Machine Learning_ Assignment_1

QUES 1: What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

Answer: B) 4.

QUES 2: In which of the following cases will K-Means clustering fail to give good results?

Answer: D) 1,2 and 4.

QUES 3: The most important part of_____ is selecting the variables on which clustering is based.

Answer: A): Interpreting and Profiling clusters

QUES 4: The most commonly used measure of similarity is the _____or its square.

Answer: A) Euclidean distance.

QUES 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.

ANSWER: B: Divisive clustering.

QUES 6: Which of the following is required by K-means clustering?

Answer: D) All answers are correct.

QUES 7: The goal of clustering is to-

Answer: A) Divide the data points into groups.

QUES 8. Clustering is a—

Answer: B) Unsupervised learning.

QUES 9: Which of the following clustering algorithms suffers from the problem of convergence at local optima?

Answer: D) All of the above.

Ques 10: Which version of the clustering algorithm is most sensitive to outliers?

Answer: A) K-means clustering algorithm

QUES 11: Which of the following is a bad characteristic of a dataset for clustering analysis?

Answer: D) All of the above.

Ques 12: For clustering, we do not require—

Answer: A) Labeled data.

Ques 13: 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. First, we have to select the variables upon which we base our clusters.

Ques 14: . 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.

Ques 15: What is cluster analysis and its types?

Answer: Cluster Analysis is the process to find similar groups of objects in order to form clusters. It is an unsupervised machine learning-based algorithm that acts on unlabeled data. A group of data points would comprise together to form a cluster in which all the objects would belong to the same group.

Types of Cluster Analysis are:

a) Centroid Based Clustering.
b) Density Based Clustering
c) Distribution Based Clustering
d) Hierarchical Based Clustering.