YBIGTA 신입세션 박하늘

1. Simple Linear Regression

단순 선형 회귀에서는 y와 x의 선형 관계를 다음과 같이 가정한다.

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

- y_i: 종속변수
- x_i: 독립변수
- β_0: 절편
- β_1: 기울기
- ϵ_i: 오차항

목표는 오차 제곱합 SSE를 최소화하는 β_0와 β_1를 추정하는 것이다.

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

해당 SSE를 β_0 과 β_1 로 각각 편미분을 진행하면 다음과 같다.

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

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

다음은 β_0에 대한 방정식을 구하는 과정이다.

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

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

좌변에 β_0만 남기고 정리를 하면

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

평균을 활용하면 다음과 같이 표현할 수 있다.

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

해당 방정식을 β_1에 대한 방정식에 대입하여 정리하면

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

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

즉 다음의 값을 유도할 수 있다.

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

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

2. Multiple Linear Regression

다중 선형 회귀를 다음과 같이 표현할 수 있다.

$$y = X\beta + \epsilon$$

- y: $n \times 1$ 벡터 (종속 변수)
- X: $n \times p$ 행렬 (설명 변수, predictor variables)
- β: $p \times 1$ 벡터 (회귀 계수)
- ϵ: $n \times 1$ 벡터 (오차항)

SSE를 정의하고 전개하면 다음과 같다.

$$J(\beta) = \|y - X\beta\|^2 = (y - X\beta)^\top (y - X\beta)$$

$$J(\beta) = y^\top y - 2y^\top X\beta + \beta^\top X^\top X\beta$$

J에 대해 β 로 편미분을 계산하면 다음과 같다.

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

이를 0으로 두고 β를 풀면:

$$-2X^\top y + 2X^\top X\beta = 0$$

$$X^\top X\beta = X^\top y$$

해당 식은 정규방정식이며, 이를 통해 β를 추정하면 다음과 같다.

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