

MACHINE LEARNING

Submitted by DARSHIK A S

- 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

- 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
 - a) 1 and 2
 - b) 2 and 3
 - c) 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**

- 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

- 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

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**

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

8) Clustering is a

- a) Supervised learning
- b) Unsupervised learning**
- c) Reinforcement learning
- d) None

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

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

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

12) For clustering, we do not require

- a) Labeled data
- b) Unlabeled data
- c) Numerical data
- d) Categorical data

13) How is cluster analysis calculated?

There are several techniques to calculate cluster analysis:

1. K-Means: It is a technique which groups elements into a number of clusters specified by the user. These clusters are represented by their centroids.
2. Agglomerative Hierarchical Clustering: It groups elements into clusters based on their similarity. Initially it considers all elements as separate clusters and pairs of clusters are successively merged until all clusters have been merged into one single cluster.
3. DBSCAN: It is a density based clustering algorithm, it groups dense regions of data in space into a single cluster.

14) How is cluster quality measured?

There are several methods to assess the quality of clusters based on the data that was clustered itself. Some of them are:

1. Silhouette Coefficient: It compares the average distance to elements in the same cluster with the average distance to elements in other clusters.
2. Dunn Index: It is the ratio between the minimal inter cluster distance to maximal intra cluster distance. It aims to identify dense and well separated clusters.

15) What is cluster analysis and its types?

Cluster analysis is an approach to group or cluster objects based on similarities. The similarity between the objects belonging to the same group is maximal and minimal between other groups.

Types of clustering:

1. Hierarchical clustering
2. Centroid based clustering
3. Distribution based clustering
4. Grid based clustering