

# **Machine Learning**

Q1- What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

**Answer-b) 4**

Q2- In which of the following cases will K-Means clustering fail to give good results?

**Answer- d) 1, 2 and 4**

Q3- The most important part of is selecting the variables on which clustering is based.

**Answer-a) interpreting and profiling clusters**

Q4- The most commonly used measure of similarity is the or its square.

**Answer-a) Euclidean distance**

Q5- \_\_\_\_\_ 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- a)Divisive clustering**

Q6- Which of the following is required by K-means clustering

**Answer- ) All answers are correct**

Q7- The goal of clustering is to

**Anwer- d) All of the above**

**Q8- Clustering is a**

**Answer- Unsupervised learning**

**Q9- Which of the following clustering algorithms suffers from the problem of convergence at local optima?**

**Answer-a) K- Means clustering**

**Q10- Which version of the clustering algorithm is most sensitive to outliers?**

**Answer- a) K-means clustering algorithm**

**Q11- Which of the following is a bad characteristic of a dataset for clustering analysis?**

**Answer-d) All of the above**

Q12- For clustering, we do not require

**Answer- a) Labeled data**

**Q13- How is cluster analysis calculated?**

Answer-It can be shown that there is no absolute “best” criterion for cluster validation. Consequently, it is the user who must supply the criterion for validating the cluster.

Q14- 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.

Q15- What is cluster analysis and its Types?

Answer- Cluster analysis is an exploratory analysis that tries to identify structures within the data. Cluster analysis is also called segmentation analysis. More specifically, it tries to identify homogenous groups of cases if the grouping is not previously known. Because it is exploratory, it does not make any distinction between dependent and independent variables.

The clustering approaches can be broadly divided into two categories: **Agglomerative** and **Divisive**.

## ***Types of Cluster Analysis-***

**1)K-Means Clustering-** K-Means is a clustering approach in which the data is grouped into K distinct non-overlapping clusters based on their distances from the K centers. The value of **K** needs to be specified first and then the algorithm assigns the points to exactly one cluster. This works on Divisive (Top-Down) Approach.

### **2) Hierarchical clustering-**

Here, we are going to discuss the bottom-up (or Agglomerative) approach of cluster building. We start by defining any sort of similarity between the data points. Generally, we consider the Euclidean distance. The points which are closer to each are more similar than the points which are farther away. The Algorithm starts with considering all points as separate clusters and then grouping points together to form clusters.

### **3) DBSCAN(Density Based Spatial Clustering of Applications with Noise)**

it is an unsupervised machine learning algorithm. This algorithm defines clusters as continuous regions of high density.