

Question.3-12

대부분의 dataset에는 noise가 섞여있다. $y = 2x$ 에서 이상적으로 만들어진 Dataset이 다음과 같을 때

$$D = \{(1,2), (2,4), (3,6)\}$$

noise에 의해 실제로 수집한 dataset이 다음과 같이 왜곡되었다고 가정하자.

$$D = \{(1,0.5), (2,5), (3,5)\}$$

θ 가 2이었을때, 즉 predictor가 target function일때 각 data sample에 의해 θ 가 update되는 값들을 구하고,

noise가 학습에 미치는 영향을 분석하시오.

주어진 상황에서 noise가 없다면 dataset

$$D = \{(1,2), (2,4), (3,6)\}$$

이러한 model의 prediction은 각각 2, 4, 6이므로 정확하게 예측한다. 즉 학습이 끝난 상태이다.

하지만 θ 가 2인 상황에서 왜곡된 dataset의 data sample들에 대한 loss와 θ 의 변화는 다음과 같다.

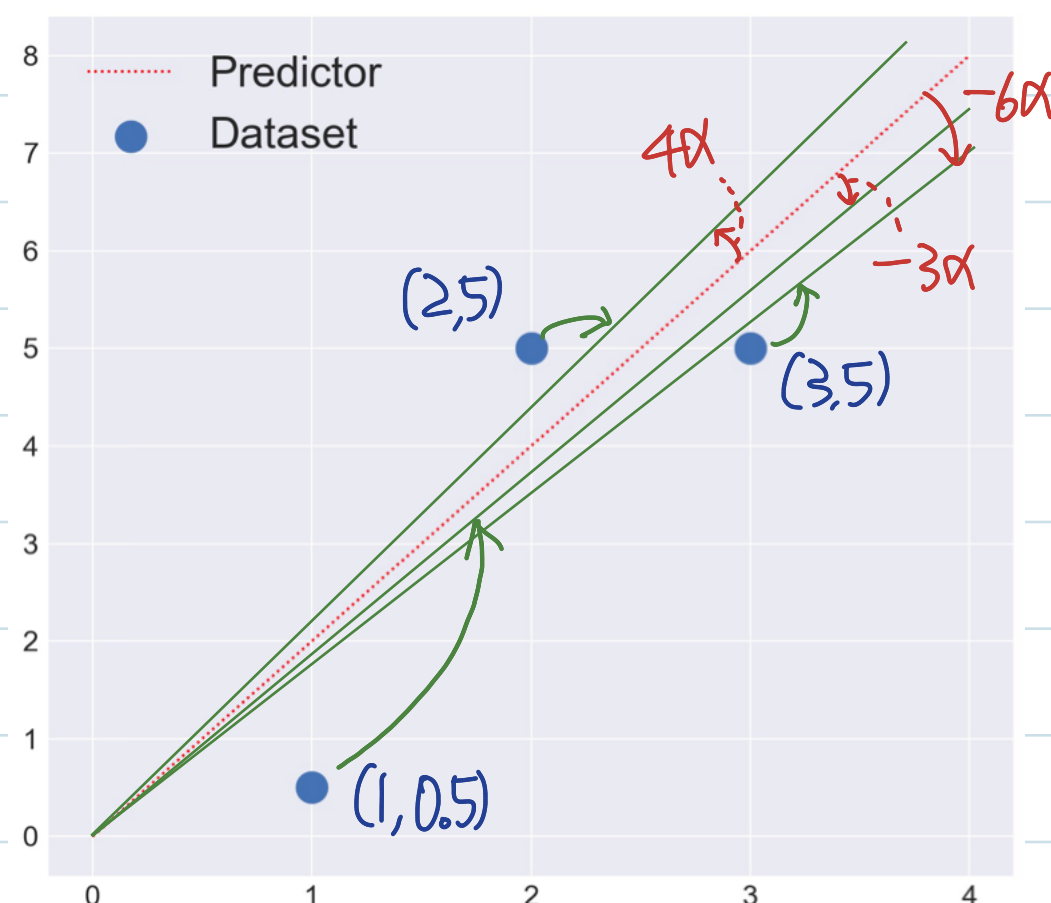
사래의 연산에서 왜곡된 data point의 영향을 보아위해 실제 θ 는 update시키지 않는다.

data point (1, 0.5) $L = (y - \theta x)^2 = (0.5 - 2 \cdot 1)^2 = 2.25$ $\Delta\theta = 2\alpha x(y - \theta x) = 2\alpha \cdot 1(0.5 - 2 \cdot 1) = -3\alpha$

data point (2, 5) $L = (y - \theta x)^2 = (5 - 2 \cdot 2)^2 = 1$ $\Delta\theta = 2\alpha x(y - \theta x) = 2\alpha \cdot 2(5 - 2 \cdot 2) = 4\alpha$

data point (3, 5) $L = (y - \theta x)^2 = (5 - 2 \cdot 3)^2 = 1$ $\Delta\theta = 2\alpha x(y - \theta x) = 2\alpha \cdot 3(5 - 2 \cdot 3) = -6\alpha$

이 $\Delta\theta$ 를 graph에 표현하면



위와 같다. 위에서 알 수 있듯이 각 data point는 target $\theta(\theta^*)$ 인 2에서 멀어질수록 θ 가 학습된다.

2보다 실제 이상적인 y보다 큰 값을 왜곡된 data sample들은 θ^* 보다 큰 θ 를 만들고,

이상적인 y보다 작은 값을 왜곡된 data sample들은 실제 θ^* 보다 작은 θ 를 만든다.

따라서 data가 왜곡되면 predictor를 왜곡된 쪽으로 학습시키게 된다.