MACHINE LEARNING

In Q1 to Q7, only one option is correct, Choose the correct option:

1. What is the advantage of hierarchical clustering over K-means clustering?
   A) Hierarchical clustering is computationally less expensive
   B) In hierarchical clustering you don't need to assign number of clusters in beginning
   C) Both are equally proficient   D) None of these
   Ans:- (B) In hierarchical clustering you don't need to assign number of clusters in beginning

2. Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?
   A) max_depth   B) n_estimators
   C) min_samples_leaf   D) min_samples_splits
   Ans:-(A) max_depth

3. Which of the following is the least preferable resampling method in handling imbalance datasets?
   A) SMOTE   B) RandomOverSampler
   C) RandomUnderSampler   D) ADASYN
   Ans:- (C) RandomUnderSampler

4. Which of the following statements is/are true about "Type-1" and "Type-2" errors?
   1. Type1 is known as false positive and Type2 is known as false negative.
   2. Type1 is known as false negative and Type2 is known as false positive.
   3. Type1 error occurs when we reject a null hypothesis when it is actually true.
   A) 1 and 2   B) 1 only
   C) 1 and 3   D) 2 and 3
   Ans:-(B) 1 only

5. Arrange the steps of k-means algorithm in the order in which they occur:
   1. Randomly selecting the cluster centroids
   2. Updating the cluster centroids iteratively
   3. Assigning the cluster points to their nearest center
   A) 3-1-2   B) 2-1-3
   C) 3-2-1   D) 1-3-2
   Ans:-(A) 3-1-2

6. Which of the following algorithms is not advisable to use when you have limited CPU resources and time, and when the data set is relatively large?
   A) Decision Trees   B) Support Vector Machines
   C) K-Nearest Neighbors   D) Logistic Regression
   Ans:-(C) K-Nearest Neighbors

7. What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Interaction Detection) Trees?
   A) CART is used for classification, and CHAID is used for regression.
   B) CART can create multiway trees (more than two children for a node), and CHAID can only create binary trees (a maximum of two children for a node).
   C) CART can only create binary trees (a maximum of two children for a node), and CHAID can create multiway trees (more than two children for a node)
   D) None of the above
   Ans:-()

In Q8 to Q10, more than one options are correct, Choose all the correct options:

8. In Ridge and Lasso regularization if you take a large value of regularization constant(lambda), which of the following things may occur?
   A) Ridge will lead to some of the coefficients to be very close to 0

MACHINE LEARNING

B) Lasso will lead to some of the coefficients to be very close to 0

C) Ridge will cause some of the coefficients to become 0

D) Lasso will cause some of the coefficients to become 0.

Ans:-(A) Ridge will lead to some of the coefficients to be very close to 0

(B) Lasso will lead to some of the coefficients to be very close to 0

MACHINE LEARNING

9. Which of the following methods can be used to treat two multi-collinear features?
    A) remove both features from the dataset
    B) remove only one of the features
    C) Use ridge regularization   D) use Lasso regularization

10. After using linear regression, we find that the bias is very low, while the variance is very high. What are the possible reasons for this?
    A) Overfitting   B) Multicollinearity
    C) Underfitting   D) Outliers

Q10 to Q15 are subjective answer type questions, Answer them briefly.

11. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?
12. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.
13. What is the difference between SMOTE and ADASYN sampling techniques?
14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?
15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief.