MACHINE LEARNING – WORKSHEET (CLUSTERING)

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=d) All of the above

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

Ans=a) Time series data

1. Netflix’s movie recommendation system uses

Ans=a) Supervised learning

1. The final output of Hierarchical clustering is\_

Ans=d) All of the above

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

Ans=c) initial guess as to cluster centroids

1. Which is the following is wrong?

Ans=b) k-means clustering tries to group n observations into k clusters

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

Ans=d) 1, 2 and 3

1. Which of the following are true?

Ans=a) 1 only

1. 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=b) 4

1. 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 segments.

1. 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:

Ans=)A

1. 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

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly

1. What is the importance of clustering?

Ans=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.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

1. How do you cluster a profile?

Ans=Profiling involves generating descriptions of the clusters with reference to the input variables you used for the cluster analysis. Profiling acts as a class descriptor for the clusters and will help you to 'tell a story' so that you can understand this information and use it across your business.

1. 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 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.