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?
   A) 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?

    Answer: Regularization is a technique used in machine learning to prevent overfitting and improve generalization performance. Regularization introduce additional constraints or penalties to model's objective

    Function, which helps control the complexity of the model and prevent it from becoming too sensitive to the training data. The goal is to find a balance between fitting the training data well and maintaining good generalization, reducing overfitting and improving the model's ability to generalize to unseen data.

    Several types of regularization techniques used some are:

    1.Lasso

    2. Ridge

    3.elastic net

    4.Dropout

    5.Early stopping

14. Which particular algorithms are used for regularization?

    Answer: Several algorithms and regularization techniques can be used depending on the specific problem and the nature of data. Some commonly used algorithms are:

    1.**Linear Regression** :This models to prevent overfitting and control magnitude of the coefficients.
    2. **Logistic regression**: Logistic Regression model can benefit from linear regression to avoid overfitting and improve generalization.
    3.**Support vector Machines:** SVM models can be regularized using techniques such as Ridge regularization to control the complexity of the decision boundary.
    4.**Decision Trees**: Decision trees can be regularized using techniques like pruning, where branches of tree are removed based on certain criteria to prevent overfitting.
    5.**Random forests:** Random forests, which are an ensemble of decision trees can benefit from regularization techniques like feature subsampling and bootstrap bagging to improve generalization.

15. Explain 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 obtained from the linear regression equation. It represents the unexplained or residual variation in the dependent variable that the linear regression model cannot account for.

    The linear regression equation is typically represented as:

    $Y=B_0+B_1X_1+B_2X_{2.....}+\varepsilon$

    Where:

    -y is the dependent variable or the target variable being predicted.

    -$X_1X_{2...}$ are independent variable or features used to predict the dependent variable.

    -B_0B_1…. Are the coefficients or weights associated with each independent variable.

    -ε represents the error term or residual.