Introduction to Linear Regression (with Math and Python)

Evan Misshula

June 9, 2025

Why Learn Linear Regression?

- Linear regression is the most basic and interpretable predictive model.
- It helps understand relationships between variables.
- Applications: housing prices, student performance, stock trends.

The Linear Model

Let $\hat{y} = w_0 + w_1 x_1 + \cdots + w_p x_p$. In vector form: $\hat{y} = \mathbf{x}^\top \mathbf{w}$.

Terms and Concepts

- y: dependent variable (response)
- $\mathbf{x}$: feature vector (independent variables)
- $\mathbf{w}$: coefficients (weights)
- w_0: intercept
- $\hat{y}$: predicted value
- Loss Function

Mean Squared Error (MSE)

$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ Minimizing this loss gives the best linear fit.

- Analytical Solution

Derivation

We want to minimize the MSE:

$$L(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) \tag{1}$$

Take the gradient with respect to $\mathbf{w}$:

$$\nabla_{\mathbf{w}} L = -2\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) \tag{2}$$

Set the gradient to zero:

$$\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) = 0 \tag{3}$$

$$\Rightarrow \mathbf{X}^\top\mathbf{y} = \mathbf{X}^\top\mathbf{X}\mathbf{w} \tag{4}$$

$$\Rightarrow \hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \tag{5}$$