# 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?
   1. Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

Ans:- D)Both And B

1. Which of the following statement is true about outliers in linear regression?
   1. Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

Ans:- A) Linear regression is sensitive to outliers

1. A line falls from left to right if a slope is ?
   1. Positive B) Negative

C) Zero D) Undefined

Ans:- B) Negative

1. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation

C) Both of them D) None of these

Ans:- B) correlation

1. Which of the following is the reason for over fitting condition?
   1. High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

Ans:-C) :Low bias and high variance

1. If output involves label then that model is called as:
   1. Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

Ans:- B) Predicted model

1. Lasso and Ridge regression techniques belong to ?
   1. Cross validation B) Removing outliers

C) SMOTE D) Regularization

Ans:- D) Regularization

1. To overcome with imbalance dataset which technique can be used?
   1. Cross validation B) Regularization

C) Kernel D) SMOTE

Ans:- D)SMOOTH

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses to make graph?
   1. TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

Ans:- A) TPR and FPR

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
   1. True B) False

ANS:- B) False

1. Pick the feature extraction from below:
   1. Construction bag of words from a email
   2. Apply PCA to project high dimensional data
   3. Removing stop words
   4. Forward selection

# Ans: B) Apply In Q12, more than one options are correct, choose all the correct options:

PCA to project high dimensional data.

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. We don’t have to choose the learning rate.
   2. It becomes slow when number of features is very large.
   3. We need to iterate.
   4. It does not make use of dependent variable.

Ana:- B) It become slow when number of feature is very large.

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

1. Explain the term regularization?

Ans: - Regularization is a technique used in machine learning and statistical modeling to prevent overfitting and improve the generalization performance of models. It involves introducing additional constraints or penalties to the model's objective function, aiming to strike a balance between fitting the training data well and avoiding excessive complexity.

The primary goal of regularization is to prevent overfitting, which occurs when a model learns the training data too well, capturing noise and random fluctuations rather than the underlying patterns or relationships. Overfitting leads to poor performance on new, unseen data, as the model becomes too specialized and fails to generalize.

Regularization methods typically add a penalty term to the loss function, which influences the model's optimization process. The penalty term acts as a control mechanism that discourages the model from assigning overly large weights or coefficients to features. By penalizing large weights, regularization encourages simpler models with more parsimonious representations, reducing the risk of overfitting.

There are different types of regularization techniques, including:

Regularization helps to control model complexity, prevent overfitting, and improve the model's

**L1 Regularization (Lasso):** It adds the sum of the absolute values of the coefficients as a penalty term. L1 regularization promotes sparsity by driving some coefficients to exactly zero, effectively performing feature selection.

**L2 Regularization (Ridge):** It adds the sum of the squared coefficients as a penalty term. L2 regularization encourages small but non-zero coefficients, effectively shrinking the coefficient values towards zero.

**Elastic Net Regularization:** It combines both L1 and L2 regularization, providing a balance between feature selection and coefficient shrinkage.

**Dropout:** A regularization technique commonly used in neural networks, where randomly selected neurons are temporarily "dropped out" during training, preventing them from contributing to the model's output. This encourages the network to learn more robust and less dependent representations.

1. Which particular algorithms are used for regularization?

Ans:-

Regularization techniques can be applied to various machine learning algorithms to control model complexity and prevent overfitting. Some of the common algorithms that can benefit from regularization include:

1. Linear Regression: Regularization techniques like Ridge Regression and Lasso Regression are commonly used to regularize linear regression models.
2. Logistic Regression: Regularization methods such as Ridge Regression and Lasso Regression can be applied to logistic regression models to prevent overfitting and improve performance.
3. Support Vector Machines (SVM): SVMs can benefit from regularization techniques, such as the use of L1 or L2 penalties, to control the complexity of the decision boundaries and improve generalization.
4. Neural Networks: Regularization techniques like L1 and L2 regularization, as well as dropout, can be applied to neural network models to prevent overfitting and improve their generalization performance.
5. Decision Trees: Regularization can be achieved in decision trees by applying techniques like pruning, which involves removing branches or nodes from the tree to reduce complexity and improve generalization.
6. Random Forests: Regularization techniques, such as limiting the maximum depth of individual trees or using feature subsampling, can be employed in random forests to control model complexity and enhance generalization.
7. Gradient Boosting: Gradient boosting algorithms like XGBoost and LightGBM often have regularization parameters that can be tuned to prevent overfitting and improve model performance.

These are just a few examples, and regularization techniques can be adapted to various other algorithms as well. The specific regularization methods applied depend on the algorithm and the problem at hand, with techniques like L1 regularization (Lasso), L2 regularization (Ridge), or a combination of both (Elastic Net) being commonly used.

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

Ans:- In the context of linear regression, the term "error" refers to the discrepancy or the difference between the actual observed values of the dependent variable and the predicted values generated by the linear regression model.

In linear regression, the goal is to find the best-fit line that minimizes the overall error or the distance between the predicted values and the actual values. This error is commonly known as the "residual" or "prediction error" and represents the deviation between the observed data and the model's predictions.

Mathematically, for each data point (xᵢ, yᵢ), where xᵢ is the input or independent variable and yᵢ is the corresponding observed output or dependent variable, the error or residual (εᵢ) can be calculated as:

εᵢ = yᵢ - ŷᵢ

Where:

* εᵢ represents the error or residual for the i-th data point.
* yᵢ is the actual observed value of the dependent variable for the i-th data point.
* ŷᵢ is the predicted value of the dependent variable generated by the linear regression model for the i-th data point.

The goal of linear regression is to find the line that minimizes the sum of squared errors (SSE) or the sum of squared residuals. This is commonly achieved through methods like Ordinary Least Squares (OLS), where the coefficients (slope and intercept) of the linear equation are estimated in a way that minimizes the sum of squared residuals.

By minimizing the error or the residual, the linear regression model aims to find the best-fit line that closely captures the relationship between the independent variable(s) and the dependent variable, allowing for accurate predictions and inference.

Top of Form

Regenerate response

Bottom of Form