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?

Ans- All of the above

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

Ans- None

3. Netflix's movie recommendation system uses-

Ans- Reinforcement learning and unsupervised learning.

4. The final output of hierarchical clustering is-

Ans- The tree representing how close the data point is to each other.

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

Ans- None

6. Which of the following metrics, do we have for finding dissimilarity between hierarchical clustering?

Ans- k-nearest neighbor is same as k-means

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

Ans- i. Single-link ii. Complete-link iii. Average-link

d.1,2 and 3

8. Which of the following are true?

Ans- 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. Which of the following task might clustering be a suitable approach?

Ans- B. given a database of information about your user, automatically group them in different market segment.

11. Given six points with the following attribute: which of the following clustering representation and dendrogram depicts the use of MIN or single link proximity in hierarchical clustering:

Ans- a

12. Given , six point with the following attributes: which of the following clustering representation and dendrogram depicts the use of MAX or complete link proximity function in hierarchical clustering.

Ans- b

Q13 to Q14 are subjective answer type question. Answer them in their own words briefly.

13. What is the importance of clustering?

Ans- The purpose of clustering and classification algorithm is to make sense of and extract value from large sets of structured and unstructured data. If you're working with huge volumes of unstructured data, it may only make sense to try to partition the data into some sort of logical groupings before attempting to analyze it.

Clustering and classification allow you to take a sweeping glace of your data and then form some logical structures based on what you find there before going deeper into the nuts-and-bolts analysis.

In their simplest form, *clusters* are sets of data points that share similar attributes, and *clustering algorithms* are the method that group these data points into different clusters based on their similarities. But you'll see clustering algorithms

used for disease classification in medical science, but you'll also see them used for customer classification in marketing research and for environment health risk assessment in environment engineering.

14. How can I improve my clustering performance?

Ans- 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 well separated clusters, the performance of k-means depends completely on the goodness of the initialization using simple furthest point heuristic(Maxmin) reduces the clustering error of k-means from 15%to 6% on average.