

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:

a) 1 and 2

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

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?

a) K- Means clustering

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

a) Labelled data

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

13. How is cluster analysis calculated?

A. Cluster analysis calculation steps:

1. Calculate distances
2. Links the clusters
3. Choose a solution by selecting the right number of clusters

14. How is cluster quality measured?

A. Cluster analysis is calculated using the silhouette score.

15. What is cluster analysis and its types?

A. Cluster Analysis is a statistical method of processing data. It sorts data in groups or clusters on the basis of their properties. Data with similar properties are grouped together.

Types of cluster analysis:

1. Hierarchical Clustering:

- Agglomerative method: Single objects are grouped into clusters.
- Divisive Method: A single cluster is divided into smaller clusters.

2. Centroid Based Clustering: Clusters are formed using centroids or a central data point

Data or object similar to the centroid becomes a part of that cluster

K-means technique is used for this type of clustering

3. Density Based Clustering:

Here Clusters are identified by the area of density that are higher than the remaining dataset.

Here each data point must have minimum data points around it to be a part of the cluster.

Data points which do not become a part of a cluster are called noise similar to outliers.