

Machine Learning Assignment 2

QUES 1: Movie Recommendation systems are an example of:

Answer. a) 2 Only

QUES 2. Sentiment Analysis is an example of:

Answer. i) Regression

QUES 3. Can decision trees be used for performing clustering?

Answer. a) True

QUES 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:

Answer. A) 1 only

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

Answer 5. B) 1

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

Answer. B) No

QUES 7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

Answer. A) Yes

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

Answer 8. d) All of the above

QUES 9. Which of the following algorithms is most sensitive to outliers?

Answer. A) K-means clustering algorithm

QUES 10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

Answer. D) All of the above

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

Answer 11. D) All of the above

QUES 12. Is K sensitive to outliers?

Answer: Yes. 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.

QUES 13. Why is K means better?

Answer. K means better because of the following reasons:

1. Relatively simple to implement.
2. Scales to large data sets.
3. Guarantees convergence.
4. Can warm-start the positions of centroids.
5. Easily adapts to new examples.
6. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

QUES 14. Is K means a deterministic algorithm?

Answer. No. K-Means is a non-deterministic algorithm.