ASSIGNMENT – 2

MACHINE LEARNING SOLUTIONS

1. Movie Recommendation systems are an example of: i) Classification ii) Clustering iii) Regression Options: a) 2 Only b) 1 and 2 c) 1 and 3 d) 2 and 3

   Ans B) CLASSFICATION AND CLUSTERING

2. Sentiment Analysis is an example of: i) Regression ii) Classification iii) Clustering iv) Reinforcement Options: a) 1 Only b) 1 and 2 c) 1 and 3 d) 1, 2 and 4

   Ans D)

3. Can decision trees be used for performing clustering?
   a) True b) False

   Ans 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: i) Capping and flooring of variables ii) Removal of outliers Options: a) 1 only b) 2 only c) 1 and 2 d) None of the above

   ANS A) CAPPING AND FLOORING OF VARIABLES

5. What is the minimum no. of variables/ features required to perform clustering?
   a) 0 b) 1 c) 2 d) 3

ANS) B) ONE

6. For two runs of K-Mean clustering is it expected to get same clustering results?
   a) Yes b) No

   ANS) B) NO

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes b) No c) Can't say d) None of these

   ANS) A) YES

8. Which of the following can act as possible termination conditions in K-Means? i) For a fixed number of iterations. ii) Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum. iii) Centroids do not change between successive iterations. iv) Terminate when RSS falls below a threshold.
   Options: a) 1, 3 and 4 b) 1, 2 and 3 c) 1, 2 and 4 d) All of the above

   ANS) D) ALL OF THE ABOVE

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm b) K-medians clustering algorithm c) K-modes clustering algorithm d) K-medoids clustering algorithm

   ANS) A) K-MEANS CLUSTERING

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning): i) Creating different models for different cluster groups. ii) Creating an input feature for cluster ids as an ordinal variable. iii) Creating an input feature for cluster centroids as a continuous variable. iv) Creating an input feature for cluster size as a continuous variable.
Options: a) 1 only b) 2 only c) 3 and 4 d) All of the above

ANS) **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?
a) Proximity function used b) of data points used c) of variables used d) All of the above

ANS) **D) ALL OF THE ABOVE**

SUBJECTIVE QUESTION

12. Is K sensitive to outliers?

ANS) K- is very much sensitive to the outliers as it affects the overall centroid of the clusters and tends to shift the means of the cluster to other position which may result to false predictions.

13. Why is K means better?

ANS) K Means is very fast and easy to understood unsupervised learning algorithm. Its robustness

Makes its popular among data scientist. K means also utilises support which provides following advantages.

a) It gives good results
b) It is already implemented in the software
c) Number of clusters has to be fixed before
d) Dependent of the initialisation parameters and the chosen distances.

14.Is K means a deterministic algorithm?

Ans no , k means is a non deterministic algorithm and it is its major drawback. K means take random data and hence result will be varied in every run which influences the quality of the resulting clusters