MACHINE LEARNING

1. Movie Recommendation systems are an example of:

   Answer: b) 1 and 2 – Classification and Clustering

2. Sentiment Analysis is an example of:

   Answer: d) 1,2 and 4 – Regression, Classification and Reinforcement

3. Can decision trees be used for performing clustering?

   Answer: 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:

   Answer: 1 only - Capping and flooring of variables

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

   Answer: b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?

   Answer: b) No

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

   Answer: Yes

8. Which of the following can act as possible termination conditions in K-Means?

   Answer: d) All of the above

9. Which of the following algorithms is most sensitive to outliers?

   Answer: a) K-means clustering algorithm

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

Answer: 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?

Answer: 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

Answer: 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?

Answer: . K- means clustering is a data segmentation technique. It divides the data into parts/clusters where each cluster tries to have data points which are very close to each other. i.e minimum variance in the cluster . That's why K means is better.

14. Is K means a deterministic algorithm?

Answer: K means is not a deterministic algorithm . It is one of the most significant drawback of K means. K means starts with a random set of data points as initial centroids . This random selection influences the quality of the resulting cluster .