Ordinary Least Squares Estimation (OLS)

1 단순 선형 회귀 (Simple Linear Regression)

단순 선형 회귀는 하나의 독립 변수 x와 종속 변수 y 간의 관계를 예측하는 모델입니다. 모델은 다음과 같습니다:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{(i = 1, 2, ..., n)}$$

여기서: - y_i는 종속 변수(판매량 등), - x_i는 독립 변수(광고비용 등), - β_0는 상수항 (intercept), - β_1은 기울기 (slope), - ϵ_i는 오차 항입니다.

1.1 Sum of Squared Errors (SSE)

단순 선형 회귀에서의 SSE는 다음과 같이 정의됩니다:

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

1.2 SSE 최소화를 위한 미분

SSE를 최소화하려면, SSE를 각 회귀 계수에 대해 미분하여 0으로 설정해야 합니다.

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

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

위의 방정식들을 풀면, 회귀 계수 β_0와 β_1는 다음과 같이 계산됩니다.

1.3 회귀 계수 추정

회귀 계수 β_1와 β_0는 각각 다음과 같이 구할 수 있습니다:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

여기서 $\bar{x}$는 x 값들의 평균, $\bar{y}$는 y 값들의 평균입니다.

2 다중 선형 회귀 (Multiple Linear Regression)

다중 선형 회귀는 여러 개의 독립 변수 $x_1, x_2, \ldots, x_p$를 사용하여 종속 변수 y를 예측하는 모델입니다. 모델은 다음과 같이 정의됩니다:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \quad \text{(i = 1, 2, ..., n)}$$

2.1 행렬 형태로 표현

다중 선형 회귀 모델은 다음과 같이 행렬 형태로 표현할 수 있습니다:

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

여기서: - $\mathbf{y}$는 $n \times 1$ 크기의 종속 변수 벡터, - X는 $n \times (p+1)$ 크기의 디자인 행렬로, 각 행은 독립 변수와 상수 항을 포함한 벡터입니다. - $\beta = [\beta_0, \beta_1, \ldots, \beta_p]^T$는 회귀 계수 벡터입니다.

2.2 Sum of Squared Errors (SSE)

다중 선형 회귀에서의 SSE는 다음과 같이 정의됩니다:

$$SSE = \|\mathbf{y} - X\beta\|^2 = (\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta)$$

2.3 SSE 최소화를 위한 미분

SSE를 최소화하려면 β에 대해 미분하여 0으로 놓고 풀어야 합니다. 이를 미분한 식은 다음과 같습니다:

$$\frac{\partial SSE}{\partial \beta} = -2X^T(\mathbf{y} - X\beta) = 0$$

2.4 Normal Equation 도출

위 미분 방정식을 풀면, 최적의 회귀 계수 β를 구할 수 있는 Normal Equation은 다음과 같습니다:

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

따라서, 회귀 계수 β의 추정 값은 다음과 같이 구할 수 있습니다:

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

이 식은 **최소제곱법(OLS)**의 핵심이며, 다중 선형 회귀에서 최적의 회귀 계수 벡터를 계산하는 데 사용됩니다.