Quiz_Introduction to Data Science

Name *

Akash Kumar Yadav

Roll No. *

R271223114

SAP ID *

500124804

Programme and Branch *

MCA

Batch *

B3

Which of the following can act as possible termination conditions in K-Means?

1. For a fixed number of iterations.
2. The assignment of observations to clusters does not change between iterations, except for cases with a bad local minimum.
3. Centroids do not change between successive iterations.
4. Terminate when RSS falls below a threshold. (1 Point) *

- ○ 1, 2, 3
- ○ 1, 2, 4
- ○ 2, 3, 4
- ◉ All of the above

Principal component is a technique for (1 Point) *

- ○ Feature selection

- [x] Dimensionality reduction
- [ ] Exploration
- [ ] None of the aboive

Out of 200 emails, a classification model correctly predicted 150 spam emails and 30 ham emails. What is the accuracy of the model? (1 Point) *

- [ ] 10 %
- [x] 90 %
- [ ] 80 %
- [ ] None of the above

What is the minimum no. of variables/ features required to perform clustering? (1 Point) *

- [ ] 0
- [x] 1
- [ ] 2

- ○ 3

Any hypothesis find an approximation of the target function over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples. This is called _____. (1 Point) *

- ○ Hypothesis
- ◉ Inductive hypothesis
- ○ Learning
- ○ Concept learning

Which of the following algorithm is most sensitive to outliers? (1 Point) *

- ◉ K-means clustering algorithm
- ○ K-medians clustering algorithm
- ○ K-modes clustering algorithm
- ○ K-medoids clustering algorithm

Which of the following is a data reduction technique? (1 Point) *

- ○ Clustering
- ◉ Classification
- ○ Sampling
- ○ Regression

Which of the following is not a type of decision tree node? (1 Point) *

- ○ Root node
- ○ Leaf node
- ○ Decision node
- ◉ Branch node

Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabeled data? (1 Point) *

- ○ Supervised learning

- (●) Unsupervised learning
- ( ) Hybrid learning
- ( ) Reinforcement learning

Which of the following is a classification algorithm? (1 Point) *

- (●) K-means
- ( ) Decision tree
- ( ) Apriori
- ( ) None of the above

Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than the desirable number of data points?

1. Capping and flouring of variables
2. Removal of outliers

(1 Point) *

- (●) 1 only
- ( ) 2 only

- ( ) 1 and 2 only
- ( ) None of the above

Can decision trees be used for performing clustering? (1 Point) *

- (•) True
- ( ) False

Which is a type of machine learning where a target feature, which is of categorical type, is predicted for the test data on the basis of the information imparted by the training data? (1 Point) *

- ( ) Unsupervised Learning
- ( ) Supervised Regression
- ( ) Supervised Classification
- (•) Categorical Attribute

For two runs of K-Mean clustering, is it expected to get the same clustering results? (1 Point) *

- ( ) Yes

- [x] No

Is it possible that the assignment of observations to clusters does not change between successive iterations in K-Means? (1 Point) *

- [x] Yes
- [ ] No
- [ ] Maybe
- [ ] None of above

Which of the following is not a measure of similarity used in clustering? (1 Point) *

- [ ] Euclidean distance
- [ ] Manhattan distance
- [ ] Cosine similarity
- [x] Entropy

In feature extraction, some of the commonly used ___ are used for combining the original features. (1 Point) *

- (●) Operators
- ( ) Delimiters
- ( ) Words
- ( ) All of the above

Which of the following options is true about the kNN algorithm? (1 Point) *

- ( ) It can be used only for classification
- ( ) It can be used only for regression
- (●) It can be used for both classification and regression
- ( ) It is not possible to use for both classification and regression

This type of interpretation of probability tries to quantify the uncertainty of some event and thus focuses on information rather than repeated trials. (1 Point) *

- (●) Frequency interpretation of probability

- ○ Gaussian interpretation of probability
- ○ Machine learning interpretation of probability
- ○ Bayesian interpretation of probability

A good feature has which of the following characteristics? (1 Point) *

- ○ It should be known at prediction time
- ○ It should be related to the objective
- ○ It should be numeric with meaningful magnitude
- ◉ All of the above