MACHINE LEARNING 2

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:

c) 1 and 3

2. Sentiment Analysis is an example of:
i) Regression
ii) Classification
iii) Clustering
iv) Reinforcement
Options:

d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

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

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

b) 1

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

b) No

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

d) None of these

ASSIGNMENT – 2

MACHINE LEARNING

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

d) All of the above

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

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?

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?
Ans:-The k-means algoriythm is sensitive to the outliers .in this paper we propose a robust two stage k means clustering algorithms based
on the observationpoint mechanism,which can accurately discover the cluster centers without the disturbance of outliers.

13. Why is K means better?
Ans:-Guarantees convergence.can warm of centeroids.easily adapts to new examples.Generalizes to clusters of different shapes and sizes,
such as elliptical clusters.

14. Is K means a deterministic algorithm?
Ans:-The non deterministic nature of k means is due to its random selection of data points as initial centroid .method

we propose an improved ,density based version of k means,which involves a novel and systematic method for selecting initial centroids