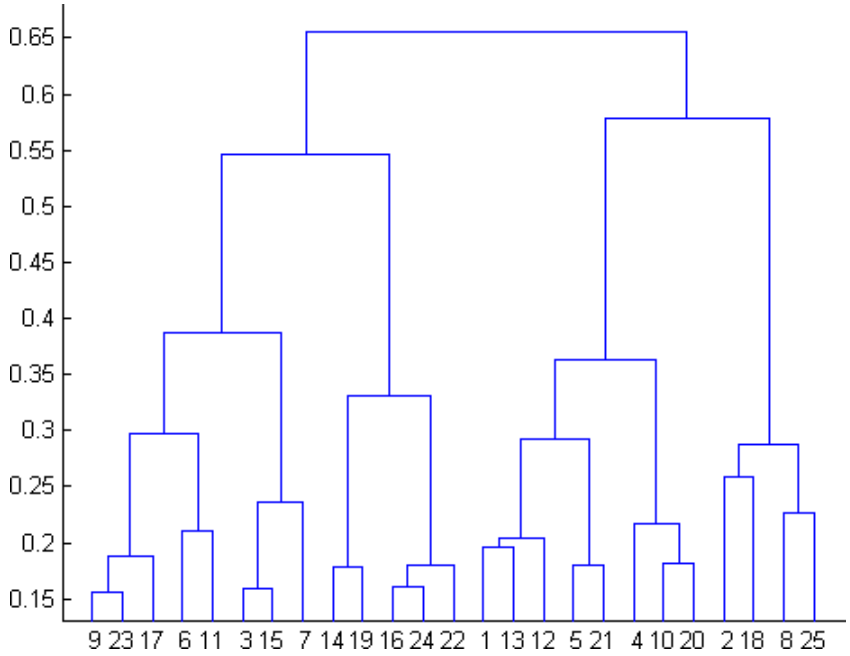
MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

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?
   1. Data points with outliers
   2. Data points with different densities
   3. Data points with round shapes
   4. Data points with non-convex shapes

   Options:

   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

MACHINE LEARNING

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-
   d) All of the above
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?
    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-
    b) Unlabeled data

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.

13. How is cluster analysis calculated?

ANS. The cluster analysis follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.

14. How is cluster quality measured?

ANS. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set. The silhouette coefficient and other intrinsic measures can also be used in the elbow method to heuristically derive the number of clusters in a data set by replacing the sum of within-cluster variances.

15. What is cluster analysis and its types?

ANS. Cluster analysis is the task of grouping a set of data points in such a way that they can be characterized by their relevancy to one another. These types are Centroid Clustering, Density Clustering, Distribution Clustering, and Connectivity Clustering.