

Machine Learning

Assignment 2

Q1) Movie Recommendation systems are an example of:

- i) Classification
- ii) Clustering
- iii) Regression

Options:

- A. 2 Only
- B. 1 and 2
- C. 1 and 3
- D. 2 and 3

Ans : Movie Recommendation systems are an example of Clustering and reinforcement learning

Q2) Sentiment Analysis is an example of:

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

Options:

- A. 1 Only
- B. 1 and 2
- C. 1 and 3
- D. 1, 2 and 4

Ans : D

Q3) Can decision trees be used for performing clustering?

- A. True
- B. False

Ans : A

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

Options:

- A. 1 only
- B. 2 only
- C. 1 and 2
- D. None of the above

Ans : A

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

- A. 0
- B. 1
- C. 2
- D. 3

Ans : B

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

- A. Yes
- B. No

Ans : B

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

- A. Yes
- B. No
- C. Can't say
- D. None of these

Ans : A

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

Options:

- A. 1, 3 and 4
- B. 1, 2 and 3
- C. 1, 2 and 4
- D. All of the above

Ans : D

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

- A. K-means clustering algorithm
- B. K-medians clustering algorithm
- C. K-modes clustering algorithm
- D. K-medoids clustering algorithm

Ans : A

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.

Options:

- A. 1 only
- B. 2 only
- C. 3 and 4
- D. All of the above

Ans : D

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

- A. Proximity function used
- B. of data points used
- C. of variables used
- D. All of the above

Ans : D

Q12) Is K sensitive to outliers?

Ans : Yes, K means is sensitive to outliers . The k-means algorithm takes the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the points are packed well together, the average is important. However, if there is outliers, this can affect the average calculation of the whole cluster. As a result, this will shift our cluster center closer to the outlier.

Here the data points are assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroid would be minimum. K-means clustering algorithm is sensitive to outliers because the mean is easily influenced by extreme values.

Q13) Why is K means better?

Ans : It is easy to implement k-means and identify unknown groups of data from complex and large data sets. The results are presented in an easy and simple manner. K-means algorithm can easily adjust to the changes. Its efficiency is good.

Compared to hierarchical algorithms, k-means produce tighter clusters especially with globular clusters.

K-means analysis improves clustering accuracy and ensures information about a particular problem domain is available.

This mode of clustering works great when dealing with spherical clusters. It operates with an assumption of joint distributions of features since each cluster is spherical. All the clusters features or characters have equal variance and each is independent of the other.

Q14) Is K means a deterministic algorithm ?

Ans : No, k-means is not deterministic algorithm .

As deterministic algorithm is an algorithm that is purely determined by its inputs, where no randomness is involved in the model. Deterministic algorithms will always come up with the same result given the same inputs whereas K-Means is non-deterministic in nature is due to its random selection of data points as initial centroids.