

Worksheet Set 2

Machine Learning

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

Correct Answer: B) Classification & Clustering

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

Correct Answer: D) Regression, Classification & Reinforcement

Q3. Can decision trees be used for performing clustering?

- a) True b) False

Correct 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:

- 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

Correct Answer: A) Capping & Flooring of Variables

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

a) 0 b) 1 c) 2 d) 3

Correct Answer: B) 1

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

a) Yes b) No

Correct Answer: B) No

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

Correct Answer: 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 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

Correct Answer: D) All the above

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

Correct Answer: A) K-means algorithm

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

Correct Answer: D) All the above

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

Correct Answer: D) All the above

Q12. Is K sensitive to outliers?

K means clustering algorithm is sensitive to outliers because a mean can be influenced easily by extreme values. Mean can be greatly influenced by the outlier and thus cannot represent the correct cluster centre. Instead of this we can use the K-medoids which is more robust to outliers and noises.

Q13. Why K Means is better?

Advantages of K Means

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

Q14. Is K Means a deterministic algorithm?

K-Mean is a non-deterministic algorithm which means a compiler cannot solve the problem in polynomial time and doesn't know clearly the next step. If we run the algorithm several times on the same data, could give the different results.

