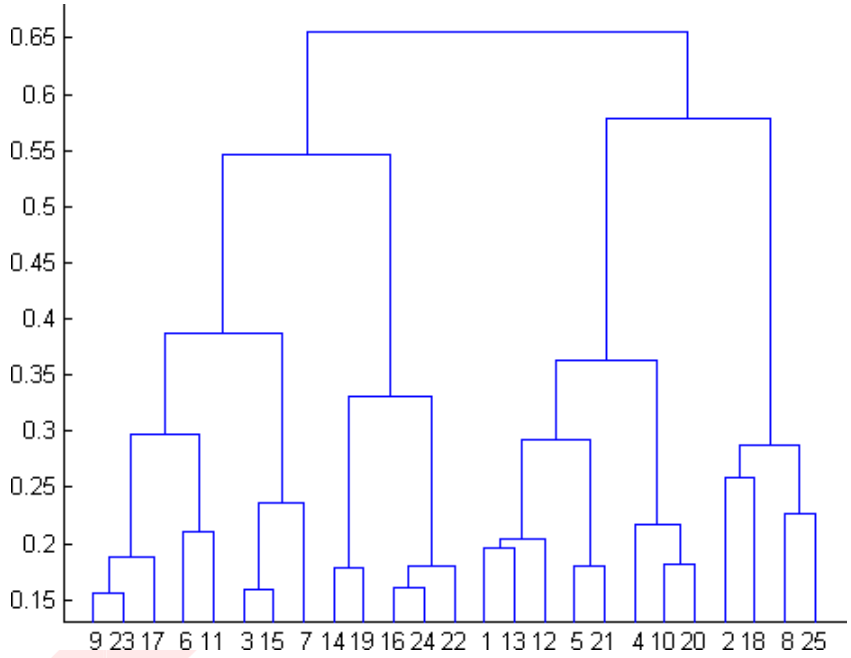
MACHINE LEARNING

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:



a) 8

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:

   a) 2 and 4

3. The most important part of____is selecting the variables on which clustering is based.
   a) interpreting and profiling clusters

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.

   a) Divisive clustering

6. Which of the following is required by K-means clustering?
   a) All answers are correct
7. The goal of clustering is to-
   a) All of the above
8. Clustering is a-

   a) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?
   a) K- Means clustering
10. Which version of the clustering algorithm is most sensitive to outliers?
    a) K-means clustering algorithm
    b)
11. Which of the following is a bad characteristic of a dataset for clustering analysis-
    a) All of the above
12. For clustering, we do not require-
    a) Labeled data

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

13. How is cluster analysis calculated?
14. How is cluster quality measured?
15. What is cluster analysis and its types?

Q13 TO Q15 Answers :-

13. This is calculated as the sum of squared distances between data points and the centers of the clusters they belong to. Inertia quantifies the within-cluster variation. Another popular metric is the silhouette coefficient, which attempts to summarize both within-cluster and between-cluster variation.

14. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

16. 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.