Answers

Question no. 1

Answer: A) Least Square Error

Question no. 2

Answer: A) Linear regression is sensitive to outliers

Question no. 3

Answer: B) Negative

Question no. 4

Answer: B) Correlation

Question no. 5

Answer: C) Low bias and high variance

Question no. 6

Answer: B) Predictive model

Question no. 7

Answer: D) Regularization

Question no. 8

Answer: D) SMOTE

Question no. 9

Answer: A) TPR and FPR

Question no. 10

Answer: B) False

Question no. 11

Answer: A) Construction bag of words from a email

Question no. 12

Answers:

A) We don't have to choose the learning rate.

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

Question no. 13

Answer:

Regularization is a technique used in machine learning and statistical modeling to prevent overfitting, which occurs when a model learns not only the underlying pattern in the training data but also the noise and random fluctuations. This can lead to poor performance on new, unseen data. Regularization adds a penalty to the loss function that the model is trying to minimize, discouraging it from becoming too complex.

There are two main types of regularization:

L1 Regularization (Lasso): Adds a penalty equal to the absolute value of the magnitude of coefficients. This can lead to some coefficients being exactly zero, effectively performing feature selection.

L2 Regularization (Ridge): Adds a penalty equal to the square of the magnitude of coefficients. This tends to shrink the coefficients but does not set any to zero, helping to keep all features in the model but with reduced impact.

By adding these penalties, regularization helps to create simpler models that generalize better to new data.

Question no. 14

Answer: The main algorithms used for regularization in machine learning are:

1. Lasso Regression (L1 Regularization):

Adds the absolute value of the magnitude of the coefficients as a penalty term to the loss function.

Encourages sparsity, leading to some coefficients being exactly zero, which can effectively perform feature selection.

2. Ridge Regression (L2 Regularization):

Adds the square of the magnitude of the coefficients as a penalty term to the loss function.

Shrinks the coefficients but does not set any of them to zero, which helps to keep all features in the model while reducing their impact.

3. Elastic Net:

Combines both L1 and L2 regularization penalties.

Offers a compromise between Lasso and Ridge regression, providing both feature selection and coefficient shrinkage.

4. Regularized Logistic Regression:

Applies L1 or L2 regularization to logistic regression, which is useful for binary classification tasks.

5. Support Vector Machines (SVMs):

Can incorporate regularization terms to control the complexity of the decision boundary.

These algorithms help prevent overfitting by controlling the complexity of the model and ensuring better generalization to new, unseen data.

Question no. 15

Answer:

In a linear regression equation, the term **"Error"** (also known as residual) refers to the difference between the actual observed values and the values predicted by the model. This error quantifies how far off the predictions are from the actual data points.

Mathematically, for a given data point i:

Error I = $y_i - \hat{y}_i$

where:

- y_i is the actual observed value of the dependent variable for the i-th data point.
- $\hat{y}_i$ is the predicted value for the i-th data point, given by the linear regression model.

In a linear regression model, the predicted value, $\hat{y}_i$ is calculated using the regression equation:

$\hat{y}_i = \beta_0 + \beta_1 x_i + \epsilon$

where:

- β_0 is the intercept.
- β_1 is the slope.
- x_i is the independent variable for the i-th data point.
- ϵ is the error term.

The error I represents the residual for each data point, which measures how far off the model's prediction is from the actual observed value. These residuals are crucial for evaluating the model's performance and guiding the optimization process during training to minimize the overall error.

Types of Errors in Linear Regression:

1. Residual Sum of Squares (RSS).

2. Mean Squared Error (MSE).

3. Root Mean Squared Error (RMSE).

4. Mean Absolute Error (MAE).

In conclusion, the error in a linear regression equation is a critical component that helps in assessing and improving the model's predictive performance, Model Evaluation, Model Improvement, Bias-Variance Trade off.