Machine learning

(1) (b) 4
(2) (d) 1,2 and 4
(3) (d) Formulating the clustering problem
(4) (a) Euclidean distance
(5) (b) Divisive clustering
(6) (d) All answers are correct
(7) (a) Divide the data points into groups
(8) (b) Unsupervised learning
(9) (d) All of the above
(10) (a) K means clustering algorithm
(11) (d) All of the above
(12) (a) Labeled data

(13) How is cluster analysis calculated?

Ans - The hierarchical cluster analysis follows three basic steps:

1) Calculate the distances,
2) Link the clusters, and
3) Choose a solution by selecting the right number of clusters.

(14) How is cluster quality measured?

Ans.- To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

The quality of a clustering result depends on both the similarity measure used by the method and its implementation. The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns. There is a separate "quality" function that measures the "goodness" of a cluster.

(15) What is cluster analysis and its types?

Ans - Cluster analysis is a multivariate data mining technique whose goal is to groups objects (products, respondents, or other entities) based on a set of user selected characteristics or attributes. It is the basic and most important step of data mining and a common technique for statistical data analysis and it is used in many fields such as data compression, machine learning, pattern recognition, information retrieval etc.

Hierarchical Cluster Analysis

In this method, first, a cluster is made and then added to another cluster (the most similar and closest one) to form one single cluster. This process is repeated until all subjects are in one cluster. This particular method is known as **Agglomerative method**

The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.

Centroid-based Clustering

In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set. K-Means method of clustering is used in this method, where K are the cluster centers and objects are assigned to the nearest cluster centers.

Distribution-based Clustering

It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster .This type of clustering can capture some complex properties of objects like correlation and dependence between attributes.