MACHINE LEARNING ASSIGNMENT 3)

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

1) Which of the following is an application of clustering?

   Ans-**a. Biological network analysis**

2) 2. On which data type, we cannot perform cluster analysis?

   Ans- **d. None**

3) Netflix's movie recommendation system uses

   Ans c. Reinforcement learning and Unsupervised learning

4) . The final output of Hierarchical clustering is

   Ans-.a The number of cluster centroids

5) Which of the step is not required for K-means clustering?

   Ans a. A distance metric

6) Which is the following is wrong?

   Ans- c. k-nearest neighbour is same as k-means

7) Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

   Ans- d. 1, 2 and 3

8) Which of the following are true?

   Ans- a. 1 only

9) In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?

   Ans a. 2

10) For which of the following tasks might clustering be a suitable approach?

   Ans b) Given a database of information about your users, automatically group them into different market segment

11) . Given, six points with the following attributes:

Ans- a.

12) Given, six points with the following attributes: Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

Ans-b

13) What is the importance of clustering?

Ans:

Clustering methods (like Hierarchical method, Partitioning, Density-based method, Model-based clustering, and Grid-based model) help in grouping the data points into clusters, using the different techniques are used to pick the appropriate result for the problem, these clustering techniques helps in grouping the data points into similar categories, and each of these subcategories is further divided into subcategories to assist the exploration of the queries output.

Clusters are geographic concentrations of interconnected companies or institutions that manufacture products or deliver services to a particular field or industry. Clusters typically include companies in the same industry or technology area that share infrastructure, suppliers, and distribution networks. Supporting firms that provide components, support services, and raw materials come together with likeminded firms in related industries to develop joint solutions and combine resources to take advantage of market opportunities. These are groups of related businesses and organizations—sometimes direct competitors, but more often operating in a complementary manner. They may comprise more than just one industry classification, and a true cluster is more than just a supplier producer-buyer model.

14. How can I improve my clustering performance?

Ans –

Feature selection refers to algorithms that select a subset of the input feature set that are most important to a given problem. On the other hand, methods that create new features based on the transformation or combination of original features are termed feature extraction. Incorporating feature selection and extraction into a clustering or classification system is one of the important steps in designing and building these systems. This method often carries several advantages. In this research, we apply feature selection and extraction algorithms to the data before applying our clustering algorithm (CCA) to improve its efficiency as well as to cut down its execution cost