Simple Linear Regression

Definition - Simple Linear Regression is a fundamental algorithm in machine learning used to model the relationship **between a single input feature and a continuous output**. It is often employed in predictive modeling tasks where you want to predict a target value based on a known input, assuming a linear relationship between the two.

Equation

$$y = B_0 + B_1.x$$

- y is the predicted output (target variable)
- x is the input feature (independent variable)
- B_0 is the intercept (the value of y when x = 0)
- B_1 is the slope (rate of change of y with respect to x)

Feature Scaling - Feature scaling is **usually not required** in Simple Linear Regression.

The model doesn't rely on the scale of the input feature because it calculates the best-fitting line using the coefficients (B_0, B_1), which are adjusted based on the feature's values.

However, in some cases where input values vary greatly in magnitude, scaling can improve numerical stability, especially when working with very large or very small numbers.

How It Works

- Using the training data, the algorithm finds the best-fit line using the Ordinary Least Squares method. This involves calculating the coefficients B_0 (intercept) and B_1 (slope) of the equation.
- The OLS minimizes the sum of squared residuals (the squared difference between actual and predicted values)
- Once the model is trained, it can be used to predict the output (dependent variable y) for new or unseen input data. For any given input x, the model uses the equation to compute predictions.

Ordinary Least Squares

In Machine Learning, **Ordinary Least Squares (OLS)** is one of the optimization techniques used to minimize the error between predicted and actual values. OLS is particularly used to estimate the best-fit line (model parameters) by minimizing the sum of squared residuals (the differences between the predicted and observed values).

The loss function (or cost function) that OLS tries to minimize is the **Mean Squared Error (MSE)**, which is common in machine learning models for regression. The loss function can be expressed as:

$$J(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- y_i is the actual value for the i-th data point
- y^_i is the predicted value for the i-th data point
- n is the total number of data points
- J(B_0, B_1) is the loss function that needs to be minimized

OLS works by adjusting the model parameters B_0 and B_1 for the most accurate values to minimize the squared difference between actual and predicted values. In machine learning terms, this is called **minimizing the cost function**.

Without using OLS (or a similar optimization technique), we would be essentially guessing the slope and intercept which is leading to inaccurate predictions.