MACHINE LEARNING ASSIGNMENT 1

1. b. 4

2. d. 1,2 and 4

3. d. formulating the clustering problem

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

10. a. K-means clustering algorithm

11. d. All of the above

12. a. Unlabelled data

13. Cluster analysis is calculated by using a variety of algorithms to group data points into clusters based on their similarity. The algorithms used to calculate cluster analysis vary depending on the type of data being analysed and the desired outcome. Common algorithms used for cluster analysis include k-means clustering, hierarchical clustering, and density-based clustering.

14. Cluster quality is typically measured using a metric called the silhouette coefficient. This metric measures how closely related the data points in a cluster are to each other, and how distinct they are from the data points in

other clusters. The silhouette coefficient ranges from -1 to 1, with higher values indicating better cluster quality.

15. Cluster analysis or clustering, is a method of data mining that groups similar data points together. The goal of cluster analysis is to divide a dataset into groups (or clusters) such that the data points within each group are more similar to each other than to data points in other groups. It is an unsupervised machine learning-based algorithm that acts on unlabelled data.

In simple words, A cluster is nothing but a collection of similar data which is grouped together.

For example, a streaming service may collect the following data about individual:

- Minutes watched per day

- Total viewing session per week

- Number of unique shows viewed per month

Using these metrics, a streaming service can perform cluster analysis to identify high usages and low usage user so that they can know who they should spend most of their advertising money on.

Types of Clustering Methods

1. Partitioning Clustering: It is a type of clustering that divided the data into non-hierarchical group. It is also known as the centroid-based method. The most common example of partitioning clustering is the K-Mean Clustering algorithm.

2. Density Based clustering: This method connects the highly-dense areas into cluster, and arbitrarily shaped distributions are formed as long as the dense

region can be connected. This algorithm does it by identifying different clusters in the dataset and connects the areas of high densities into clusters.

3. Distribution Model-Based Clustering: In this method, the data is divided based on the probability of how a dataset belongs to a particular distribution.

4. Hierarchical Clustering: In this technique, the dataset is divided into clusters to create a tree-like structure, which is also called a dendrogram.