FLIP ROBO TECHNOLOGIES

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
C) Both of them
B) Correlation
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 MODEL

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation
C) SMOTE
B) Removing outliers
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
C) Sensitivity and Specificity
B) Sensitivity and precision
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 -- APPLY PCA TO PROJECT HIGH DIMENSIONAL DATA

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 -- OPTION A AND B

13. Explain the term regularization?

ANSWER --
Regularization is a technique used in machine learning and statistics to prevent overfitting by adding a penalty to the loss function that the model is trying to minimize. This penalty discourages the model from fitting the noise in the training data, promoting simpler models that generalize better to unseen data.

There are several common forms of regularization:

Lasso Regularization (L1 Regularization):

Adds the absolute value of the coefficients as a penalty term to the loss function.
Encourages sparsity in the model, meaning some coefficients may become zero, effectively selecting a simpler model with fewer features.
Ridge Regularization (L2 Regularization):

Adds the square of the coefficients as a penalty term to the loss function.
Shrinks the coefficients but does not set them to zero, which helps to prevent multicollinearity and reduces model complexity.
Elastic Net:

Combines both Lasso and Ridge regularization, adding both L1 and L2 penalties to the loss function.
Useful when there are multiple features that are correlated with each other.

14. Which particular algorithms are used for regularization?

ANSWER --
Linear Regression with Regularization
Logistic Regression with Regularization
Support Vector Machines (SVM)
Neural Networks
Decision Trees and Ensemble Methods
Bayesian Methods

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

ANSWER --

The error (ϵ) represents the difference between the actual value of the dependent variable and the predicted value produced by the model.

Types of Errors:
Residuals: These are the individual errors for each observation in the dataset. They can be calculated for each data point and help analyze how well the model fits.

Mean Squared Error (MSE): This is a common metric used to quantify the overall error in the model. It is calculated by averaging the squares of the residuals:

Significance of Error:
The error term captures all the factors that influence the dependent variable but are not included in the model. This may include omitted variables, measurement errors, or inherent randomness in the data.

Analyzing the errors can provide insights into model performance and help identify potential improvements. For example, systematic patterns in the residuals may indicate that the model is missing important features or that a non-linear model might be more appropriate