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- Hierarchical clustering is computationally less expensive

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- 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
   ANS- SMOTE

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- 1 and 2

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) 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

   ANS- Support Vector Machines

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).

MACHINE LEARNING

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- CART can create multiway trees (more than two children for a node), and CHAID can only createbinary trees (a maximum of two children for a node).

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.

ANS-   Ridge 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
      ANS- remove both features from the dataset

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- Overfitting

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?
    ANS- For a feature having a large number of unique feature values or categories, one-hot encoding is not a great choice. There are various other techniques to encode the categorical (ordinal or nominal) features

For categorical variables where no such ordinal relationship exists, the integer encoding is not enough.

In fact, using this encoding and allowing the model to assume a natural ordering between categories may result in poor performance or unexpected results (predictions halfway between categories).

In this case, a one-hot encoding can be applied to the integer representation. This is where the integer encoded variable is removed and a new binary variable is added for each unique integer value.

In the "*color*" variable example, there are 3 categories and therefore 3 binary variables are needed. A "1" value is placed in the binary variable for the color and "0" values for the other colors.

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

- Choose Proper Evaluation Metric. The accuracy of a classifier is the total number of correct predictions by the classifier divided by the total number of predictions. ...
- Resampling (Oversampling and Undersampling) ...
- SMOTE. ...
- BalancedBaggingClassifier. ...
- Threshold moving.

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

MACHINE LEARNING

ANS-

SMOTE: Synthetic Minority Over sampling Technique (SMOTE) algorithm applies KNN approach where it selects K nearest neighbors, joins them and creates the synthetic samples in the space. The algorithm takes the feature vectors and its nearest neighbors, computes the distance between these vectors. The difference is multiplied by random number between (0, 1) and it is added back to feature. SMOTE algorithm is a pioneer algorithm and many other algorithms are derived from SMOTE.

ADASYN: ADAptive SYNthetic (ADASYN) is based on the idea of adaptively generating minority data samples according to their distributions using K nearest neighbor. The algorithm adaptively updates the distribution and there are no assumptions made for the underlying distribution of the data. The algorithm uses Euclidean distance for KNN Algorithm. 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. The latter generates the same number of synthetic samples for each original minority sample.

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 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.

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

- ANS- Regression

- Why we require Evaluation Metrics

- Mean Absolute Error(MAE)

- Mean Squared Error(MSE)

- RMSE

- RMSLE

- R squared

- Adjusted R Squares

- EndNote

Regression

Regression is a type of Machine learning which helps in finding the relationship between independent and dependent variable.

MACHINE LEARNING

In simple words, Regression can be defined as a Machine learning problem where we have to predict discrete values like price, Rating, Fees, etc.

Why We require Evaluation Metrics?

Most beginners and practitioners most of the time do not bother about the model performance. The talk is about building a well-generalized model, Machine learning model cannot have 100 per cent efficiency otherwise the model is known as a biased model. which further includes the concept of overfitting and underfitting.

Mean Absolute Error(MAE)

MAE is a very simple metric which calculates the absolute difference between actual and predicted values.

To better understand, let's take an example you have input data and output data and use Linear Regression, which draws a best-fit line.

Mean Squared Error(MSE)

MSE is a most used and very simple metric with a little bit of change in mean absolute error. Mean squared error states that finding the squared difference between actual and predicted value.

So, above we are finding the absolute difference and here we are finding the squared difference.

Root Mean Squared Error(RMSE)

As RMSE is clear by the name itself, that it is a simple square root of mean squared error.

Root Mean Squared Log Error(RMSLE)

Taking the log of the RMSE metric slows down the scale of error. The metric is very helpful when you are developing a model without calling the inputs. In that case, the output will vary on a large scale.

To control this situation of RMSE we take the log of calculated RMSE error and resultant we get as RMSLE.

Adjusted R Squared

The disadvantage of the R2 score is while adding new features in data the R2 score starts increasing or remains constant but it never decreases because It assumes that while adding more data variance of data increases.

MACHINE LEARNING

EndNote

I hope it was easy to catch all the important 6 metrics we have discussed. There is not anyone metric that always performs well and helps to build the generalized model.

There can be situations where you have to use different evaluation metrics and if you are a beginner then you should try all these metrics which will help you to get a better understanding of each to evaluate when you can use which metric.