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.

FLIP ROBO

MACHINE LEARNING

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

13. Explain the term regularization?

Regularization is a machine learning technique that reduces overfitting and increases model accuracy. It operates by including a penalty term in the cost function of the model, which helps to simplify the model and increase generalizability. Regularization can be applied to different types of models, such as linear regression, logistic regression, neural networks, etc. There are different types of regularization methods, such as L1, L2, dropout, etc., which have different effects on the model parameters and performance.

14. Which particular algorithms are used for regularization?

- Lasso regression: This algorithm uses the L1 regularization technique, which adds a penalty term proportional to the absolute values of the model parameters to the cost function. This technique tends to shrink some of the coefficients to zero, resulting in a sparse model that selects only the most relevant features12.
- Ridge regression: This algorithm uses the L2 regularization technique, which adds a penalty term proportional to the squared values of the model parameters to the cost function. This technique tends to reduce the magnitude of all the coefficients, but does not eliminate them completely. This technique is useful when there are many correlated features12.
- Elastic net: This algorithm uses a combination of L1 and L2 regularization techniques, which adds a penalty term that is a weighted sum of the absolute and squared values of the model parameters to the cost function. This technique balances the advantages and disadvantages of both lasso and ridge regression, and can handle both sparse and correlated features13.

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

The term error in linear regression equation refers to the difference between the observed value of the output variable and the predicted value by the model. It is also known as the residual12. The error can be calculated by subtracting the predicted value from the observed value for each data point, and then squaring each of these differences. The mean of these squared differences is called the mean square error (MSE)23. The MSE is a measure of how well the model fits the data. A lower MSE indicates a better fit. The standard error of the regression line is another measure of the error that indicates how much the regression equation over- or under-predicts