Linear regression through Optimization

Chris Cornwell

September 11, 2025

Outline

Outline

Measuring the fitness of linear models

Least Squares loss function

On given data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_P, y_P)$, the measure that is commonly used for how well a linear regression model, $f_{b,\mathbf{w}}(\mathbf{x}) = b + \mathbf{x}^T\mathbf{w}$, fits the data is the Least Squares loss (cost) function.

- *This is P times the Mean Squared Error.*

Following notation from textbook, this loss function is

$$g(b, \mathbf{w}) = \sum_{p=1}^{P} \left(f_{b,\mathbf{w}}(\mathbf{x}_p) - y_p\right)^2 .$$

Meaning of Least Squares loss

For $1 \leq p \leq P$, since $f_{b,\mathbf{w}}(\mathbf{x}_p) = \hat{y}_p$, the quantity $|f_{b,\mathbf{w}}(\mathbf{x}_p) - y_p|$ is the vertical distance from (x_p, y_p) to the point predicted by the linear model, $(x_p, \hat{y}_p)$.

Additionally, the length of the vector $\mathbf{y} - \hat{\mathbf{y}}$ (which is the distance from $\mathbf{y}$ to the point determined by $\tilde{\mathbf{w}}$, in the column space of our feature matrix) is equal to

$$\sqrt{(\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + \ldots + (\hat{y}_P - y_P)^2} = \sqrt{g(b, \mathbf{w})}.$$

We see that minimizing $g(b, \mathbf{w})$ is the same as minimizing that distance, which will give us the $\hat{\mathbf{y}}$ in the column space that makes $\mathbf{y} - \hat{\mathbf{y}}$ be orthogonal to the column space.

Minimizing the Least Squares loss

The data $\{(\mathbf{x}_p, y_p)\}_{p=1}^{P}$ is fixed. How do we solve the problem

$$\underset{b,\mathbf{w}}{\text{minimize}} \; g(b, \mathbf{w})?$$

We can use methods from calculus, specifically the first order condition – that we want all partial derivatives equal to zero.

- Note, what are the variables of the function g? They are the parameters $b, w_1, w_2, \ldots, w_N$.

Minimizing the Least Squares loss

The data $\{(\mathbf{x}_p, y_p)\}_{p=1}^{P}$ is fixed. How do we solve the problem

$$\underset{b,\mathbf{w}}{\text{minimize}} \; g(b,\mathbf{w})?$$

We can use methods from calculus, specifically the first order condition – that we want all partial derivatives equal to zero.

- Note, what are the variables of the function g? They are the parameters $b, w_1, w_2, \ldots, w_N$.

Next: a review of calculus minimization techniques.