Machine Learning Worksheet-2 (Internship-23, Konatala Mohit ID-34)

1. Movie Recommendation systems are an example of:

i) Classification ii) Clustering iii) Regression

a) 2 only

2.Sentiment Analysis is an example of:

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

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:

i) Capping and flooring of variables ii) Removal of outliers

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? :

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

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): 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.

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

12. Is K sensitive to outliers?

K is extremely sensitive to outliers because K is a mean value of the data and mean value is influenced by extreme values (outliers).

13. Why is K means better?

K mean is better because it is relatively simple to implement, it Scales to large data sets, guarantees convergence, can warm-start the positions of centroid, easily adapts to new examples and generalizes to clusters of different shapes and sizes, such as elliptical clusters.

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.