**Machine Learning Questions**

1. Spam filtering for e-mails can be seen as an example of which of the following types of approaches of data mining?

A) reduction

B) association

C) cause-and-effect modeling

D) classification

Ans: D

2. \_\_\_\_\_\_\_\_ is a collection of techniques that seek to group or segment a collection of objects or observations into subsets, such that those within each subset are more closely related to one another than objects assigned to different subsets.

A) Association rule mining

B) Discriminant analysis

C) Cluster analysis

D) Logistic regression

Answer: C

3. Which of the following is the first stage of joining clusters in agglomerative hierarchical clustering?

A) by separating clusters into finer groups

B) by joining two clusters farthest away from each other

C) by joining two clusters that are not at a Euclidean distance

D) by joining two clusters that are closest to each other

Ans: D

4. In classification, which of the following would be considered as a categorical variable of interest for a credit approval decision for a requester?

A) age of the requester

B) income of the requester

C) revolving balance of the requester

D) reject or accept credit approval

Answer: D

5. Which of the following data sets provides the most realistic estimate of the performance of a model on completely unseen data?

A) validation data set

B) test data set

C) training data set

D) linear regression data set

Answer: B

6. Which of the following data sets provides the most realistic estimate of the performance of a model on completely unseen data?

A) validation data set

B) test data set

C) training data set

D) linear regression data set

Answer: B

7. \_\_\_\_\_\_\_\_ is the ratio of the number of transactions that include all items in the consequent as well as the antecedent to the number of transactions that include all items in the antecedent.

A) Lift

B) Logit

C) Support for the association rule

D) Confidence of the association rule

Answer: D

8. Divisive clustering method is different from agglomerative clustering methods in that divisive clustering methods \_\_\_\_\_\_\_\_.

A) can only have a pair of subjects in each cluster

B) separate objects into a particular cluster in one step

C) separate *n* objects successively into finer groupings

D) can only have a single subject in each cluster

Answer: C

9. ) \_\_\_\_\_\_\_\_ is a collection of techniques that seek to group or segment a collection of objects or observations into subsets, such that those within each subset are more closely related to one another than objects assigned to different subsets.

A) Association rule mining

B) Discriminant analysis

C) Cluster analysis

D) Logistic regression

Answer: C

10. U-botit.com is an electronic commerce company that sells music online. It keeps a tab of what genre of music their registered customers buy. If u-botit.com were to use the data mining approach of association, which of the following actions would it take?

A) send recommendations to customers based on their buying habits

B) conduct surveys to customers to gauge customer satisfaction

C) classify the customers based on the genre of music they listen to

D) provide discounts to certain genre buyers where sales are less

Answer: A