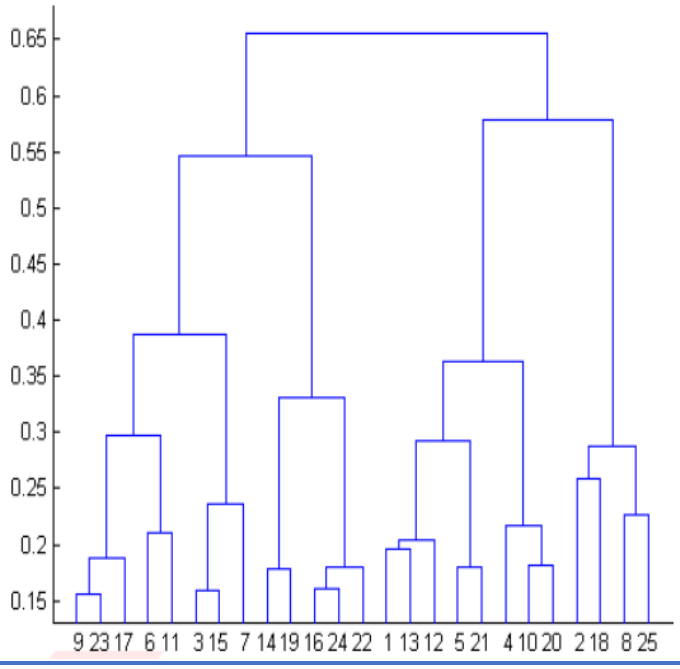
Parth Makwana (DS0722)

Q1 to Q12

1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:



b) 4

2. In which of the following cases will K-Means clustering fail to give good results?

d) 1, 2 and 4

3. The most important part of is selecting the variables on which clustering is based.

d) formulating the clustering problem

4. The most commonly used measure of similarity is the or its square.

a) Euclidean distance

5. ____is a clustering procedure where all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller and smaller clusters.

b) Divisive clustering

6. Which of the following is required by K-means clustering?

d) All answers are correct

7. The goal of clustering is to-

a) Divide the data points into groups

8. Clustering is a-

b) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

d) All of the above

10. Which version of the clustering algorithm is most sensitive to outliers?

Parth Makwana (DS0722)

a) K-means clustering algorithm

11. Which of the following is a bad characteristic of a dataset for clustering analysis-

d) All of the above

12. For clustering, we do not require-

a) Labeled data

Q13 to Q15

13. How is cluster analysis calculated?

Cluster analysis groups related items together using different algorithms to identify the “clusters.” These clusters are latent variables, meaning they aren’t directly measured but instead are inferred from the relationship items have with each other. Cluster analysis is the approach used in card sorting when you want to know how closely products, content, or functions relate from the users’ perspective.

14. How is cluster quality measured?

If all the data objects in the cluster are highly similar then the cluster has high quality. We can measure the quality of Clustering by using the Dissimilarity/Similarity metric in most situations. But there are some other methods to measure the Qualities of Good Clustering if the clusters are alike.

15. What is cluster analysis and its types?

Cluster Analysis is the process to find similar groups of objects in order to form clusters. It is an unsupervised machine learning-based algorithm that acts on unlabelled data. A group of data points would comprise together to form a cluster in which all the objects would belong to the same group.

Types of Cluster analysis:
Hierarchical Cluster Analysis
Centroid-based Clustering
Distribution-based Clustering
Density-based Clustering
Applications and Examples