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.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Regularization is a technique used in machine learning and statistics to prevent overfitting and improve the generalization performance of a model. Overfitting occurs when a model learns to capture noise or random fluctuations in the training data, leading to poor performance on unseen data.

Regularization works by adding a penalty term to the objective function or loss function of a machine learning algorithm. This penalty term discourages overly complex models by penalizing large parameter values or model complexity. As a result, regularization helps to control the trade-off between fitting the training data well and maintaining simplicity in the model.

There are two common types of regularization techniques:

1. **L1 Regularization (Lasso Regression)**:

- In L1 regularization, the penalty term is proportional to the absolute values of the model coefficients.
- L1 regularization encourages sparsity in the model by shrinking some coefficients to exactly zero, effectively performing feature selection.
- L1 regularization is useful when dealing with high-dimensional datasets with many irrelevant features.

2.L2 Regularization (Ridge Regression):

In L2 regularization, the penalty term is proportional to the square of the model coefficients.

- L2 regularization shrinks the coefficients towards zero but does not typically reduce them to exactly zero.
- L2 regularization helps to reduce the magnitudes of the coefficients and improve the numerical stability of the model.
- It is particularly effective when there are correlated features in the dataset.

The strength of the regularization is controlled by a hyperparameter, usually denoted as

λ

λ (lambda). By adjusting the value of

λ

λ, the balance between fitting the training data and regularization can be fine-tuned. Cross-validation is often used to select the optimal value of

λ

λ that minimizes the generalization error of the model.

14. Which particular algorithms are used for regularization?

Several machine learning algorithms incorporate regularization techniques to prevent overfitting and improve model generalization. Some of the commonly used algorithms that employ regularization include:

1. **Linear Regression with Lasso (L1 regularization) or Ridge (L2 regularization)**:
    - In linear regression, Lasso regression (L1 regularization) and Ridge regression (L2 regularization) are used to penalize the magnitude of the coefficients to prevent overfitting.
2. **Logistic Regression with L1 or L2 regularization**:
    - Similar to linear regression, logistic regression can also use L1 or L2 regularization to penalize the coefficients and prevent overfitting in binary classification tasks.
3. **Support Vector Machines (SVM) with L2 regularization**:
    - Support Vector Machines (SVM) can be regularized using L2 regularization to control the margin width and prevent overfitting.
4. **Neural Networks with Dropout regularization**:
    - Dropout regularization is a technique used in neural networks to randomly drop a fraction of neurons during training, which helps prevent co-adaptation of neurons and reduces overfitting.
5. **ElasticNet Regression**:
    - ElasticNet regression is a combination of L1 and L2 regularization, which simultaneously applies both penalties to the coefficients. It is useful when there are correlated features and helps to select relevant features while reducing multicollinearity.
6. **Tree-based models with pruning**:
    - Decision trees and ensemble methods like Random Forest and Gradient Boosting Machines (GBM) can be regularized by pruning the trees or limiting their depth to prevent overfitting.

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

In the context of linear regression, the term "error" refers to the difference between the observed values of the dependent variable and the values predicted by the linear regression model. This difference is also known as the residual.

In a simple linear regression model, which involves only one independent variable, the relationship between the independent variable

X

X and the dependent variable

Y

Y is described by the equation:

$Y=\beta 0+\beta 1X+\epsilon$

Here:

- YY is the dependent variable.

- XX is the independent variable.
- $\beta 0 \beta 0$ is the intercept (the value of YY when XX is zero).
- $\beta 1 \beta 1$ is the slope (the change in YY for a unit change in XX).
- $\epsilon\epsilon$ is the error term.

The error term (ϵ) captures the variability in Ythat is not explained by the linear relationship with X. It represents the influence of all other factors, variables, or randomness that affect Y but are not included in the model. These factors could include measurement errors, unobserved variables, or random fluctuations.

The goal of linear regression is to estimate the coefficients $\beta 0$ and $\beta 1$ that minimize the sum of squared errors (residuals) between the observed values of Y and the values predicted by the model. This is typically achieved using the method of least squares, which finds the line that best fits the observed data by minimizing the total squared distance between the data points and the regression line.