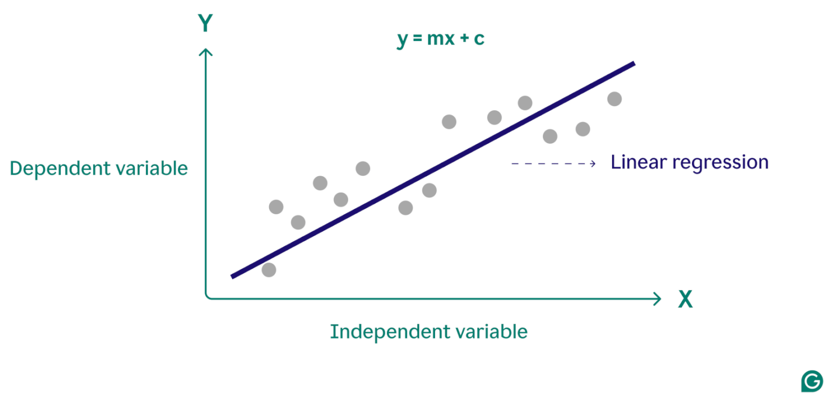
Linear Regression

Introduction

Linear regression is one of the most fundamental algorithms in supervised learning, used to model the relationship between input features and a continuous output. It assumes a linear relationship between independent variables and the dependent variable, making it one of the simplest yet most widely used predictive models. Despite its simplicity, linear regression provides powerful insights into data patterns and is a baseline for more complex models.



Mathematical Foundation

1. Linear Model

Given a feature vector $x = (x_1, x_2, \ldots, x_n)$, weights $w = (w_1, w_2, \ldots, w_n)$, and bias b, linear regression models the relationship as:

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b = w^T x + b \tag{1}$$

where $\hat{y}$ represents the predicted output.

2. Cost Function

To measure how well the model's predictions match the actual values, we use the Mean Squared Error cost function:

$$C(w,b) = \frac{1}{2N}\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2 \tag{2}$$

Where:

- N is the number of training samples.
- $\hat{y}^{(i)}$ is the predicted output for the i-th sample.
- $y^{(i)}$ is the actual target value for the i-th sample.

3. Optimization via Gradient Descent

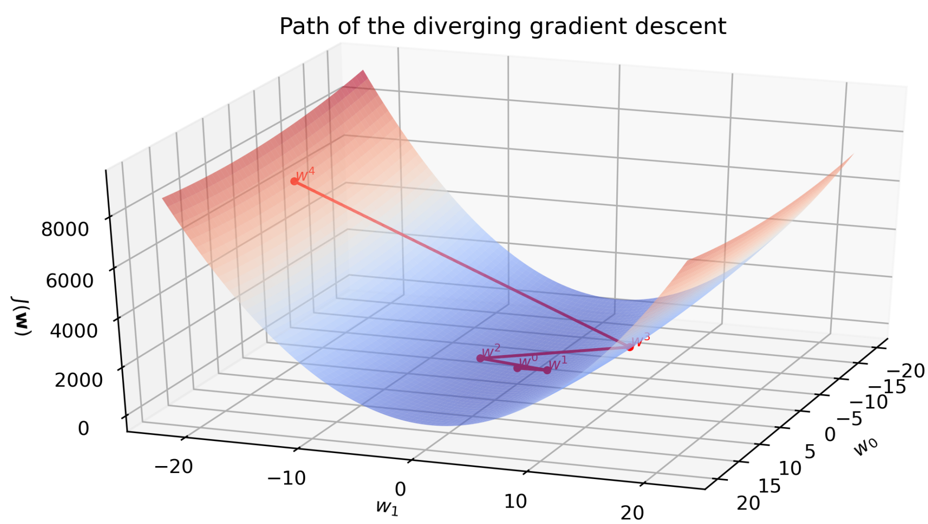
The goal is to minimize the cost function by adjusting the model parameters. The gradient descent algorithm updates the weights and bias iteratively:

$$w_j \leftarrow w_j - \alpha\frac{\partial C}{\partial w_j}, \quad j = 1, 2, \ldots, n \tag{3}$$

$$b \leftarrow b - \alpha\frac{\partial C}{\partial b} \tag{4}$$

Where α is the learning rate. The partial derivatives are computed as:

$$\frac{\partial C}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})x_j^{(i)}, \quad \frac{\partial C}{\partial b} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)}) \tag{5}$$



Algorithm Steps

1. Initialize weights and bias to small random values or zeros.
2. Repeat until convergence:
   (a) Compute predictions $\hat{y}$ using the current weights and bias.
   (b) Calculate the gradient of the cost function.
   (c) Update the weights and bias using gradient descent.
3. Stop when the changes in parameters are sufficiently small or a set number of iterations is reached.

Key Characteristics

- Type: Regression.
- Model: Linear relationship between inputs and outputs.
- Cost Function: Mean Squared Error.
- Optimization: Uses gradient descent.

Strengths

- Simple to implement and easy to interpret.
- Computationally efficient, even for large datasets.
- Provides valuable insights into relationships between variables through learned coefficients.

Weaknesses

- Assumes a linear relationship between features and the target variable, which may not always be valid.
- Sensitive to outliers, as they can disproportionately impact the model's predictions.
- Struggles to capture complex, non-linear patterns in data.

Applications

Linear regression is widely used in various fields due to its interpretability and efficiency:

- Economics: Used for forecasting economic indicators such as GDP growth and inflation rates.
- Real Estate: Helps predict housing prices based on features like location, size, and amenities.
- Healthcare: Assists in predicting patient outcomes based on medical variables.
- Marketing: Used for sales forecasting and understanding the impact of advertising campaigns.

Conclusion

Linear regression is one of the most fundamental and widely used models in machine learning. While its simplicity makes it easy to interpret and apply, its assumptions about linearity can limit its effectiveness for more complex tasks. Despite this, it remains a powerful tool for analyzing data and serves as a strong baseline model for regression problems.