MACHINE LEARNING

Question) Movie Recommendation systems are an example of:

i) Classification
ii) Clustering
iii) Regression

Answer) Clustering

Question) Sentiment Analysis is an example of:

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

Answer) Regression, Classification and Reinforcement

Question) Can decision trees be used for performing clustering ?

Answer) True

Question) 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

Answer) Capping and flooring of variables

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

Answer) 1

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

Answer) No

Question) s it possible that Assignment of observations to clusters does not change between successive iterations in K-Means ?

Answer) Yes

Question) 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.

Answer) all options are possible termination conditions in K-Means

Question) 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

Answer) K-means clustering algorithm

Question) 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.

Answer) All options are correct

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

Answer) All options are correct

Question) Is K sensitive to outliers ?

Answer) The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. The group of points in the right form a cluster, while the rightmost point is an outlier.

Question) Why is K means better ?

Answer) 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.

Question) Is K means a deterministic algorithm ?

Answer) 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.