MACHINE LEARNING

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

Answer) 4

Explanation: The decision of the no. of clusters that can best depict different groups can be chosen by observing the dendrogram. The best choice of the no. of clusters is the no. of vertical lines in the dendrogram cut by a horizontal line that can transverse the maximum distance vertically without intersecting a cluster.

Question) 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

Answer) 1, 2 and 4

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

Answer) formulating the clustering problem

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

Answer) Euclidean Distance

Question) ____ 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) Divisive clustering

Question) Which of the following is required by K-means clustering ?

Answer) a) Defined distance metric

b) Number of clusters

c) Initial guess as to cluster centroids

Question) The goal of clustering is to –

Answer)) Divide the data points into groups

Question) Clustering is a –

Answer) Unsupervised learning

Question) Which of the following clustering algorithms suffers from the problem of convergence at local optima ?

Answer) a) K- Means clustering

b) Hierarchical clustering

c) Diverse clustering

Question) Which version of the clustering algorithm is most sensitive to outliers ?

Answer) K-means clustering algorithm

Question) Which of the following is a bad characteristic of a dataset for clustering analysis-

Answer) a) Data points with outliers

b) Data points with different densities

c) Data points with non-convex shapes

Question) For clustering, we do not require –

Answer) Labeled data

Question) How is cluster analysis calculated ?

Answer) 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.

Question) How is cluster quality measured ?

Answer) To measure a cluster's fitness within a clustering, we can compute the average silhouette coefficient value of all objects in the cluster. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

Question) What is cluster analysis and its types ?

Answer) Cluster analysis is the task of grouping a set of data points in such a way that they can be characterized by their relevance to one another. These types are Centroid Clustering, Density Clustering Distribution Clustering, and Connectivity Clustering.