ASSIGNMENT – 1

MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:

 i) Classification
 ii) Clustering
 iii) Regression

Answer: b) 1 and 2

2. Sentiment Analysis is an example of:

 i) Regression
 ii) Classification
 iii) Clustering
 iv) Reinforcement

Answer: i) Regression

3. Can decision trees be used for performing clustering?

Answer: 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

Answer: c) 1 and 2

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

Answer: b) 1

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

Answer: b) No

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

Answer: 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.

Answer: d) All of the above

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

Answer: a) K-means clustering algorithm

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.

Answer: 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?

Answer: 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?

Answer: Yes, the K-means clustering algorithm is sensitive to outliers.

13. Why is K means better?

Answer: The k-means does cluster analysis efficiently.

14. Is K means a deterministic algorithm?

Answer: No, K means is not a deterministic algorithm