Machine Learning Worksheet

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

Ans : 6

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

Ans : 1,2 and 4

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

Ans : formulating the clustering problem

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

Ans : Euclidean

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

Ans : Divisive clustering

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

Ans : Defined distance metric,Number of clusters,Initial guess as to cluster centroids

7)The goal of clustering is

Ans : Divide the data points into groups,Classify the data point into different classes,Predict the output values of input data points

8)Clustering is a-

Ans : Un supervised learning

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

Ans : k means clustering

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

Ans : k means clustering algorithm

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

Ans : Data points with different densities

12)For clustering, we do not require

Ans :Labeled data

13)How is cluster analysis calculated?

Ans:

How is cluster quality measured?

What is cluster analysis and its types