MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering
   iii) Regression
   Options:

   ANS: 1 and 2 (Classification and Clustering)

2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification
   iii) Clustering
   iv) Reinforcement
   Options:

   a) 1 Only

3. Can decision trees be used for performing clustering?
   a) True

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
   i) Capping and flooring of variables
   ii) Removal of outliersOptions:
   a) 1 only

5. What is the minimum no. of variables/ features required to perform clustering?

   b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
   b) No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes

8. Which of the following can act as possible termination conditions in K-Means?
   i) For a fixed number of iterations.
   ii) Assignment of observations to clusters does not change between iterations. Except for cases witha bad local minimum.
   iii) Centroids do not change between successive iterations.
   a) Terminate when RSS falls below a threshold.
   Options:
   d) All of the above

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm

MACHINE LEARNING

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

    d) All of the above

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
    d)All of the above

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?

Yes, the K-means clustering algorithm is sensitive to outliers because extreme numbers can readily change the mean, K-medoids clustering is a K-means variation that is more resistant to noise and outliers. It indicate the centre of a cluster by a real point within the cluster rather than by the mean point. Medoid is the object in the cluster with the fewest total distances to other points and the most central location.

13. Why is K means better?

The benefits of k-means:
(i) Implementing it is not too difficult.
(ii) It scales to enormous data sets.
(iii) Convergence is ensured by it.
(iv) can reset the placements of centroids.
(V) It easily adjusts to new examples.
(Vi) It generalises to clusters of various sizes and shapes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

The basic *k*-means clustering is a non-deterministic algorithm. This implies that using the algorithm on the same data multiple times could have various outcomes. However, to ensure consistent results, FCS Express performs *k*-means clustering using a deterministic method.