Question. 4-10

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

$$\mathcal{D}_1 = \{(-3,0), (3,6)\}$$

$$\mathcal{D}_2 = \{(-10,-7), (10,13)\}$$

Question. 4-09과 마찬가지로 $\mathcal{D}_1, \mathcal{D}_2$ 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 α $=0.1$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) Q.4-9와 Q.4-10의 결과를 토대로 Q.4-10의 $\mathcal{D}_1, \mathcal{D}_2$에서 $\vec{\theta}$가 발산하는 이유를 설명하시오.

1) $D_1 = \{(-3,0), (3,6)\}$으로 학습을 진행하는 경우

$(x,y) = (-3,0)$에 의해 $\theta_1 := -1 + 2\cdot(0.1)\cdot(-3)(0-(-1)(-3)+1) = 0.2$

$\theta_0 := -1 + 2\cdot(0.1)(-2) = -1.4$

$(x,y) = (3,6)$에 의해 $\theta_1 := 0.2 + 2(0.1)(3)(6-(0.2)\cdot 3 + 1.4) = 4.28$

$\theta_0 := -1.4 + 2(0.1)(6-0.6+1.4) = -0.04$

$\therefore (\theta_1, \theta_0) = (4.28, -0.04)$로 update되고

이때 target $\vec{\theta}^*$까지의 L2-norm은 $\sqrt{(1-4.28)^2+(3-(-0.04))^2} \fallingdotseq 4.47$

$D_2 = \{(-10,-7), (10,13)\}$으로 학습을 진행하는 경우

$(x,y) = (-10,-7)$에 의해 $\theta_1 := -1 + 2\cdot(0.1)\cdot(-10)\cdot(-7-(-1)(-10)+1) = 31$

$\theta_0 := -1 + 2\cdot(0.1)\cdot(-7-10+1) = -4.2$

$(x,y) = (10,13)$에 의해 $\theta_1 := 31 + 2\cdot(0.1)\cdot(10)\cdot(13-(31)\cdot(10)+4.2) = -554.6$

$\theta_0 := -4.2 + 2\cdot(0.1)\cdot(13-310+4.2) = -62.76$

$\therefore (\theta_1, \theta_0) = (-554.6, -62.76)$으로 update되고

이때 target $\vec{\theta}^*$까지의 L2-norm은 $\sqrt{(1-(-554.6))^2+(3-(-62.76))^2} \fallingdotseq 559.48$

2) Question. 4-9에선 $|\alpha|<1$이었기 때문에 $|x|$가 증가할수록 학습속도가 빨라지는 것을 확인할 수 있었다.

그러나 Question. 4-10에선 $|\alpha|>1$의 조건을 같은 learning rate와 initial $\vec{\theta}$를 사용하더라도

θ_0의 update양이 $2\alpha x(y-\theta_1 x-\theta_0)$이기 때문에 θ_1이 크게 발산한다.

또한 θ_0의 update양은 $2\alpha(y-\theta_1 x-\theta_0)$로 θ_1보다는 적지만 x의 영향을 받기 때문에 θ_0도 발산하게된 것이다.