Question. 4-04

$|x| \geq 1$인 Dataset $\mathcal{D}$가 다음과 같이 주어졌다.

$$\mathcal{D} = \{(1,7),(5,15),(-1,3),(-5,-5)\}$$

Dataset을 $y = 2x + 5$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}=(\theta_1,\theta_0)$ =(1,1)이고, learning rate α =0.1로 주어졌을 때 다음 질문에 답하시오.

1) Square error를 Loss function으로 사용하였을 때, $\frac{\partial \mathcal{L}}{\partial \theta_1}$과 $\frac{\partial \mathcal{L}}{\partial \theta_0}$을 구하시오.

2) θ_1과 θ_0에 대한 gradient vector의 norm을 구하시오.

3) θ_1 ,θ_0의 Gradient Descent Method를 구하시오.

4) 각 Data sample에 의한 θ_1 ,θ_0의 Update 양을 구하시오.

5) 4)에서의 결과를 통해 $x^{(i)}$가 γ 배 되었을 때, θ_1, θ_0가 Update 되는 양의 변화를 구하시오.

1) $L = (y-\hat{y}) = (y-\theta_1 x-\theta_0)^2$ 이므로

$$\frac{\partial L}{\partial \theta_1} = 2(y-\theta_1 x-\theta_0)(-x) = -2x(y-\theta_1 x-\theta_0) = -2x(y-\hat{y})$$

$$\frac{\partial L}{\partial \theta_0} = 2(y-\theta_1 x-\theta_0)\cdot(-1) = -2(y-\theta_1 x-\theta_0) = -2(y-\hat{y})$$

2) L2 norm $= \sqrt{\left(\frac{\partial L}{\partial \theta_1}\right)^2 + \left(\frac{\partial L}{\partial \theta_0}\right)^2} = \sqrt{\{-2x(y-\hat{y})\}^2 + \{-2(y-\hat{y})\}^2}$

$= \sqrt{(4x^2+4)(y-\hat{y})^2} = 2(y-\hat{y})^2\sqrt{x^2+1}$

3) $\theta_1 := \theta_1 - \alpha\frac{\partial L}{\partial \theta_1} = \theta_1 - \alpha\cdot(-2x)(y-\hat{y}) = \theta_1 + 2\alpha x(y-\hat{y}) = \theta_1 + 2\alpha x(y-\theta_1 x-\theta_0)$

$\theta_0 := \theta_0 - \alpha\frac{\partial L}{\partial \theta_0} = \theta_0 - \alpha(-2)(y-\hat{y}) = \theta_0 + 2\alpha(y-\hat{y}) = \theta_0 + 2\alpha(y-\theta_1 x-\theta_0)$

4) $(x,y)=(1,7)$ 일 때 $\Delta\theta_1 = 2\alpha x(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot 1\cdot(7-1-1) = 1$

$\Delta\theta_0 = 2\alpha(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot(7-1-1) = 1$

즉, $x=1$ 일때 θ_1과 θ_0는 동일하게 update 된다.

$(x,y) = (5,15)$ 일 때 $\Delta\theta_1 = 2\alpha x(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot 5\cdot(15-5-1) = 9$

$\Delta\theta_0 = 2\alpha(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot(15-6) = 1.8$

즉, $x=5$ 일때 θ_1의 update양이 θ_0의 update 양의 5배이다.

$(x,y)=(-1,3)$ 일때 $\Delta\theta_1 = 2\alpha x(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot(-1)\cdot(3+1-1) = -0.6$

$\Delta\theta_0 = 2\alpha(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot(3) = 0.6$

즉, $x=-1$ 일 때 θ_1과 θ_0 크기는 같으나 반대 방향으로 학습된다.

$(x,y)=(-5,-5)$일 때 $\Delta\theta_1 = 2\alpha x(y-\theta_1 x-\theta_0) = 2\cdot 0.1\cdot(-5)(-5+5-1) = 1$

$\Delta\theta_0 = 2\alpha(y-\theta_1 x-\theta_0) = 2\cdot 0.1(-5+5-1) = -0.2$

즉, $x=-5$일 때 θ_1의 update 양은 θ_0의 update 양의 -5배이다.

5) θ_1의 업데이트 양은 $2\alpha x(y-\theta_1 x-\theta_0)$이고 θ_0의 업데이트 양은 $2\alpha(y-\theta_1 x-\theta_0)$
이므로 x가 r배 되었을 때 θ_1은 $2\alpha r x(y-\theta_1 r x-\theta_0)$ 만큼 변화하고
θ_0는 $2\alpha(y-\theta_1 r x-\theta_0)$ 만큼 변화한다.

즉, x가 r배가 되면 θ_1과 θ_0의 업데이트 양의 차이도 r배가 된다.