MACHINE LEARNING – WORKSHEET 2 (CLUSTERING)

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

1. Movie Recommendation systems are an example of:

Ans=2. Clustering Only 2.

1. Sentiment Analysis is an example of:

Ans=d) 1, 2 and 3

1. Can decision trees be used for performing clustering?

Ans=b) False

1. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

Ans=b) Removal of outliers Only 2.

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

Ans=b) 1

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

Ans=a) Yes

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

Ans=a) Yes

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

Ans=d) All of the above

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

Ans=d) 1 and 3

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

Ans=a. K-means clustering algorithm

1. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

Ans=f) All of the above

1. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

Ans=e) All of the above

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly

1. Is K sensitive to outliers?

Ans=The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers. Instead of using the mean point as the center of a cluster, K-medoids uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points. Figure [1](https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_426#Fig1_426) shows the difference between mean and medoid in a 2-D example. The group of points in the right form a cluster, while the rightmost point is an outlier. Mean is greatly influenced by the outlier and thus cannot represent the correct cluster center, while medoid is robust to the outlier and correctly represents the cluster center.

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. To implement and to run.

K-means clustering algorithm can be significantly **improved** by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improve the results of the initialization technique.

15.Is K means a deterministic algorithm?

Ans=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. However, to ensure consistent results, FCS Express performs k-means clustering using a deterministic method.