Q1-Q11: Multiple Choice Questions

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: D) Both A and B

Both Least Squares Error and Maximum Likelihood methods can be used to find the best fit line in linear regression. Least Squares is the most common method, but Maximum Likelihood can also be used in this context.

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

Outliers can significantly affect the results of linear regression because they can skew the best fit line.

3. **A line falls from left to right if a slope is _____?**

- A) Positive

- B) Negative

- C) Zero

- D) Undefined

Answer: B) Negative

A negative slope indicates that the line falls as you move from left to right.

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: B) Correlation

Correlation measures the symmetric relationship between two variables, unlike regression which is typically used to predict the dependent variable based on the independent variable.

5. **Which of the following is the reason for overfitting 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

Overfitting occurs when a model has low bias but high variance, meaning it fits the training data very well but generalizes poorly to new data.

6. **If output involves label then that model is called as:**

- A) Descriptive model
- B) Predictive model
- C) Reinforcement learning
- D) All of the above

Answer: B) Predictive model

A model that involves labels and is used to predict future outcomes based on those labels is called a predictive model.

7. **Lasso and Ridge regression techniques belong to _____?**

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

Answer: D) Regularization

Lasso and Ridge regression are techniques used to regularize a model, which helps to prevent overfitting.

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

- A) Cross validation
- B) Regularization
- C) Kernel

- D) SMOTE

Answer: D) SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) is used to handle imbalanced datasets by creating synthetic samples for the minority class.

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

The AUC-ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (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

For a better model, the area under the ROC curve (AUC) should be larger, indicating better performance.

11. **Pick the feature extraction from below:**

- A) Construction bag of words from an email

- B) Apply PCA to project high dimensional data

- C) Removing stop words

- D) Forward selection

Answer: B) Apply PCA to project high dimensional data

PCA (Principal Component Analysis) is a method used for feature extraction by reducing the dimensionality of the data.

Q12: Multiple Answers

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.

Answers: A) We don’t have to choose the learning rate.

B) It becomes slow when number of features is very large.

The Normal Equation is a closed-form solution and does not require iteration or a learning rate but can become computationally expensive with a large number of features.

Q13-Q15: Subjective Questions

13. **Explain the term regularization?**

Answer: Regularization is a technique used in machine learning to prevent overfitting by adding a penalty to the loss function based on the complexity of the model. It helps to constrain or shrink the coefficients of the model, encouraging simpler models that are less likely to fit noise in the training data. Common regularization methods include Lasso (L1 regularization) and Ridge (L2 regularization).

14. **Which particular algorithms are used for regularization?**

Answer: Common algorithms used for regularization include:

- **Lasso Regression (L1 Regularization):** Adds a penalty equivalent to the absolute value of the magnitude of coefficients.

- **Ridge Regression (L2 Regularization):** Adds a penalty equivalent to the square of the magnitude of coefficients.

- **Elastic Net:** Combines both L1 and L2 regularization penalties.

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

Answer: In linear regression, the error term (or residual) represents the difference between the observed value of the dependent variable and the value predicted by the regression model. It captures the variation in the dependent variable that is not explained by the independent variables. Mathematically, it is the difference between the actual value and the predicted value: $\text{Error} = \text{Actual Value} - \text{Predicted Value}$. The goal of linear regression is to minimize the sum of the squared errors across all observations.