MACHINE LEARNING ASSIGNMNENT

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

ANSWER: B] 4

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

ANSWER:D] 1,2,4

3. The most important part of _is selecting the variables on which clustering is based.

ANSWER: D] formulating the clustering problem

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

ANSWER: A] EUCLIDEAN 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.
ANSWER: C] AGGLOMERATIVE CLUSTERING

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

ANSWER: D] ALL ANSWERS ARE CORRECT

7. The goal of clustering is to-
ANSWER: D] ALL OF THE ABOVE

8. Clustering is a-

ANSWER: B] UNSUPERVISED LEARNING

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

ANSWER: A] K MEANS CLUSTERING

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

ANSWER: A] K-MEANS CLUSTERING ALGORITHM

11. Which of the following is a bad characteristic of a dataset for clustering analysis

ANSWER: D] ALL OF THE ABOVE

12. For clustering, we do not require

ANSWER: A] LABELED DATA

13. How is cluster analysis calculated?

ANSWER: The cluster analysis is a exploratory method to identify structures in data. It is a Method which finds homogenous groups and makes cluster not letting all the data mixed. The cluster analysis can be done by:
1) k-means cluster
2) hierarchical cluster

K- mean Clustering: kmean clustering is a method to quickly cluster large data sets, the Number of clusters is to be predefined, this method is useful to test different models with Different number of clusters

Hierarchical clustering: hierarchical clustering is a method which generates a series of Clusters from 1 to n for each case as an individual cluster. It also works for variables Opposed to cases. It can handle nominal, ordinal and scale data .

14. How is cluster quality measured?

ANSWER:
The Silhouette Index measure the distance between each data point, the centroid of the cluster it was assigned to and the closest centroid belonging to another cluster. when the silhouette quotient value is near 1 the cluster is compact and far from other clusters.

15. What is cluster analysis and its types
ANSWER: The cluster analysis is a exploratory method to identify structures in data. It is a Method which finds homogenous groups and makes cluster not letting all the data mixed. The cluster analysis can be done by:
1) k-means cluster
2) hierarchical cluster

K- mean Clustering: kmean clustering is a method to quickly cluster large data sets, the Number of clusters is to be predefined, this method is useful to test different models with Different number of clusters

Hierarchical clustering: hierarchical clustering is a method which generates a series of Clusters from 1 to n for each case as an individual cluster. It also works for variables Opposed to cases. It can handle nominal, ordinal and scale data