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

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
   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.

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?
   If the Categorical feature is Ordinal . In ordinal we can use Dummy variable technique and use to drop First or last in pd.get_dummies() method.

12. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.
   Resampling (Oversampling and Undersampling)
   This technique is used to upsample or downsample the minority or majority class. When we are using an imbalanced dataset, we can oversample the minority class using replacement. This technique is called oversampling. Similarly, we can randomly delete rows from the majority class to match them with the minority class which is called undersampling. After sampling the data we can get a balanced dataset for both majority and minority classes. So, when both classes have a similar number of records present in the dataset, we can assume that the classifier will give equal importance to both classes.
   SMOTE
   Synthetic Minority Oversampling Technique or SMOTE is another technique to oversample the minority class. Simply adding duplicate records of minority class often don't add any new information to the model. In SMOTE new instances are synthesized from the existing data. If we explain it in simple words, SMOTE looks into minority class instances and use k nearest neighbor to select a random nearest neighbor, and a synthetic instance is created randomly in feature space.

13. What is the difference between SMOTE and ADASYN sampling techniques?

   Oversample using Adaptive Synthetic (ADASYN) algorithm. This method is similar to SMOTE but it generates different number of samples depending on an estimate of the local distribution of the class to be oversampled.

14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?
   GridSearchCV is a technique for **finding the optimal parameter values from a given set of parameters in a grid**. It's essentially a cross-validation technique. The model as well as the parameters must be entered. After extracting the best parameter values, predictions are made

   No,it is preferable as If the hyper-parameters range or number is high, the possibilities can be in millions and it will take so much time to finish.

MACHINE LEARNING

RandomizedSearchCV solves the drawbacks of GridSearchCV, as it goes through only a fixed number of hyperparameter settings. It moves within the grid in a random fashion to find the best set hyperparameters. This approach reduces unnecessary computation.

15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief

a) Mean Squared Error: 1 / N * sum for i to N (y_i – yhat_i)^2

It is differentiable, converge faster but not robust to outliers

b) Root Mean Squarred Error: sqrt(1 / N * sum for i to N (y_i – yhat_i)^2)

c) Mean Absolute Error: 1 / N * sum for i to N |y_i – yhat_i|

It is robust to outliers

d) R2: 1 – (Squared sum error of regression line) / (Squared sum error of mean line)

R2 score is a metric that tells the performance of your model, not robust to larger number of features that is why we use the adjusted R2 instead , also Adjusted R2 used for population datasets.