FLIP ROBO

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
   A) Least Square Error
   B) Maximum Likelihood
   C) Logarithmic Loss
   D) Both A and B

2. Which of the following statement is true about outliers in linear regression?
   A) Linear regression is sensitive to outliers
   B) linear regression is not sensitive to outliers
   C) Can't say
   D) none of these

3. A line falls from left to right if a slope is_____?
   A) Positive
   B) Negative
   C) Zero
   D) Undefined

4. Which of the following will have symmetric relation between dependent variable and independent variable?
   A) Regression
   B) Correlation
   C) Both of them
   D) None of these

5. Which of the following is the reason for over fitting condition?
   A) High bias and high variance
   B) Low bias and low variance
   C) Low bias and high variance
   D) none of these

6. If output involves label then that model is called as:
   A) Descriptive model
   B) Predictive modal
   C) Reinforcement learning
   D) All of the above

7. Lasso and Ridge regression techniques belong to_____?
   A) Cross validation
   B) Removing outliers
   C) SMOTE
   D) Regularization

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation
   B) Regularization
   C) Kernel
   D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses_____to make graph?
   A) TPR and FPR
   B) Sensitivity and precision
   C) Sensitivity and Specificity
   D) Recall and precision

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    A) True
    B) False

11. Pick the feature extraction from below:
    A) Construction bag of words from a email
    B) Apply PCA to project high dimensional data
    C) Removing stop words
    D) Forward selection

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
    A) We don't have to choose the learning rate.
    B) It becomes slow when number of features is very large.
    C) We need to iterate.
    D) It does not make use of dependent variable.

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

14. Which particular algorithms are used for regularization?

There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

- **Ridge Regularization**

  Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

  This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

- **Lasso Regression**

  It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

  Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients

15. Explain the term error present in linear regression equation?

An error term is the sum of the deviations of each actual observation from a model regression line. Regression analysis is used to establish the degree of correlation between two variables, one independent and one dependent, the result of which is a line that best "fits" the actually observed values of the dependent value in relation to the independent Variable or Variables.

The error term is a measure of how accurately the regression model reflects the actual relationship between the independent and dependent variable or variables. The error term can indicate either that the model can be improved, such as by adding in another independent variable that explains some or all of the difference, or by randomness, meaning that the dependent and independent variable or variables are not correlated to any greater degree.

MACHINE LEARNING

There are actually two types of error terms commonly used in regression analysis: absolute error and relative error. Absolute error is the error term as previously defined, the difference between the actually observed values of the independent variable and the results predicted by the model. Derived from this, relative error is defined as the absolute error divided by the exact value predicted by the model.