MACHINE LEARNING
ANSWERS

Q1. Movie Recommendation systems are an example of: i) Classification ii) Clustering iii) Regression Options

ANS: a) 2 Only

Q2. Sentiment Analysis is an example of: i) Regression ii) Classification iii) Clustering iv) Reinforcement

ANS: d) 1, 2 and 4

Q3. Can decision trees be used for performing clustering?

ANS: 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: i) Capping and flooring of variables ii) Removal of outliers

ANS: a) 1 only

Q5. What is the minimum no. of variables/ features required to perform clustering?

ANS: b) 1

Q6. For two runs of K-Mean clustering is it expected to get same clustering results?

ANS: b) No

Q7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

ANS: a) Yes

Q8. 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 witha bad local minimum. iii) Centroids do not change between successive iterations. iv) Terminate when RSS falls below a threshold.

ANS: d) All of the above

Q9. Which of the following algorithms is most sensitive to outliers?

ANS: a) K-means clustering

Q10. 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.

ANS: d) All of the above

Q11. 11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

ANS: d) All of the above

Q12. Is K sensitive to outliers?

ANS: 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.

Q13. 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.

Q14. .Is K means a deterministic algorithm?

ANS: NO, K-Means is a Non- Deterministic Algorithm. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.