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:
   Answer : Clustering

2. Sentiment Analysis is an example of:
   Regression / Classification / Reinforcement
   Answer :1, 2 and 4

3. Can decision trees be used for performing clustering?
   Answer : 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:
   Answer : Capping and flooring of variables

5. What is the minimum no. of variables/ features required to perform clustering?
   Answer :1

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

7. Is it possible that Assignment of observations to clusters does not change between successiveiterations in K-Means?
   Answer : Yes

8. Which of the following can act as possible termination conditions in K-Means?
   1. For a fixed number of iterations.
   2. Assignment of observations to clusters does not change between iterations. Except for caseswitha bad local minimum.
   3. Centroids do not change between successive iterations.
   4. Terminate when RSS falls below a threshold.
   Answers : All of the above

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

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regressionmodel (Supervised Learning):
    Creating different models for different cluster groups.
    Creating an input feature for cluster ids as an ordinal variable.
    Creating an input feature for cluster centroids as a continuous variable.
    Creating an input feature for cluster size as a continuousvariable.Options:
    Answer : 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?
    a) Proximity function used
    b) of data points used
    c) of variables used
    Answer : All of the above

12. Is K sensitive to outliers?

    Answer : K-means is sensitive to outliers because it uses the mean of the points in a cluster as its centroid. Outliers can greatly affect the mean of a cluster and therefore change the position of the centroid, which can cause points to be assigned to different clusters. This can lead to clusters that are not as compact or tight as they could be, and also increase the overall error of the algorithm.

13. Why is K means better?

    Answer: K-means is a popular clustering algorithm due to its simplicity, efficiency, and scalability. Some of the reasons why K-means is considered to be a better algorithm include:

    Simplicity, Efficiency, Scalability, Clusters with similar characteristics, Good with spherical clusters

14. Is K means a deterministic algorithm?
    Answer : K-means is a deterministic algorithm, which means that it will always produce the same results given the same input data and initial conditions. The algorithm starts by randomly initializing the centroids of the clusters, and then iteratively reassigns points to clusters and recalculates the centroid until convergence. The final cluster assignments and centroid locations will be the same every time the algorithm is run with the same input data and initial conditions.