Question. 4-09

2개의 Dataset $\mathcal{D}_1$과 $\mathcal{D}_2$가 다음과 같이 주어졌다.

$$\mathcal{D}_1 = \{(-0.5,2.5),(0.5,3.5)\}$$

$$\mathcal{D}_2 = \{(-1,2),(1,4)\}$$

$\mathcal{D}_1, \mathcal{D}_2$ 모두 $y = x + 3$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 $\alpha=0.1$로 주어졌을 때 다음 질문에 답하시오.

1) Loss에 대한 Update Equation을 이용하여 1번의 epoch동안 $\mathcal{D}_1$, $\mathcal{D}_2$ 각각 $\vec{\theta}$의 변화를 구하시오.

2) $\mathcal{D}_1$과 $\mathcal{D}_2$ 중 어느 Dataset이 $\vec{\theta}$를 target $\vec{\theta}^*$에 더 가깝게 하였는지 판단하고 이유를 설명하시오.

1) $\mathcal{L}=(y-\hat{y})^2=(y-\theta_1x+\theta_0)^2$이므로 $\frac{\partial\mathcal{L}}{\partial\theta_1}=2(y-\theta_1x+\theta_0)\cdot(-x)=-2x(y-\theta_1x-\theta_0)$

$\frac{\partial\mathcal{L}}{\partial\theta_0}=2(y-\theta_1x+\theta_0)(-1)=-2(y-\theta_1x-\theta_0)$이다.

$\therefore\ \theta_1:=\theta_1-\alpha\cdot\frac{\partial\mathcal{L}}{\partial\theta_1}=\theta_1+2\alpha x(y-\theta_1x-\theta_0)$

$\theta_0:=\theta_0-\alpha\frac{\partial\mathcal{L}}{\partial\theta_0}=\theta_0+2\alpha(y-\theta_1x-\theta_0)$이다.

$\mathcal{D}_1=\{(-0.5,2.5),(0.5,3.5)\}$으로 학습을 진행하는 경우

$(x,y)=(-0.5,2.5)$에 의해 $\theta_1:=-1+2\cdot(0.1)(-0.5)(2.5-(-1)(-0.5)+1)=-1.3$

$\theta_0:=-1+2\cdot(0.1)(2.5-0.5+1)=-0.4$

$(x,y)=(0.5,3.5)$에 의해 $\theta_1:=-1.3+0.2(0.5)(3.5-(-1.3)(0.5)+0.4)=-0.845$

$\theta_0:=-0.4+0.2(3.5+(1.3)(0.5)+0.4)=0.51$

$\therefore(\theta_1,\theta_0)=(-0.84,0.51)$으로 update 되고

이때 target $\vec{\theta}^*$ 까지의 L2-norm은 $\sqrt{(1-(-0.84))^2+(3-0.51)^2}\doteq 3.10$

$\mathcal{D}_2=\{(-1,2),(1,4)\}$으로 학습을 진행하는 경우

$(x,y)=(-1,2)$에 의해 $\theta_1:=-1+2\cdot(0.1)\cdot(-1)\cdot(2-(-1)(-1)-(-1))=-1.4$

$\theta_0:=-1+2(0.1)(2-1+1)=-0.6$

$\therefore(\theta_1,\theta_0)=(-1.4,-0.6)$으로 update 되고

$(x,y)=(1,4)$에 의해 $\theta_1:=-1.4+2\cdot(0.1)\cdot1\cdot(4-(-1.4)(1)-(-0.6))=-0.2$

$\theta_0:=-0.6+2\cdot(0.1)\cdot(4+2)=0.6$

$\therefore(\theta_1,\theta_0)=(-0.2,0.6)$이 된다.

이때 target $\vec{\theta}^*$ 까지의 L2-norm은 $\sqrt{(1-(-0.2))^2+(3-0.6)^2}\doteq 2.68$ 이다.

2) $\mathcal{D}_1=\{(-0.5,2.5),(0.5,3.5)\}$에 의해 update된 $\vec{\theta}$의 $\vec{\theta}^*$까지의 norm $=3.10$

$\mathcal{D}_2=\{(-1,2),(1,4)\}$에 의해 update된 $\vec{\theta}$의 $\vec{\theta}^*$까지의 norm $=2.68$로

$\mathcal{D}_2$에서 $\vec{\theta}$가 $\mathcal{D}_1$에서 보다 약 0.42만큼 더 $\vec{\theta}^*$ 근접했다.

이는 $0<|x|<1$ 인 $\mathcal{D}_1$의 Data point에 의해 x의 영향에 비례한 θ_1의 update 속도가 느려졌기 때문이다.