FLIP ROBO

ASSIGNMENT – 2

MACHINE LEARNING

1. Movie Recommendation systems are an example of:

i) Classification

ii) Clustering

iii) Regression

Options: b) 1 and 2

2. Sentiment Analysis is an example of:

i) Regression

ii) Classification

iii) Clustering

iv) Reinforcement

Options: d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

Options: 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:

Options: a) 1 only

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

Options: b) 1

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

Options: b) No

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

Options: 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: 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

Options: d) 1 and 3

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: 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?

Options: 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?

Yes, k-means can be quite sensitive to outliers in your data set. The reason is simply that k-means tries to optimize the sum of squares**. And thus a large deviation (such as of an outlier) gets a lot of weight.**

13. Why is K means better?

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

14. Is K means a deterministic algorithm.

K-Means is a non-deterministic algorithm. This means that a compiler cannot solve the problem in polynomial time and doesn't clearly know the next step. This is because some problems have a great degree of randomness to them. These algorithms usually have 2 steps — 1)Guessing step 2)Assignment step. On similar lines is the K-means algorithm. The K-Means algorithm divides the data space into K clusters such that the total variance of all data points with respect to the cluster mean is minimized.