Parth Makwana (DS0722)

MACHINE LEARNING

Q1 to Q11

1. Movie Recommendation systems are an example of:
   a) 2 Only

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:
   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?
   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

12. Is K sensitive to outliers?

- Because the mean, as a statistic, is generally sensitive to outliers.
- The mean of 2,2,2,3,3,3,4,4,4 is 3
- If we add a single 23 to that, the mean becomes 5, which is larger than any of the other values.
- Since in k-means, you'll be taking the mean a lot, you wind up with a lot of outlier-sensitive calculations.
- That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers

13. Why is K means better?

- k-means is one of the simplest algorithm which uses unsupervised learning method to solve known clustering issues. It works really well with large datasets.

14. Is K means a deterministic algorithm?

- The basic k-means clustering is based on a non-deterministic algorithm.
- This means that running the algorithm several times on the same data, could give different results.