Machine learning

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:- **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

Answer:- **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

   Answer:- **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

   Answer:- **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

   Answer:- **Low 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

   Answer:- **predictive modal**

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers C) SMOTE D) Regularization

Answer:- **Regularization**

8. To overcome with imbalance dataset which technique can be used?

   A) Cross validation B) Regularization C) Kernel D) SMOTE

   Answer:- **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

   Answer:- **TPR and FPR**

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

    A) True B) False

    Answer:- **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

    Answer:- **Construction bag of words from a email**

12. 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:- **A, B, D.**

13. Explain the term regularization?

    Answer:- **Regularization is a technique used to prevent overfitting of a model by adding a penalty term to the loss function. The goal of regularization is to reduce the complexity of the model and make it more generalizable to new data.**
    There are two main types of regularization techniques used in machine learning: L1 regularization (also known as Lasso regularization) and L2 regularization (also known as Ridge regularization). L1 regularization adds a penalty proportional to the absolute value of the model weights, while L2 regularization adds a penalty proportional to the square

of the model weights. Regularization works by adding a penalty term to the cost function that the model is trying to minimize during training. This penalty term is usually a function of the model weights, and it is designed to make the model prefer smaller weights. The intuition behind this is that smaller weights correspond to a simpler model, which is less likely to overfit to the training data.

14. Which particular algorithms are used for regularization?

Answer:- Regularization can be applied to various machine learning algorithms, such as linear regression, logistic regression, neural networks, and support vector machines (SVMs). In linear regression, two popular regularization techniques are Lasso and Ridge regression. In logistic regression, L1 and L2 regularization can be applied. In neural networks, dropout regularization and L2 regularization are commonly used. In SVMs, L2 regularization is often applied.

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

Answer:- In linear regression, the term "error" refers to the difference between the actual values of the dependent variable and the predicted values generated by the linear regression model. It is also known as the residual.
The linear regression equation can be written as: $y = \beta0 + \beta1*x + \varepsilon$ where y is the dependent variable, x is the independent variable, β0 and β1 are the coefficients, and ε is the error term.