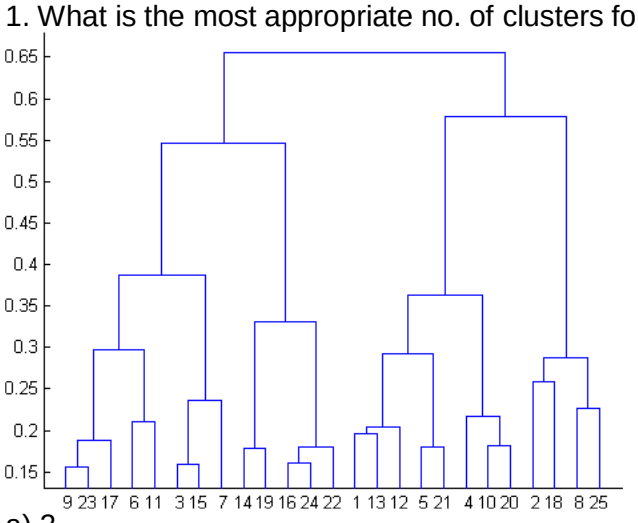
MACHINE LEARNING

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



a) 2
b) 4
c) 6
d) 8

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:
a) 1 and 2
b) 2 and 3
c) 2 and 4
d) 1, 2 and 4

d) 1, 2 and 4

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

a) interpreting and profiling clusters
b) selecting a clustering procedure
c) assessing the validity of clustering
d) formulating the clustering problem

d) formulating the clustering problem

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

a) Euclidean distance
b) city-block distance
c) Chebyshev's distance
d) Manhattan distance

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.
a) Non-hierarchical clustering
b) Divisive clustering
c) Agglomerative clustering
d) K-means clustering

b) Divisive clustering

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

a) Defined distance metric
b) Number of clusters
c) Initial guess as to cluster centroids
d) All answers are correct

d) All answers are correct

7. The goal of clustering is to-
a) Divide the data points into groups
b) Classify the data point into different classes
c) Predict the output values of input data points
d) All of the above

a) Divide the data points into groups

8. Clustering is a-
a) Supervised learning
b) Unsupervised learning
c) Reinforcement learning
d) None

Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?
a) K- Means clustering
b) Hierarchical clustering
c) Diverse clustering
d) All of the above

a) K- Means clustering

10. Which version of the clustering algorithm is most sensitive to outliers?
a) K-means clustering algorithm
b) K-modes clustering algorithm
c) K-medians clustering algorithm
d) None

a) K-means clustering algorithm
as its centroid based

11. Which of the following is a bad characteristic of a dataset for clustering analysis-
a) Data points with outliers
b) Data points with different densities
c) Data points with non-convex shapes
d) All of the above

d) All of the above

12. For clustering, we do not require-
a) Labeled data
b) Unlabeled data
c) Numerical data
d) Categorical data

a) Labeled data

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

13. How is cluster analysis calculated?

Cluster analysis known as classification – it divides the data in the segment for the analysis – makes different group make clusters for analysis.
as we have 3 cluster analysis or we can say unsupervised method – given below

1. *k*-means clustering
2. Hierarchical clustering
3. Diverse clustering

14. How is cluster quality measured?

1. Dissimilarity/Similarity metric
2. Cluster completeness
3. Ragbag
4. Small cluster preservation

15. What is cluster analysis and its types?

The process of making a group of abstract objects into classes of similar objects is known as clustering.

Clustering Methods:

1. Model-Based Method
2. Hierarchical Method
3. Constraint-Based Method
4. Grid-Based Method
5. Partitioning Method
6. Density-Based Method