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

2. Which of the following statement is true about outliers in linear regression?
   A) Linear regression is sensitive to outliers Answer 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 Answer
   C) Zero D) Undefined

4. Which of the following will have symmetric relation between dependent variable and independent variable?
   A) Regression B) Correlation Answer
   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 Answer D) none of these

6. If output involves label then that model is called as:
   A) Descriptive model B) Predictive modal Answer
   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 Answer

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

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 Answer 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 Answer

11. Pick the feature extraction from below:
    A) Construction bag of words from a email Answer
    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. Answer
    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?

It's a technique used in machine learning to mitigate against overfitting of the model to the training data thereby improving the generalization performance of the model. Overfitting occurs when the model learns the noise or irrelevant patterns in the training data, and as a result, performs poorly on unseen data. Regularization adds a penalty term to the loss function of the model, which helps to discourage overfitting by controlling the complexity of the model.

It also introduces a bias-variance tradeoff. The penalty term added to the loss function can cause the model to underfit if the regularization strength is too high, or to overfit if it is too low. It is important to choose an appropriate regularization strength that balances the bias and variance of the model

14. Which particular algorithms are used for regularization?

Two common types of regularization algorithm techniques used in machine learning are L1 and L2 .

L1, also known as Lasso regularization, adds a penalty term to the loss function that is proportional to the absolute value of the coefficients. It encourages the model to reduce the number of non-zero coefficients, effectively performing feature selection by shrinking the coefficients of the less important features to zero.

L2, also known as Ridge regularization, adds a penalty term to the loss function that is proportional to the square of the coefficients. It encourages the model to reduce the magnitude of the coefficients, effectively performing feature shrinkage by spreading the coefficient values across all the features.

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

Term error is the difference between the predicted values of the dependent variable and the actual values of the dependent variable for a given set of independent variables. The term error is also called the residual and is denoted by the letter e.

Linear regression equation is:

$Y = a + bX + e$

Y is the dependent variable, X is the independent variable, a and b are the intercept and slope coefficients, and e is the error term

It depicts the part of the variation in the dependent variable that cannot be attributed to the relationship with the independent variable(s).