Mathematical Explanation for Linear Regression Implementation

Introduction

This document provides a mathematical overview of a simple implementation of **Linear Regression**. Linear Regression aims to identify the best-fit line for a dataset by minimizing the error between predicted and actual values using the **least-squares regression** method.

Key Equations

The equation of a line is represented as:

$$y = mx + b$$

where:

- m: The slope of the line, representing the rate of change.
- b: The y-intercept, representing the value of y when $x = 0$.

Slope m

The slope m is computed as:

$$m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Where:

- x_i and y_i: Individual data points from X_{train} and y_{train}, respectively.
- $\bar{x}$: The mean of X_{train}.
- $\bar{y}$: The mean of y_{train}.

Intercept b

After calculating m, the y-intercept b is determined using the formula:

$$b = \bar{y} - m \cdot \bar{x}$$

Detailed Steps in the Code

Numerator for m

The numerator for m is calculated as:

$$\text{num} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Denominator for m

The denominator for m is computed as:

$$\text{den} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Final Formula for m

By combining the numerator and denominator, the slope m is:

$$m = \frac{\text{num}}{\text{den}}$$

Intercept b

The y-intercept b is calculated as:

$$b = \bar{y} - m \cdot \bar{x}$$

Prediction Formula

To predict outcomes using the trained model, the following equation is applied:

$$\hat{y} = m \cdot x + b$$

Where:

- $\hat{y}$: The predicted value.
- x: The input feature value.

Conclusion

This mathematical formulation forms the basis of the Linear Regression implementation. The process involves calculating the slope and intercept to define the best-fit line, which is then used to make predictions for unseen data.