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

   Answer: A) Least Square Error

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

   Answer: A) Linear regression is sensitive to outliers

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

   Answer: B) Negative

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

   Answer: B) Correlation

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

   Answer: C) Low bias and high variance

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

   Answer: B) Predictive model

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

   Answer: D) Regularization

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

MACHINE LEARNING

C) Kernel   D) SMOTE

Answer: 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

   Answer: A) TPR and FPR

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

    Answer: 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

    Answer: B) Apply PCA to project high dimensional data

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.

    Answer: B) It becomes slow when number of features is very large.

MACHINE LEARNING

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

13. Explain the term regularization?

Answer: Regularization is a technique used in machine learning and statistical modeling to prevent overfitting and improve the generalization performance of models. Overfitting occurs when a model fits the training data too closely, capturing noise and random fluctuations in the data, which leads to poor performance on new, unseen data. Two common types of regularization techniques are L1 regularization (Lasso) and L2 regularization (Ridge). L1 regularization adds the absolute values of coefficients to the penalty term, encouraging some coefficients to become exactly zero, effectively performing feature selection. L2 regularization adds the squares of coefficients to the penalty term, which discourages the coefficients from becoming too large.

14. Which particular algorithms are used for regularization?

Answer: Several machine learning algorithms use regularization techniques to prevent overfitting and improve generalization. Some of the common algorithms that incorporate regularization are:

1. **Linear Regression with Ridge and Lasso:** Ridge regression (L2 regularization) and Lasso regression (L1 regularization) are extensions of linear regression that add a penalty term to the regression objective to prevent overfitting.
2. **Logistic Regression with Ridge and Lasso:** Similar to linear regression, logistic regression can also be regularized using Ridge and Lasso techniques.

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

Answer: In the context of linear regression, the term "error" refers to the difference between the actual observed values of the dependent variable and the values predicted by the linear regression model. This difference is also known as the "residual" or "residual error."

Mathematically, the linear regression equation can be represented as:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$

- y is the actual observed value of the dependent variable.
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the coefficients that the linear regression model aims to estimate.
- $x_1, x_2, \ldots, x_n x_1, x_2, \ldots, x_n$ are the predictor variables (features) used in the model.
- ϵ represents the error term.