Machine Learning Set2

Q1-Movie Recommendation systems are an example of
Answer-d) 2 and 3

Q2-Sentiment Analysis is an example of:
Answer-d) 1, 2 and 4

Q3-Can decision trees be used for performing clustering?
Answer-a) True

Q4-Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
Answer-i) Capping and flooring of variables

Q5-What is the minimum no. of variables/ features required to perform clustering?
Answer-b) 1

Q6-For two runs of K-Mean clustering is it expected to get same clustering results?
Answer-b) No

Q7-Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means
Answer-a)Yes

Q8-Which of the following can act as possible termination conditions in K-Means?
Answer-d) All of the above

Q9-Which of the following algorithms is most sensitive to outliers
Answer- a) K-means clustering algorithm

Q10-How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
Answer-d) All of the above

Q11-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

Q12- Is K sensitive to outliers?
Answer-Yes, K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center.

Q13-Why is K means better?
Answer-K means is better –as

- it's Easy to Implement
- Can work on Large Datasets
- Guarantees convergence
- Easily adapts to new examples
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

Q14- Is K means a deterministic algorithm?
Answer- No, K means is non-deterministic algorithm, as it randomly select data points as initial centroids .this means that running the algorithm several times on the same data, could give different results