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:
b) 1 and 2

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

12. Is K sensitive to outliers?

The *K*-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values and thus, cannot represent the correct cluster centre.

13. Why is K means better?

K mean is better because-
1) If variables are huge, then K-Means most of the times computationally faster than hierarchical clustering, if we keep k smalls.
2) K-Means produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

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.
The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.