ASSIGNMENT – 2

MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. 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

Answer :- a) 2 Only

---

2. 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

Answer :- d) 1, 2 and 4

---

3. Can decision trees be used for performing clustering?
a) True b) False

Answer :- 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
Options:
a) 1 only b) 2 only
c) 1 and 2 d) None of the above

Answer :- a) 1 only

---

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

a) 0 b) 1

c) 2 d) 3

Answer :- b) 1

---

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

a) Yes b) No

Answer :- b) No

---

7. 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

Answer :- 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.

Options:

a) 1, 3 and 4 b) 1, 2 and 3

c) 1, 2 and 4 d) All of the above

Answer :- d) All of the above

---

9. 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

Answer: - 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.
Options:
a) 1 only b) 2 only
c) 3 and 4 d) All of the above

Answer :- 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?
a) Proximity function used
b) of data points used
c) of variables used
d) All of the above

Answer :- d) All of the above

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?
Answer: - Yes, K is sensitive to outliers. If we rescale the data by using normalization or standards, the outcome will be different.

---

13. Why is K means better?
Answer: - K-Means is very popular clustering algorithm in data mining is easy to code and understand. Some key features are as follows:

- 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.

---

14. Is K means a deterministic algorithm?

Answer: - No, k means is non-deterministic algorithm. The non-deterministic nature of K-mean is due to its random selection of data points as initial centroids. The key idea of the algorithm is to select data points which are belong to dense regions and which are adequately separated in future space as the initial centroids.

----------********----------