Question. 4-01

Dataset이 다음과 같이 주어졌다고 하자,

$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)})\} = \{(1,3), (3,7)\}$

위의 Dataset은 $y = 2x + 1$에서부터 만들어졌기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$가 $\theta_1 = -1, \theta_0 = -1$일 때, 다음 질문에 답하시오.

1) 각 Data sample에 대해 Square error를 이용한 Loss를 구하시오.

2) Loss function이 θ_1, θ_0에 대해 각각에 대해 몇 차 식인지 구하고 Convexity를 확인하시오.

3) 각 Data sample에 대해 $\vec{\theta} = (\theta_1, \theta_0) = (1,0)$일 때와 $\vec{\theta} = (\theta_1, \theta_0) = (2,1)$일 때의 Loss를 각각 구하고 비교하시오.

4) θ_1이 +1 커진 경우와 θ_0가 +1 커진 경우 중 어느 경우가 Loss가 더 많이 감소하는지 비교하시오.

1) $\hat{y} = \theta_1 x + \theta_0$ 이므로 $\mathcal{L}(\theta_1, \theta_0) = (y - \hat{y})^2 = (y - \theta_1 x - \theta_0)^2$이다.

$(\theta_1, \theta_0) = (-1, -1)$에서 $(x, y) = (1, 3)$이라면 $\mathcal{L}(\theta_1, \theta_0) = (3 - (-1) \cdot 1 - (-1))^2 = 25$

$(x, y) = (3, 7)$이라면 $\mathcal{L}(\theta_1, \theta_0) = (7 - (-1) \cdot 3 - (-1))^2 = 121$ 이 된다.

이때 $\theta_1 x$는 x에 비례하여 커지고, θ_0는 x와 무관하다.

즉, θ_1은 bias가 없는 Single-variate linear regression과 마찬가지로 Loss를 x에 비례하게 변화시키고 θ_0는 x와 무관하게 Loss를 변화시킨다.

2) $\mathcal{L} = (y - \hat{y})^2 = y^2 - 2y\hat{y} + \hat{y}^2 = y^2 - 2y(\theta_1 x + \theta_0) + \theta_1^2 x^2 + 2\theta_1\theta_0 x + \theta_0^2$ 이다.

이를 θ_1에 대하여 정리하면 $x^2\theta_1^2 + (2\theta_0 x - 2xy)\theta_1 + (y^2 - 2y\theta_0 + \theta_0^2)$으로 θ_1에 대한 이차식이 된다.

이를 다시 θ_0에 대하여 정리하면 $\theta_0^2 + (2\theta_1 x - 2y)\theta_0 + (y^2 - 2xy\theta_1 + x^2\theta_1^2)$으로 θ_0에 대한 이차식이 된다.

이때 $(x, y) = (1, 3)$을 대입하면 $(3 - \theta_1 - \theta_0)^2 = (\theta_1 + (\theta_0 - 3))^2$

$(x, y) = (3, 7)$을 대입하면 $(7 - 3\theta_1 - \theta_0)^2 = (3\theta_1 + (\theta_0 - 7))^2$ 으로 θ_1과 θ_0에 대하여 Convex하다.

3) $\mathcal{L}(\theta_1, \theta_0) = (y - (\theta_1 x + \theta_0))^2$ 이므로 $(\theta_1, \theta_0) = (1, 0)$일 때 Loss는

$(x, y) = (1, 3)$ 이라면 $(3 - (1 + 0))^2 = 4$ 이고 $(x, y) = (3, 7)$ 이라면 $(7 - (3 + 0))^2 = 16$ 이다.

즉 $(\theta_1, \theta_0) = (-1, -1)$ 일때보다 $\vec{\theta}^*$에 가까워졌을 때,

Loss가 $25 \to 4$, $121 \to 16$ 으로 감소한 모습을 확인할 수 있다.

$(\theta_1, \theta_0) = (2, 1)$일 때 Loss는 $(x, y) = (1, 3)$이라면 $(3 - 2 - 1)^2 = 0$

$(x, y) = (3, 7)$이라면 $(7 - 6 - 1)^2 = 0$ 이다.

즉, θ_1과 θ_0가 $\vec{\theta}^*$에 도달하였을 때 Loss는 x와 무관하게 0임을 알 수 있다.

4) $(\theta_1, \theta_0) = (-1, -1)$에서 θ_1이 +1 커진 경우와 θ_0가 +1 커진 경우는 각각 θ_1^*, θ_0^*에 가까워진 경우를 의미한다.
주어진 Dataset의 x가 모두 $x > 0$ 인 조건에서 $L(\theta_1, \theta_0) = (y - (\theta_1 \cdot x + \theta_0 \cdot 1))^2$ 이므로

$x > 1$ 인 경우 θ_1이 θ_1^*에 더 가까워졌을 때 Loss가 더 많이 감소한다.

반대로 $0 < x < 1$ 인 경우 θ_0가 θ_0^*에 더 가까워진 경우 Loss가 더 많이 감소한다.