

예제 1) 입력 데이터가 다음 표와 같이 주어진다.

| x1 | x2 | y |
|----|----|---|
| 2 | 10 | 9 |

선형 회귀 모델이 다음과 같이 주어진다.

$$h_{\theta}(x) = \theta^T x \quad s.t. \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

오차함수 E 가 Sum of Squared Error일 때, $\frac{\partial E}{\partial \theta_1}$ 을 구하여라.

sol)

$$\begin{aligned} E &= \frac{1}{2} \sum_{i=0}^I (h_{\theta}(x^{(i)}) - y^{(i)})^2 = \frac{1}{2} (h_{\theta}(x) - y)^2 \quad (\because I = 1) \\ &= \frac{1}{2} (\theta^T x - y)^2 = \frac{1}{2} (\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 - y)^2 \end{aligned}$$

이므로, $\frac{\partial E}{\partial \theta_1}$ 에 대입하여 계산하면

$$\begin{aligned} \frac{\partial E}{\partial \theta_1} &= \frac{\partial}{\partial \theta_1} \frac{1}{2} (\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 - y)^2 \\ &= (\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 - y) x_1 \\ &= (3 \times 1 + 2 \times 2 + 1 \times 10 - 9) \times 2 \\ &= 16 \end{aligned}$$

rmk)

이 예제는 BGD의 update rule

$$\theta_n \leftarrow \theta_n - \alpha \frac{\partial E}{\partial \theta_n}$$

의 계산 과정 중 $n=1$ 일 때의 gradient를 구하는 부분에 해당한다.

실제로 θ_n 값을 업데이트 해줄 때는 구한 gradient에 학습률 α 를 곱하여 각각 뺄셈을 수행한다.