Machine Learning

1. Movie Recommendation systems are an example of:

Ans: B

2. Sentiment Analysis is an example of

Ans: D

3. Can decision trees be used for performing clustering?

Ans: A

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?

Ans: A

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

Ans: B

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

Ans: B

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

Ans: A

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

Ans: D

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

Ans: D

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

Ans: D

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

Ans: D

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

Q12. Is K Sensitive to Outliers?

Ans: As we know that Outliers Removal is very import, suppose we need to calculate the mean of followings. Case I: [2,3,4,5,6], we get mean as a 5. Case II: [2,3,4,5,6,84], we get the mean as a 34. If any maximum value gets added to the data point, then that would create a noise, Due to variability in the measurement or can cause experimental error. Here in K means algorithm the cluster centers are decided by taking the average of all the data points that are closer to the cluster center. So, model can easily be influenced by the outliers. In Such cases we can use DB SCAN which is Density Based

Q13. Why is K Means Better?

Ans: K Means is better Algorithm as K means is very easy to implement, this can be used for large data sets, By WCSS or Elbow method we can determine the number of clusters Technically. Whereas the other Agglomerative Clustering or Hierarchical Clustering and DB SCAN works without any input.

Q14. IS K MEANS a deterministic Algorithm?

Ans: The Basic K Means Clustering is based on a non-deterministic algorithm; this Means that running the algorithm several times on the data would produce different results. This is because of the Nature of the algorithm varies because of the random selection of data points in k -means as initial centroid. The Main Idea of the algorithm is to select data points which belongs to the dense regions and the data points should be closer to the centroid and their should be a gap between each clusters.