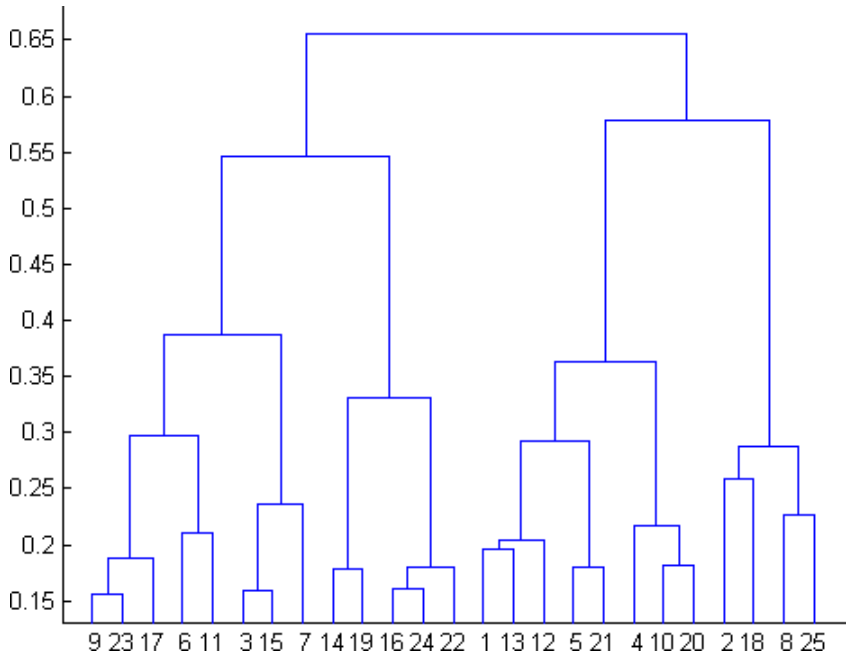
MACHINE LEARNING

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:



Answer : 2

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**

Answer:1, 2 and 4

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

Answer : formulating the clustering problem

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

Answer :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.
   Answer : Divisive clustering

6. Which of the following is required by K-means clustering?
   Answer: All answers are correct

7. The goal of clustering is to-
   Answer: Divide the data points into groups

8. Clustering is a-
   Answer : Unsupervised learning

   Which of the following clustering algorithms suffers from the problem of convergence at local optima?
   Answer :All of the above

9. Which version of the clustering algorithm is most sensitive to outliers?
   Answer: K-means clustering algorithm

10. Which of the following is a bad characteristic of a dataset for clustering analysis-
    Answer: All of the above

11. For clustering, we do not require-
    Answer:Labeled data

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

12. How is cluster analysis calculated?

Answer: It's calculated by **adding up the absolute value of the differences of the corresponding variables**, and is less likely to be influenced by a very large difference between just one of the variables

13. How is cluster quality measured?
    Answer: To measure the quality of a clustering, we can **use the average silhouette coefficient value of all objects in the data set**.

14. What is cluster analysis and its types?

Answer: Cluster Analysis is the way of organizing the data points with similar characteristics/features in one group so that they differ from the other data points of the other clusters.

Types:

1. Hierarchical Analysis
2. Centroid based
3. Distribution based
4. Density based