MACHINE LEARNING

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

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

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

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: C

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: D

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: B

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: C

MACHINE LEARNING

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.
   Ans: A & D

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
   Ans: B, C or D

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
    Ans: A

11. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?
    Ans: One-hot encoding creates d-dimensional vectors for each instance where d is the unique number of feature values in the dataset. For a feature having a large number of unique feature values or categories, one-hot encoding is not a great choice. To fight the curse of dimensionality, binary encoding might be a good alternative to one-hot encoding because it creates fewer columns when encoding categorical variables. Ordinal encoding is a good choice if the order of the categorical variables matters.

12. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.
    Ans: 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.

13. What is the difference between SMOTE and ADASYN sampling techniques?
    Ans: The key difference between ADASYN and SMOTE is that the former uses a density distribution, as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed distributions.

MACHINE LEARNING

14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?

    Ans: GridSearchCV is a library function that is a member of sklearn's model_selection package. It helps to loop through predefined hyperparameters and fit your estimator (model) on your training set. So, in the end, you can select the best parameters from the listed hyperparameters. The grid Search Cross-Validation technique is computationally expensive. The complexity of Grid Search CV increases with an increase in the number of parameters in the param grid. Thus Grid Search CV technique is not recommended for large-size datasets or param grids with a large number of components.

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

    Ans: There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are:

    a. Mean Squared Error (MSE): Another way to do so is by squaring the distance, so that the results are positive. This is done by the **MSE**, and higher errors (or distances) weigh more in the metric than lower ones, due to the nature of the power function.

    b. Root Mean Squared Error (RMSE): Root mean square error or root mean square deviation is one of the most commonly used measures for evaluating the quality of predictions. It shows how far predictions fall from measured true values using Euclidean distance.

    c. Mean Absolute Error (MAE): In short, **MAE** evaluates the absolute distance of the observations (the entries of the dataset) to the predictions on a regression, taking the average over all observations. We use the absolute value of the distances so that negative errors are accounted properly.