Gradient Descent in Linear Regression

Linear Regression Model

The hypothesis function for simple linear regression is:

$$\hat{y} = wx + b$$

where

- w is the weight (slope)
- b is the bias (intercept)

The goal is to find the best values of w and b that minimize the prediction error.

Cost Function

The error is measured using the Mean Squared Error (MSE):

$$J(w, b) = \frac{1}{n} \sum_{i=1}^{n} (wx_i + b - y_i)^2$$

Gradient Descent is used to minimize this cost function.

What is Gradient Descent?

Gradient descent is an iterative optimization algorithm that updates the model parameters in the direction of the negative gradient of the cost function.

$$\theta := \theta - \alpha \nabla J(\theta)$$

For linear regression, the parameters are $\theta = \{w, b\}$ and α is the learning rate.

Parameter Updates

The partial derivatives of the cost function are:

$$\frac{\partial J}{\partial w} = \frac{2}{n} \sum_{i=1}^{n} (wx_i + b - y_i)x_i$$

$$\frac{\partial J}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} (wx_i + b - y_i)$$

Gradient descent updates are:

$$w := w - \alpha \frac{\partial J}{\partial w}$$

$$b := b - \alpha \frac{\partial J}{\partial b}$$

Intuition

The cost function $J(w, b)$ forms a convex bowl-shaped surface. Gradient descent follows these steps:

1. Start with initial values of w and b
2. Compute the gradient (slope of the cost)
3. Move "downhill" by subtracting the gradient
4. Repeat until convergence

This process leads to the global minimum, giving the best-fitting line.