

## MACHINE LEARNING Answer Key

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

Ans: b) 1 and 2

2. Sentiment Analysis is an example of:

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

Ans: d) 1,2and4

3. Can decision trees be used for performing clustering?

Ans: A) True

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:

- i) Capping and flooring of variables  
ii) Removal of outliers

Ans: a) 1 only

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

Ans: b) 1

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

Ans: b) No

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

Ans: a) Yes

8. 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.

Ans: d) All of the above

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

Ans: a) K-means clustering algorithm

10. 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

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

12. 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.

Because the mean, as a statistic, is generally sensitive to outliers.

The mean of 2,2,2,3,3,3,4,4,4,2,2,2,3,3,3,4,4,4 is 3.3

If we add a single 23.23 to that, the mean becomes 5.5, which is larger than *any* of the other values.

Since in k-means, you'll be taking the mean a lot, you wind up with a lot of outlier-sensitive calculations.

That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers.

13. Why is K means better?

Ans: k-means is one of the simplest algorithm which uses unsupervised learning method to solve known clustering issues. It works really well with large datasets.

However, there are also drawbacks of K-Means which are:

Strong sensitivity to outliers and noise

Doesn't work well with non-circular cluster shape -- number of cluster and initial seed value need to be specified beforehand

Low capability to pass the local optimum.

Is there anything great about k-means, because it seems that the drawbacks are beyond the good things about k-means.

14. Is 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.

The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. ... The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.