closely corresponds to a weighted graph's [minimum spanning tree](http://www.dcs.gla.ac.uk/~iain/keith/data/pages/57.htm).

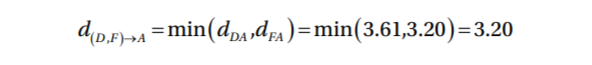
Average Link-is a compromise between the sensitivity of complete-link clustering to outliers and the tendency of single-link clustering to form long chains that do not correspond to the intuitive notion of clusters as compact, spherical objects.

Centroid-Linkage **-**Centroid-linkage is the distance between the centroids of two clusters. As the centroids move with new observations, it is possible that the smaller clusters are more similar to the new larger cluster than to their individual clusters causing an inversion in the dendrogram. This problem doesn’t arise in the other linkage methods because the clusters being merged will always be more similar to themselves than to the new larger cluster.

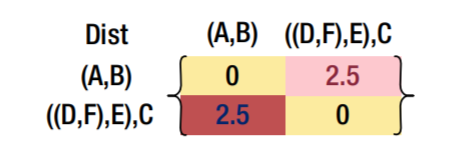


Ans. We see that the minimum distance is between D and F. Thus, we combine D and F.  The resulting matrix now looks like –

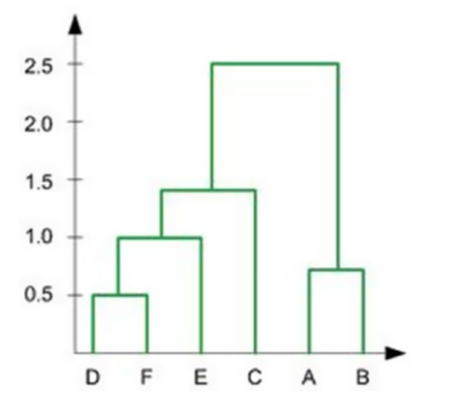




Based on the preceding calculation, we replace the missing value in the distance between {D,F} and A with minimum of the distances between DA and FA. Similarly, we would impute with other missing values. We keep proceeding like that until we are left with-



The resulting cluster can now be represented as-



Q3. Suppose that the data mining task is to cluster the following eight

points (with (x,y) representing location) into three clusters.

A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2),

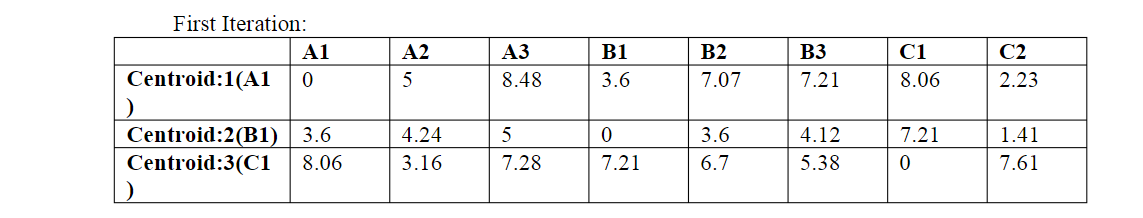
C2(4,9).The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively.

Use the k-means algorithm to show only

I. the three cluster centers after the first round execution, and

II. the final three clusters

Ans. First Iteration



The three clusters with cluster points are –

1= A1(2,10)

2= {A3(8,4), B1(5,8), B2(7,5), C2(4,9)}

3={ A2(2,5), C1(1,2)}

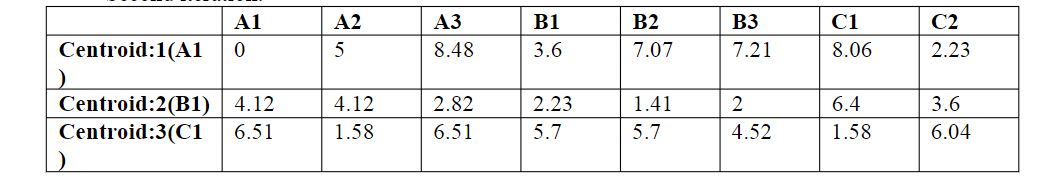
Centres –

1=(2,10)

2={(5+8+7+6+4)/5,(8+4+5+4+9)/5}

3=(1.5,3.5)

Final three Clusters –



Final Clusters are –

1= {A1,C2,B1}

2= {A3,B2,B3}

3= {A2,C1}

Q4. What is Medoid in K-Medoid Algorithm? Consider set of five objects A (0, 0), B (6, 6), C (-3,-3), D (3, 3), and E (-6,-6). Calculate the Medoid of given Data Set.

Ans.

