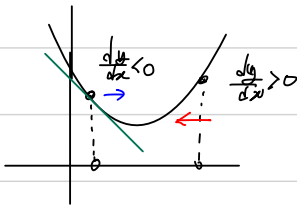


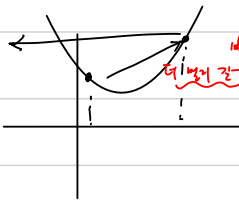
CH3 - 1 - 5

Gradient Descent Method

$\arg \min_x J$: $J(\theta)$ 를 최소 만드는 θ value를 찾는 것.



θ 가 커지는 방향.
 기울기가 음수 \rightarrow θ 값이 커져야..
 기울기가 양수 \rightarrow θ 값이 작아져야..
 θ 값이 커지는 방향.



$x := x - \frac{dJ}{dx}$
 반복해서가 좋아...
 더 많이 갈 수도 있다...
 그래서 gradient 이
 상하 값으로 알려준다.
 $\frac{dJ}{dx}$ $x := x - \frac{dJ}{dx}$

Learning Rate : 학습 속도를 조절하여 수렴에너
 조절하여 알려준다.
 hyper parameter : 'we' should control.

Loss

$$J^{(1)}(\theta) = (y^{(1)} - (\theta x^{(1)}))^2$$

$$\theta = \theta - \alpha \frac{\partial J^{(1)}(\theta)}{\partial \theta}$$

= 학습
작업

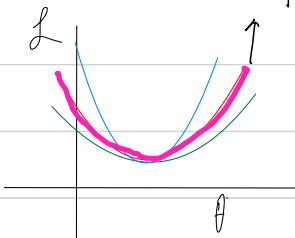
Cost

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (y^{(i)} - (\theta x^{(i)}))^2$$

$$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

작업

작업이 2번. 비용함수는 이것의 average 'Cost'이다.



2번의 파라 (Learning rate)
 hyperparameter
 조절을 해주는 것이 중요함