MACHINE LEARNING

ASSIGNMENT – 1

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

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

b) 4

2. 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
Options:

d) 1, 2 and 4

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

d) formulating the clustering problem

4. The most commonly used measure of similarity is the or its square.
a) Euclidean distance

MACHINE LEARNING

ASSIGNMENT – 1

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.

b) Divisive clustering

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

d) All answers are correct

7. The goal of clustering is to
a) Divide the data points into groups

8. Clustering is a
b) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local
optima?
a) K- Means clustering

10. Which version of the clustering algorithm is most sensitive to outliers?
a) K-means clustering algorith

11. Which of the following is a bad characteristic of a dataset for clustering analysis
d) All of the above

12. For clustering, we do not require
a) Labeled data

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.

13. How is cluster analysis calculated?
Ans:-Calculate the distances ,Link the cluster choose a solution by selecting the right number of cluster .first we have to select
the variable upon which we base our 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

15. What is cluster analysis and its types
Ans:- Cluster analysis is a multivariate data mining technique whose goal is to groups object (eg:-Products ,respondent or other entities)
based on a set of used selected characteristics or attributes.

Type of cluster Analysis :-
1.Hierarchical cluster
2.Centroid based clustering
3.Distribution based clustering
4.Density based clustering