**MACHINE LEARNIG ASSIGNMENT-2**

1. Movie Recommendation systems are an example of:
2. Classification
3. Clustering
4. Regression

Answer. A) 2 only (Clustering)

1. Sentiment Analysis is an example of:
2. Regression
3. Classification
4. Clustering
5. Reinforcement

Answer. D) 1,2 and 4

1. Can decision trees be used for performing clustering?

Answer. A) True

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:
2. Capping and flooring of variables
3. Removal of outliers

Answer. A) 1 only

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

Answer. B) 1

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

Answer. B) No

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

Answer. A) Yes

1. Which of the following can act as possible termination conditions in K-Means?
2. For a fixed number of iterations.
3. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
4. Centroids do not change between successive iterations.
5. Terminate when RSS falls below a threshold.

Answer. D) All of the above

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

Answer. A) K-means clustering algorithm

1. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
2. Creating different models for different cluster groups.
3. Creating an input feature for cluster ids as an ordinal variable.
4. Creating an input feature for cluster centroids as a continuous variable.
5. Creating an input feature for cluster size as a continuous variable.

Answer. D) 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?

Answer. D) All of the above

1. Is K sensitive to outliers?

Answer. 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 centre 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.

1. Why is K means better?

Answer. K-Means is on of the simplest unsupervised learning algorithms that solves the well known clustering problem. The procedure follows an easy way to classify a given data set through a certain number of clusters (assume K clusters) fixed a priority. The main idea is to define k centres, on for each cluster. These centroids should be placed in a smart way beacause of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending. the first step is completed and an early groupage is done. At this point we need to re-calculate K new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid.

1. Is K means a deterministic algorithm?

Answer. No, K-means is non-deterministic algorithm, this means that running the algorithm several times on the same data, could give different results. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.