MACHINE LEARNING

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering
   iii) Regression
      Options:
   a) 2 Only
   b) 1 and 2
   c) 1 and 3
   d) 2 and 3

Ans - a) 2 Only

2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification
   iii) Clustering
   iv) Reinforcement
      Options:
   a) 1 Only
   b) 1 and 2
   c) 1 and 3
   d) 1, 2 and 4

Ans - d) 1, 2 and 4

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

Ans – 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 outliers
      Options:
   a) 1 only
   b) 2 only
   c) 1 and 2
   d) None of the above

Ans – a) 1 only

5. What is the minimum no. of variables/ features required to perform clustering?
   a) 0
   b) 1
   c) 2
   d) 3

Ans – b) 1

d) v

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

Ans – b) No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes
   b) No
   c) Can't say
   d) None of these

Ans – 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.
   iv) Terminate when RSS falls below a threshold.
   Options:
   a) 1, 3 and 4
   b) 1, 2 and 3
   c) 1, 2 and 4
   d) All of the above

Ans – d) All of the above

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm
   b) K-medians clustering algorithm
   c) K-modes clustering algorithm
   d) K-medoids clustering algorithm

Ans – a) K-means clustering

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
   i)   Creating different models for different cluster groups.
   ii) Creating an input feature for cluster ids as an ordinal variable.
   iii) Creating an input feature for cluster centroids as a continuous variable.
   iv) Creating an input feature for cluster size as a continuous variable.
      Options:
   a) 1 only
   b) 2 only
   c) 3 and 4
   d) All of the above

Ans – 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?
   a) Proximity function used
   b) of data points used
   c) of variables used
   d) All of the above

Ans – d) of data points used

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

12. Is K sensitive to outliers?

Yes, K is sensitive to outliers

The k-means algorithm updates the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the points are packed nicely together, the average makes sense. However, when you have outliers, this can affect the average calculation of the whole cluster. As a result, this will push your cluster center closer to the outlier.

13. Why is K means better?

The meat of the K-means clustering algorithm is just two steps, the cluster assignment step and the move centroid step. If we're looking for an unsupervised learning algorithm that is easy to implement and can handle large data sets, K-means clustering is a good starting point.

Based on my experience using K-means clustering, the algorithm does its work quickly, even for really big data sets.

14. Is K means a deterministic algorithm?

What is understand K-Means is its non-deterministic nature. This means that running the algorithm several times on the same data, could give different results. We propose an improved, density based version of K-means,which involves a novel and symentric mentod for selecting initial centroids.

The proposed algorithm is compared with classical K-means with random initialization and improved K-means variants such as K-means++ and MinMax algorithms. It is also compared with three deterministic initialization methods.