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 ANSWER 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 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 AD) 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 B) Sensitivity and precision
   C) Sensitivity and Specificity ANSWER 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.
    C) We need to iterate.

D) It does not make use of dependent variable. ANSWER

MACHINE LEARNING

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

13. Explain the term regularization?In machine learning, regularization refers to a set of techniques used to prevent overfitting and improve the generalization of a model. The idea behind regularization is to add a penalty term to the loss function of the model, which discourages it from fitting the training data too closely and instead encourages it to learn more general patterns that are likely to hold for unseen data.Two commonly used regularization techniques in linear regression are Ridge Regression and Lasso Regression. Ridge Regression adds a penalty term that is proportional to the square of the magnitude of the coefficients, which leads to a "shrinking" of the coefficients towards zero. Lasso Regression, on the other hand, adds a penalty term that is proportional to the absolute value of the coefficients, which can lead to some coefficients being exactly zero and thus performing feature selection.Regularization is a powerful tool in machine learning and can help to prevent overfitting and improve the performance of models on new, unseen data. However, it is important to choose the regularization parameter carefully to balance the trade-off between bias and variance in the model.

14. Which particular algorithms are used for regularization? There are several algorithms that can be used for regularization, including:Ridge Regression: It adds a penalty term to the least-squares objective, which is proportional to the square of the magnitude of the coefficients.Lasso Regression: It adds a penalty term to the least-squares objective, which is proportional to the absolute value of the coefficients.Elastic Net: It combines the penalties of Ridge and Lasso regression.Dropout Regularization: It randomly drops out a subset of the neurons during training, reducing their inter-dependency Early Stopping: It stops the training process early if the validation error starts to increase.These techniques can help to prevent overfitting and improve the generalization performance of the model.

15. Explain the term error present in linear regression equation? In the context of linear regression, error refers to the difference between the actual values of the dependent variable and the predicted values obtained from the linear regression equation.The linear regression equation is used to model the relationship between the independent variables and the dependent variable, and it estimates the values of the dependent variable for a given set of independent variables. However, due to the presence of random factors and unaccounted variables, the predicted values may not be exact and will have some error associated with them.The objective of linear regression is to minimize the sum of squared errors (SSE) between the actual values of the dependent variable and the predicted values obtained from the linear regression equation. By minimizing the error, the linear regression model is able to fit the data better and make more accurate predictions