

Question. 4-07

target function $y = \theta_1^* x + \theta_0^*$ 의 Linear Regression을 위해

Prediction 모델을 $\hat{y} = \theta_1 x + \theta_0$ 로, Loss를 Square Error로 설정하였다.

이때, 다음 질문에 답하시오.

1) $\underset{(\theta_1, \theta_0)}{\operatorname{argmin}} \mathcal{L}$ 를 만족시키는 θ_1, θ_0 이 θ_1^*, θ_0^* 임을 보이시오.

2) initial $\vec{\theta} = (\theta_1, \theta_0)$ 가 $(-1, -1)$ 로 주어졌을 때, 다음과 같은 target function에서도 학습이 잘 이루어짐을 보이시오.

이때, learning rate = 0.1, data sample은 $x = 1$ 인 경우에 대하여 1 iteration동안 학습을 진행하시오.

$$y = -x + 1$$

$$y = 3x - 1$$

$$y = 3x + 5$$

1) $\mathcal{L} = (y - \hat{y})^2 = (\theta_1^* x + \theta_0^* - (\theta_1 x + \theta_0))^2$ 이고

$$(\theta_1^* x + \theta_0^* - (\theta_1 x + \theta_0))^2 \geq 0 \text{ 이므로}$$

$$\theta_1^* x + \theta_0^* - \theta_1 x - \theta_0 = 0 \text{ 일때 } \mathcal{L} = (\theta_1^* x + \theta_0^* - (\theta_1^* x + \theta_0^*))^2 = 0 \text{ 이된다.}$$

$$\text{그러므로 } \theta_1 = \theta_1^*, \theta_0 = \theta_0^* \text{에서 } \underset{\theta_1, \theta_0}{\operatorname{argmin}} \mathcal{L} \text{를 만족시킨다.}$$

2) target function이 $y = -x + 1$ 일때

$$x=1\text{일때의 } \mathcal{L}_{\text{loss}} = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (0 - (-1) \cdot 1 - (-1))^2 = 4$$

$$\theta_1 := \theta_1 + 2\alpha x (y - \theta_1 x - \theta_0) = -1 + 2 \cdot (0.1) \cdot 1 \cdot 2 = -0.6$$

$$\theta_0 := \theta_0 + 2\alpha (y - \theta_1 x - \theta_0) = -1 + 2 \cdot (0.1) \cdot 2 = -0.6$$

\therefore 1 iteration 이후 updated된 $(\theta_1, \theta_0) = (-0.6, -0.6)$ 이고

$$\text{이때 } x=1\text{일때의 } \mathcal{L}_{\text{loss}} = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (0 - (-0.6) \cdot 1 - (-0.6))^2 = 1.44 \text{ 이다.}$$

Loss가 4에서 1.44로 감소한 것으로 미루어볼때 학습이 잘 이루어지고 있음을 알 수 있다.

target function이 $y = 3x - 1$ 일 때

$$x=1\text{일때의 } \mathcal{L}_{\text{loss}} = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (2 - (-1) \cdot 1 - (-1))^2 = 16$$

$$\theta_1 := \theta_1 + 2\alpha x (y - \theta_1 x - \theta_0) = -1 + 2 \cdot (0.1) \cdot 1 \cdot 4 = -0.2$$

$$\theta_0 := \theta_0 + 2\alpha (y - \theta_1 x - \theta_0) = -1 + 2 \cdot (0.1) \cdot 4 = -0.2$$

\therefore 1 iteration 이후 updated된 $(\theta_1, \theta_0) = (-0.2, -0.2)$ 이고

$$\text{이때 } x=1\text{일때의 } \mathcal{L}_{\text{loss}} = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (2 - (-0.2) \cdot 1 - (-0.2))^2 = 5.76 \text{ 이다.}$$

Loss가 16에서 5.76으로 감소한 것으로 미루어볼때 학습이 잘 이루어지고 있음을 알 수 있다.

target function이 $y = 3x + 5$ 일 때

$$x=10\%의 Loss = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (8 - (-1) \cdot 1 - (-1))^2 = 100$$

$$\theta_1 := \theta_1 + 2\alpha x (y - \theta_1 x - \theta_0) = -1 + 2 \cdot (0.1) \cdot 1 \cdot 10 = 1$$

$$\theta_0 := \theta_0 + 2\alpha (y - \theta_1 x - \theta_0) = -1 + 2 \cdot 0.1 \cdot 10 = 1$$

$\therefore 1$ iteration 이후 update된 $(\theta_1, \theta_0) = (1, 1)$ 이고

$$\text{이때 } x=10\%의 Loss = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2 = (8 - 1 \cdot 1 - 1)^2 = 36 \text{ 이다.}$$

Loss가 100에서 36으로 감소한 것으로 미루어보아 학습이 잘 이루어지고 있음을 알 수 있다.