Ordinary Least Squares Estimation

1. Simple Linear Regression

Simple Linear Regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{for } i = 1, 2, \ldots, n,$$

where y_i is the dependent variable, x_i is the independent variable, β_0 and β_1 are the regression coefficients, and ϵ_i is the error term.

Least Squares Estimator minimizes the Sum of Squared Errors

$$SSE = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Taking partial derivatives of SSE with respect to β_0 and β_1 and setting them to zero

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0,$$

$$\frac{\partial SSE}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0.$$

Rearranging gives the normal equations

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i,$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2.$$

The solutions are

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}.$$

2. Multiple Linear Regression

Multiple Linear Regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is an $n \times 1$ vector, $\mathbf{X}$ is an $n \times p$ matrix, and $\boldsymbol{\beta}$ is a $p \times 1$ vector.

Least Squares Estimator minimizes

$$J(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Taking the derivative of $J(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and setting it to zero

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0.$$

This gives the normal equation

$$\mathbf{X}^\top \mathbf{y} = \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}.$$

Solving for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$