**MACHINE LEARNING**

**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?

d. All of the above

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

d. None

3. Netflix’s movie recommendation system uses-

c. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is-

b. The tree representing how close the data points are to each other

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

d. None

6. Which is the following is wrong?

c. k-nearest neighbour is same as k-mean

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

d. 1, 2 and 3

8. Which of the following are true?

i. Clustering analysis is negatively affected by multicollinearity of features

ii. Clustering analysis is negatively affected by heteroscedasticity

1. 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?

b. 4

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

b. Given a database of information about your users, automatically group them into different market segments.

11. Given, six points with the following attributes: Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

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.

b

**Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly**

13. What is the importance of clustering?

**Clustering** is useful for exploring data. If there are many cases and no obvious groupings, **clustering** algorithms can be used to find natural groupings.

**Clustering** can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models.

14. How can I improve my clustering performance?

K-means **clustering** algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) **the** algorithm. When **the** data has overlapping **clusters**, k-means can **improve the** results of **the** initialization technique.

**Clustering** segment the data into a similar group instead of prediction , then you **can** build a predictive model for each group.It helps in finding the pattern within the data. ... This works best with large data set.