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:

   c) 1 and 3

2. Sentiment Analysis is an example of:

   d) 1, 2 and 4

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:

   c) 1 and 2

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?

   d) All of the above

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

   a) K-means clustering algorithm

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 Means sensitive to outliers?

K means is outlier sensitive, because when outlier presents in the data set the mean will be affected and it will be resulted in the poor performance in outcome. While creating the group model will fail to group the closest data point as 1 cluster due to outlier distance .because K means always create centroid with distance.

13. Why is K means better?

- Its easy to implement and easy to understand
- It will work on the large data sets as well.
- It will best accuracy compared to other clustering algorithm
- Its easy to tune, which is we can decide number of cluster to be created.

14. Is K means a deterministic algorithm?

No, because when we run the algorithm several times on the same data set, it will give us the different outcome.