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:
A) Least Square Error

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:
A) 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:
B) 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) D) None of these

Answer:
B) 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:
C) 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:
B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularization

Answer:
D) Regularization

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

Answer:
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

Answer:
A) 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:
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

Answer:
A) Construction bag of words from an email

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:
The correct options are:
A) We don’t have to choose the learning rate.
B) It becomes slow when the number of features is very large.

Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

Answer:

Regularization:
Regularization is a technique used in machine learning to prevent overfitting by adding a penalty term to the objective function. The purpose of regularization is to encourage the model to learn a simpler and more generalizable representation of the data. It adds a penalty for complexity, typically by adding a term proportional to the magnitude of the coefficients to the loss function during training. This helps in preventing the model from fitting the noise in the training data and improves its performance on unseen data.

Algorithms Used for Regularization:
Two common algorithms used for regularization in linear regression are:

Lasso Regression (L1 Regularization): Adds the absolute values of the coefficients as a penalty term.
Ridge Regression (L2 Regularization): Adds the squared values of the coefficients as a penalty term.

Error in Linear Regression Equation: The error in the linear regression equation refers to the difference between the predicted values and the actual values in the dataset. In the context of linear regression, the most common measure of error is the residual sum of squares (RSS) or mean squared error (MSE). Mathematically, it is expressed as the sum of the squared differences between the predicted and actual values. Minimizing this error is the objective during the training of a linear regression model. The error represents how well the model fits the given data, and minimizing it leads to finding the best-fitting line.