MACHINE LEARNING
ASSIGNMENT – 1

1. (b)

2. (d) 1, 2 and 4

3. (d) Formulating the cluster program

4. (a) Euclidean Distance

5. (b) Divisive Clustering

6. (d) All Answers are correct

7. (a) Divide the data points into groups

8. (b) Unsupervised Learning

9. (d) All of the above (K-means clustering, Hierarchical, Diverse)

10. (a) K-means Clustering Algorithm

11. (d) All

12. (a) Labeled Data

13. **CLUSTER ANALYSIS CALCULATED:**

We must select the variables upon which we base our clusters.

The cluster analysis follows three basic steps-

(a) Calculate the Euclidean Distance between observed value and centroid value.

(b) Link the adjacent clusters.

(c) Finalize the calculation by selecting right number of clusters.

14. Silhouette Coefficient or silhouette score is a metric used to calculate the quality of a clustering technique. Its value ranges from -1 to 1. Whereas 1: Means clusters are well apart from each other and clearly distinguished.

15. Cluster analysis is a data analysis technique that explores the naturally occurring groups within a data set known as clusters.

Types of Clustering:
a) **Centroid Clustering:** organizes the data into non-hierarchical clusters.
b) **Density Clustering:** unsupervised machine learning methods that identify distinctive clusters in the data.
c) **Distribution Clustering:** directly relates to the use of distribution models in statistics
d) **Hierarchical Clustering:** method of cluster analysis that seeks to build a hierarchy of clusters.