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

   ans : 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

   ans: 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

   ans: 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

   ans: 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

   ans: A) High 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
   ans: B) Predictive modal

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

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
   ans: B) Sensitivity 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
    ans: A) True

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

ans: 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.

ans: A) C) D)

MACHINE LEARNING

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

13. Explain the term regularization?

ans: Regularization, meaning in the machine learning context, refers to minimizing or shrinking the coefficient estimates towards zero to avoid underfitting or overfitting the machine learning model.

we will first need to understand the concept of bias and variance.

Bias: Bias is the difference between the actual and predicted values. A machine learning model with high bias gives little consideration to the data pattern, resulting in oversimplified and underfit models.

Variance: Variance is a measure of flexibility of the model. It is the opposite of a model's bias and decides how sensitive the model is to change based on the patterns in the input data. If the variance of a model is high, it will more easily learn from the patterns in the input data and will, as a result, be more susceptible to noise in the data resulting in overfitting.

Consequently, when the bias is high, the training and testing error will be high, and when the variance is high, the training error will be low and the testing error high. There will be a point where the training and test error will be reasonably low between these two extremes. The goal of regularization techniques is to find this point of perfect balance.

Perhaps this explanation of 'bias' and 'variance' should also serve to clarify that the 'balance' we have been going on endlessly about in the previous sections, is in this case, actually a trade-off between bias and variance.

14. Which particular algorithms are used for regularization?

ans: LASSO regression, also known as L1 regularization, is a popular technique used in statistical modeling and machine learning to estimate the relationships between variables and make predictions. LASSO stands for Least Absolute Shrinkage and Selection Operator.

The primary goal of LASSO regression is to find a balance between model simplicity and accuracy. It achieves this by adding a penalty term to the traditional linear regression model, which encourages sparse solutions where some coefficients are forced to be exactly zero. This feature makes LASSO particularly useful for feature selection, as it can automatically identify and discard ir

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

ans: An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.
The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

Error Term Use in a Formula:
An error term essentially means that the model is not completely accurate and results in differing results during real-world applications