

Question.3-05

다음과 같이 하나의 data sample만 가지고 있는 Dataset이 주어졌다.

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

이때 다음 질문들에 답하시오.

단 Question.3-04와 마찬가지로 prediction model은 $\hat{y} = \theta x$ 를 사용하고, initial θ 는 1로 설정한다.

- 1) learning rate이 0.8일때, 3 iteration 동안 update되는 θ 들을 구하시오.
- 2) learning rate이 1.1일때, 3 iteration 동안 update되는 θ 들을 구하시오.
- 3) Question.3-04에서 구한 learning rate가 0.1일때 θ 의 변화와, 위의 2), 3)의 결과들에서의 θ 의 변화들을 $\mathcal{L}(\theta)$ 와 함께 graph로 나타내시오.

그리고 learning이 커지면 어떤 문제점이 발생하는지 설명하시오.

먼저 θ 를 update시키는 Question.3-04에서 구한대로 다음과 같다.

$$\theta := \theta + \alpha(4 - 2\theta)$$

- 1) learning rate이 0.8일 때 3번의 iteration 동안 θ 의 변화를 계산하면 다음과 같다.

$$1^{\text{st}} \text{ iteration: } \theta := 1 + 0.8(4 - 2) = 2.6$$

$$2^{\text{nd}} \text{ iteration: } \theta := 2.6 + 0.8(4 - 2 \times 2.6) = 1.64$$

$$3^{\text{rd}} \text{ iteration: } \theta := 1.64 + 0.8(4 - 2 \times 1.64) = 2.22$$

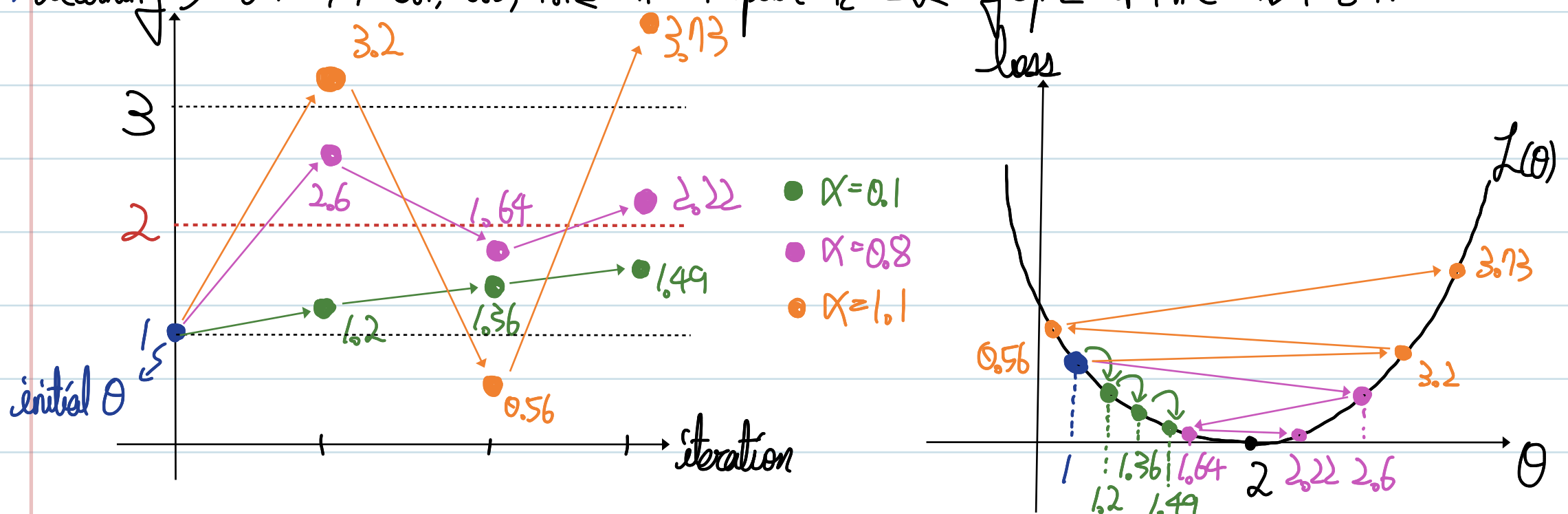
- 2) learning rate이 1.1일 때 3번의 iteration 동안 θ 의 변화를 계산하면 다음과 같다.

$$1^{\text{st}} \text{ iteration: } \theta := 1 + 1.1(4 - 2) = 3.2$$

$$2^{\text{nd}} \text{ iteration: } \theta := 3.2 + 1.1(4 - 2 \times 3.2) = 0.56$$

$$3^{\text{rd}} \text{ iteration: } \theta := 0.56 + 1.1(4 - 2 \times 0.56) = 3.73$$

- 3) learning rate이 각각 0.1, 0.8, 1.1일 때 θ 가 update되는 모습을 graph로 나타내면 다음과 같다.



위에서 알 수 있듯이 θ 가 충분히 작을 때 안정적으로 θ^* 에 가까워지지만 $\alpha=0.8$ 처럼 θ 가 어느정도 커지면 θ 의 update가 불안정해지기 시작한다 $\alpha=1.1$ 처럼 설정이상 커지게 되면 발산하게 된다. 따라서 θ 는 충분히 작게 설정해야 안정적인 학습이 가능하다.