Linear Regression

David Robinson

Simple Linear Regression involves one independent variable and one dependent variable.

$$y = \theta_0 + \theta_1 x_1$$

Multiple Linear Regression involves multiple independent variables and one dependent variable.

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

Gradient Descent

Gradient descent is an iterative optimization algorithm used to minimize the cost function $J(\theta)$ by updating the model parameters θ.

Update Rule

For each parameter θ_j,

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Optimization Process

1. Start with random initial values for θ.
2. Compute the cost function $J(\theta)$.
3. Update θ using the update rule.
4. Repeat until the cost function converges.

Normal Equation

The **Normal Equation** is an analytical method to solve for the parameters θ without using iterative methods like gradient descent.

$$\theta = (X^T X)^{-1} X^T y$$

Gradient Descent vs. Normal Equation

Gradient Descent

1. You need to choose learning rate.
2. It is an iterative approach that updates parameters until convergence.
3. Works for almost every loss function.
4. Efficient for datasets with large numbers of features, as it does not require matrix inversion.

Normal Equation

1. No need to choose learning rate.
2. It is an analytical method and provides a solution in one shot.
3. You need to compute $(X^T X)^{-1}$, an $n \times n$ matrix, where n is the number of features.
4. Time complexity is $O(n^3)$ so it is computationally expensive for a large n.
5. Only works for least squares loss.