

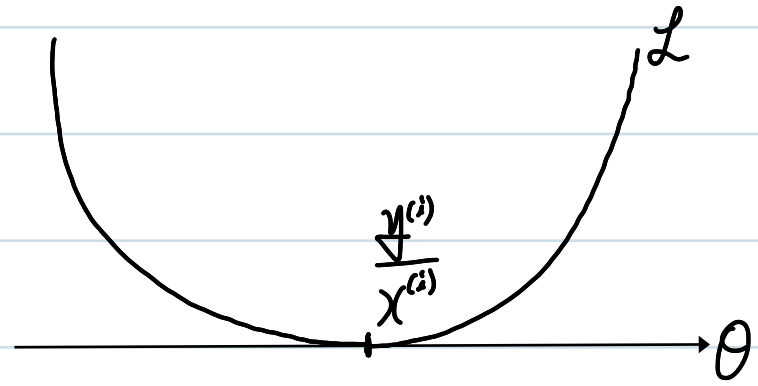
### Question.3-11

Linear regression을 통해 predictor를 학습시킬 때, iteration이 지날 때마다 loss의 감소량과

$\theta$ 가 update되는 양이 줄어드는 이유를 설명하시오.

먼저  $\hat{y} = \theta x$ 를 학습시킬 때 loss function은 다음과 같다.

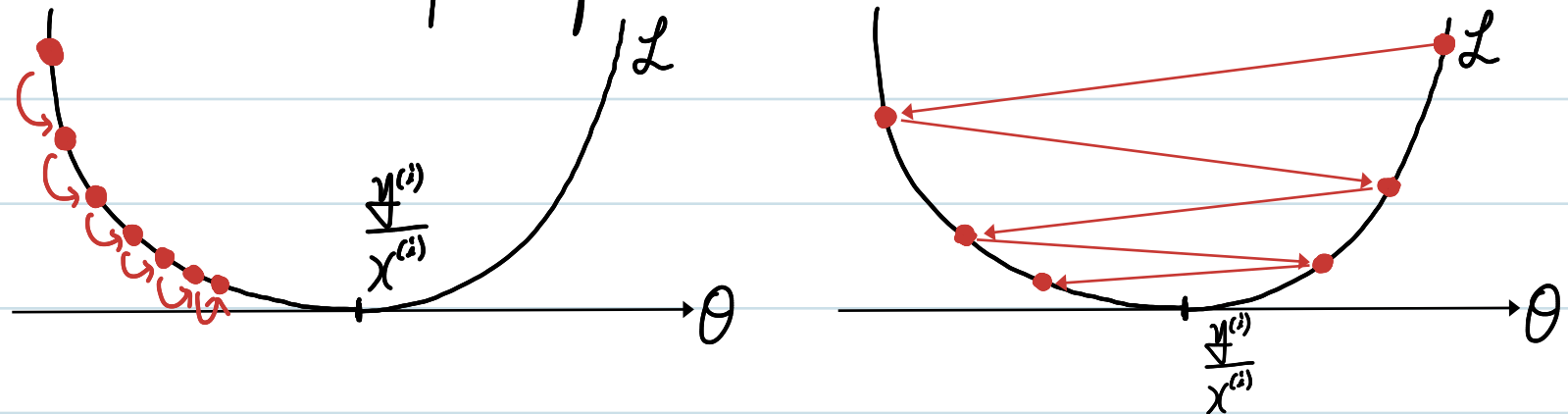
$$L = (y^{(i)} - \hat{y}^{(i)})^2 = (y^{(i)} - \theta x^{(i)})^2$$



2222 learning rate이 충분히 작다면

$$\theta := \theta - \alpha \cdot \frac{\partial L}{\partial \theta}$$

를 통해 initial theta가 어디에 있든지 optimal point로 접근한다. 이를 2개의 그림으로 표현하면 다음과 두 개가 된다.



위의 2개에서 알 수 있듯이  $\theta$ 가 optimal point에 접근할수록  $\left| \frac{\partial L}{\partial \theta} \right|$ 가 줄어들게 된다. 이때 따라  $\Delta \theta$ 는  $|2\alpha x^{(i)} (y^{(i)} - \hat{y}^{(i)})|$

로 줄어들게 된다. 즉  $\theta$ 가 update 될수록  $\Delta \theta$ 는 작아지고, 이때 따라 predictor의 계수도 점점 천천히 변하게 되며 결국 loss 감소 속도  $\frac{\Delta L(\theta)}{\text{iteration}}$ 도 감소하게 된다.