MACHINE LEARNING

Q.1- Movie recommendation systems are an example of:

Ans- a)- 2 only(classification)

Q.2- Sentiment analysis is an example of:

Ans- d)- 1,2 and 4 (regression, classification and reinforcement)

Q.3- Can decision trees be used for performing clustering?

Ans- a)- True

Q.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:

Ans – a)- 1 only

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

Ans- b)- 1 (1)

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

Ans- b)- no

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

Ans- a)- yes

Q.8- which of the following can act as possible termination conditions in K-Means?

Ans- d)- all of the above

Q.9- which of the following algorithm is most sensitive to outliers?

Ans- a)- K-Means clustering algorithm

Q.10- how can clustering (unsupervised learning) be used to improve the accuracy of linear regression model (supervised learning)?

Ans- d) all of the above

Q.11- what could be the possible reason(s) for producing two different dendrograms and agglomerative clustering algorithms for the same dataset?

Ans- d)-all of the above

Q.12 to Q 14 are subjective answers type questions, answer them in their own words briefly.

Q12- is K sensitive to outliers?

Ans- The K-Means clustering algorithm is sensitive to outliers, because a mean is easily is easily influenced by extreme values. The algorithm seeks to minimize the squared Euclidean distances between the observation and the cluster centroid to which it belongs. However, the K-Mean algorithm does not always produce the best results. It is susceptible to outliers is a data point that differs from the rest of the data points.

Q.13- what is K-means better?

Ans- convergence is guarantee with K means also centroid's positions can be warmed up. Adapts easily to new examples. Generalizes to different shapes as elliptical clusters. K means that it is easy to implement and adapt to new examples. With that we can also handle large data sets.

Q.14- 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. We can propose an improved, density-based version of k mean that includes a novel and systematic method for choosing initial centroids.