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)

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)

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

Answer (B)

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)

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)

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)

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

Answer (B)

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

Answer (D)

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)

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)

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

MACHINE LEARNING

C) Removing stop words
D) Forward selection

Answer (A)

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, D)

MACHINE LEARNING

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

13. Explain the term regularization?

Answer – In machine learning, regularization is a technique used to prevent overfitting of a model on training data. Overfitting occurs when a model is too complex and fits the training data very closely, but performs poorly on new, unseen data. Regularization helps to control the complexity of a model by adding a penalty term to the cost function that the model tries to minimize during training.

The penalty term is usually a function of the model's parameters, such as their magnitude or their variance. By adding this penalty term to the cost function, the model is encouraged to find a set of parameters that not only minimize the training error but also have a simpler structure. This helps to reduce overfitting and improve the generalization performance of the model.

There are different types of regularization techniques, such as L1 regularization (also known as Lasso), L2 regularization (also known as Ridge regression), and dropout regularization. L1 regularization adds a penalty term proportional to the absolute value of the model's parameters, which encourages sparsity in the parameter values. L2 regularization adds a penalty term proportional to the square of the model's parameters, which encourages small values for all parameters. Dropout regularization randomly sets some of the neurons in a neural network to zero during training, which helps to prevent overfitting by forcing the network to learn more robust features.

Regularization is an important technique in machine learning because it helps to improve the generalization performance of a model and prevent overfitting.

Q14. particular algorithms are used for regularization?

Answer - Regularization is a technique used in machine learning to prevent overfitting of a model on the training data. There are several algorithms used for regularization, including:

1. L1 Regularization (Lasso Regression): This algorithm adds a penalty term to the loss function of the model that is proportional to the absolute value of the model's coefficients. It forces some of the coefficients to become zero, effectively performing feature selection.

2. L2 Regularization (Ridge Regression): This algorithm adds a penalty term to the loss function of the model that is proportional to the square of the model's coefficients. It shrinks the coefficients towards zero, reducing the impact of the less important features.

3. Elastic Net Regularization: This algorithm is a combination of L1 and L2 regularization. It adds both penalty terms to the loss function of the model. It can be useful when there are many correlated features.

4. Dropout Regularization: This algorithm randomly drops out some of the nodes in a neural network during training, forcing the network to learn more robust features.

5. Early Stopping: This algorithm stops the training of the model when the performance on the validation set starts to deteriorate, preventing overfitting.

Q15. the term error present in linear regression equation?

Answer - In linear regression, the term "error" refers to the difference between the actual observed values and the predicted values by the linear regression model. It is also commonly known as "residuals".

In a linear regression equation, the goal is to fit a straight line that best represents the relationship between a dependent variable (Y) and one or more independent variables (X) based on the observed data. The linear regression equation is of the form:

$Y = \beta 0 + \beta 1 * X + \varepsilon$

where:

* Y is the dependent variable

* X is the independent variable(s)

* β0 is the intercept, which represents the expected value of Y when X is zero

MACHINE LEARNING

* β1 is the slope, which represents the change in Y for a unit change in X

* ε is the error term, which represents the unexplained variation in Y that is not accounted for by the linear relationship with X.

The error term ε captures the variability in the dependent variable Y that is not explained by the linear relationship with the independent variable(s) X. It represents the part of Y that is left unexplained by the model. It includes factors such as measurement errors, unobserved variables, and random fluctuations that affect the true relationship between Y and X. The goal of linear regression is to minimize the error term ε by finding the best-fitting line that provides the closest approximation to the observed data points.