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?
   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?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

Answer 13. Regularization is a technique used in machine learning and statistical modelling to prevent overfitting by adding a penalty to the loss function. Overfitting occurs when a model learns the noise and details in the training data to the extent that it negatively impacts its performance on new, unseen data. A model that overfits is too complex and captures patterns that do not generalize to other datasets. Regularization is an essential technique in machine learning that helps to balance the trade- off between fitting the training data well and maintaining the model's ability to generalize to new data.

Answer 14. Regularization is a technique used in machine learning to prevent overfitting by adding a penalty to the loss function. Several algorithms and methods implement regularization. The most common types of regularization include:

1. **Lasso Regularization:** Lasso stands for Least Absolute Shrinkage and Selection Operator. It adds the absolute value of the coefficients as a penalty term to the loss function.

2. **Ridge Regularization:** Ridge regularization add the squared value of the coefficients as a penalty term to the loss function.

3. **Elastic Net Regularization:** Elastic Net combines both L1 and L2 regularization penalties.

4. **Dropout Regularization:** Commonly used in neural networks, dropout involves randomly setting a fraction of the neurons to zero during training.

5. **Early Stopping:** This technique involves monitoring the model's performance on a validation set during training and stopping the training process when performance starts to degrade.

Answer 15. In the context of linear regression, the term "error" (often denotes as (\epsilon) or (e)) refers to the difference between the observed values and the values predicted by the linear regression model. Linear regression aims to model the relationship between one or more independent variables (predictors) and a dependent variable (response) by fitting a linear equation to the observed data.

Linear Regression Model:

The basic form of a linear regression equation is:

[
Y= \beta_0 + \beta_2*_2 + \Idots + \beta_n*n + \epsilon
]

Where:

- **(y) is the dependent variable (the output we are trying to predict).**
- **(\beta_0) is the intercept of the regression line (the value of (y) when all (x) values are zero).**
- **(\beta_1, \beta_2, \Idots, \beta_n) are the coefficients of the independent variables (x_1, x_2, \Idots, x_n).**
- **(\epsilon) is the error term.**