Least Squares Estimator in Simple and Multiple Linear Regression

Uhm Yoonhee

1 Simple Linear Regression

단순 선형 회귀에서 우리는 종속 변수 y_i와 독립 변수 x_i 간의 관계를 다음과 같은 식으로 모델링합니다:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

여기서:

- y_i : 종속 변수
- x_i : 독립 변수
- β_0 : 절편 (독립 변수 $x = 0$일 때 y의 예상 값)
- β_1 : 기울기 (독립 변수 x에 대해 y의 변화율)
- ϵ_i : 오차 항

우리는 이제 SSE(제곱 오차 합)를 최소화하는 β_0와 β_1의 값을 찾으려고 합니다. 각 데이터 포인트의 오차는 관측값 y_i와 예측값 $\hat{y}_i = \beta_0 + \beta_1 x_i$ 간의 차이입니다. 따라서,

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

우리는 이 합을 최소화하기 위해 β_0와 β_1에 대해 부분 미분을 하고 이를 0으로 설정하여 정상 방정식을 구합니다.

먼저, β_0에 대해 미분합니다:

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

이는 다음과 같이 단순화됩니다:

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

β_0에 대해 풀면:

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

다음으로, β_1에 대해 미분합니다:

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

이는 다음과 같이 단순화됩니다:

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

여기서 $\beta_0 = \bar{y} - \beta_1 \bar{x}$를 대입하면:

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

β_1에 대해 풀면:

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

따라서, β_0와 β_1에 대한 최소 제곱 추정값은 다음과 같습니다:

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

2 Multiple Linear Regression

다중 선형 회귀에서, 우리는 종속 변수 y_i와 여러 독립 변수 $x_{1i}, x_{2i}, \ldots, x_{ki}$ 간의 관계를 모델링합니다:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

여기서:

- y_i : 종속 변수
- $x_{1i}, x_{2i}, \ldots, x_{ki}$: 독립 변수들
- β_0 : 절편
- $\beta_1, \beta_2, \ldots, \beta_k$: 독립 변수들의 계수
- ϵ_i : 오차 항

우리는 최소 제곱법을 사용하여 $\beta_0, \beta_1, \ldots, \beta_k$를 추정하려고 합니다. 각 관측값에 대한 오차는 관측값과 예측값 간의 차이입니다. 따라서:

$$J(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ji} \right)^2$$

여기서, β_0는 절편이고, $\beta_1, \beta_2, \ldots, \beta_k$는 독립 변수 $x_1, x_2, \ldots, x_k$의 계수들입니다. 우리는 이 값들을 최소 제곱법으로 추정하려고 합니다.

행렬 형태로 다중 선형 회귀 모델을 나타내면 다음과 같습니다:

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

여기서:

- $\mathbf{y}$: 종속 변수의 관측값들을 나타내는 $n \times 1$ 열 벡터
- $\mathbf{X}$: 독립 변수들의 값을 나타내는 $n \times (k+1)$ 행렬
- $\boldsymbol{\beta}$: $\beta_0, \beta_1, \ldots, \beta_k$를 나타내는 $(k+1) \times 1$ 열 벡터
- $\boldsymbol{\epsilon}$: 오차 항을 나타내는 $n \times 1$ 열 벡터

제곱 오차 합을 행렬 형태로 나타내면 다음과 같습니다:

$$\begin{aligned} J(\boldsymbol{\beta}) &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

다음 단계는 $\boldsymbol{\beta}$에 대해 SSE를 최소화하는 것입니다. SSE에 대해 $\boldsymbol{\beta}$를 미분하면 다음과 같습니다:

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

이를 0으로 설정하여 $\boldsymbol{\beta}$가 SSE를 최소화하는 값을 찾습니다:

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

이를 단순화하면:

$$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

따라서 $\boldsymbol{\beta}$에 대한 해는 다음과 같습니다:

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