MACHINE LEARNING ASSIGNMENT – 2

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: 2 Only (A)

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: 1, 2 and 4 (D)

3. Can decision trees be used for performing clustering?

a) True

b) False

ANS: True (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:

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: 1 only (A)

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

a) 0

b) 1

c) 2

d) 3

ANS: 1 (B)

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

a) Yes

b) No

ANS: No (B)

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: Yes (A)

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 witha 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: All of the above (D)

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

i) K- Means clustering algorithm

ii) Agglomerative clustering algorithm

iii) Expectation-Maximization clustering algorithm

iv) Diverse clustering algorithm

Options:

a) 1 only

b) 2 and 3

c) 2 and 4

d) 1 and 3

ANS: 1 and 3 (D)

10. 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: K-means clustering algorithm (A)

11. 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: All of the above (D)

12. 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: All of the above (D)

13. Is K sensitive to outliers?

ANS: YES, K is sensitive to outliers because a mean is easily influenced by extreme values.

14. Why is K means better?

ANS: Other clustering algorithms with better features tend to be more expensive. In this case, k-means becomes a great solution for pre-clustering, reducing the space into disjoint smaller sub-spaces where other clustering algorithms can be applied. K-means is the simplest.Plus, most people don't need quality clusters.

15. Is K means a deterministic algorithm.

ANS: Kmeans algorithm is an iterative algorithm that tries to partition the dataset into Kpre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.