

Question.3-02

Dataset이 다음과 같이 주어졌다고 하자.

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)})\} = \{(1, 2), (2, 4), (3, 6)\}$$

위의 dataset에 대해 다음의 물음에 답하시오.

1) model을 $\hat{y} = x$ 로 설정한다면, 각 data sample들에 대한 square loss를 구하고 서로 비교하시오.

2) square error loss를 임의의 data point $(x^{(i)}, y^{(i)})$ 에 대한 algebraic equation으로 표현하고,

$(x^{(i)}, y^{(i)})$ 에 대한 square error loss를 최소로 만드는 θ 를 구하시오.

단, θ 는 변수로 사용하여 model은 $\hat{y} = \theta x$ 로 설정하시오.

3) 2)에서 구한 algebraic equation이 θ 에 대한 몇차인지 구하고, convexity를 말하시오.

1) 주어진 predictor에 $x^{(1)}=1, x^{(2)}=2, x^{(3)}=3$ 에 대한 prediction $\hat{y}^{(1)}, \hat{y}^{(2)}, \hat{y}^{(3)}$ 은 다음과 같다.

$$\hat{y}^{(1)} = \hat{y}|_{x=1} = 1 \quad \hat{y}^{(2)} = \hat{y}|_{x=2} = 2 \quad \hat{y}^{(3)} = \hat{y}|_{x=3} = 3$$

따라서 각각의 data sample들에 대한 loss는 $L^{(1)}, L^{(2)}, L^{(3)}$ 가 된다.

$$\begin{aligned} L^{(1)} &= (y^{(1)} - \hat{y}^{(1)})^2 = (2 - 1)^2 = 1 \\ L^{(2)} &= (y^{(2)} - \hat{y}^{(2)})^2 = (4 - 2)^2 = 4 \\ L^{(3)} &= (y^{(3)} - \hat{y}^{(3)})^2 = (6 - 3)^2 = 9 \end{aligned}$$

이므로 x 값이 α 배 커지면 loss는 α^2 만큼 커지는 것을 알 수 있다.

2) 변수 θ 와 임의의 data point $(x^{(i)}, y^{(i)})$ 에 대해 loss를 구해보면

$$L^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2 = (y^{(i)} - \theta \cdot x^{(i)})^2$$

이 된다. 그리고

$$y^{(i)} - \theta \cdot x^{(i)} = 0$$

실제 $L^{(i)}$ 는 최소가 된다. 즉, θ 의 값을 $x^{(i)}$ 에 대해

$$\theta = \frac{y^{(i)}}{x^{(i)}}$$

실제 $L^{(i)}$ 는 최소가 된다. 그리고 θ 는 dataset의 만들어진 $y = \alpha x$ 와 동일할 때를 의미한다.

3) 2)에서 구한 $L^{(i)}$ 에 대한 algebraic equation은 상수인 θ 에 대해

2차식이므로 convex quadratic function이 된다. graph를 그려면

오른쪽과 같다.

따라서 $L^{(i)}(\theta)$ 는 θ 에 대한 2차식의 convex한 특성을 가진다.

