BIG DATA ANALYTICS

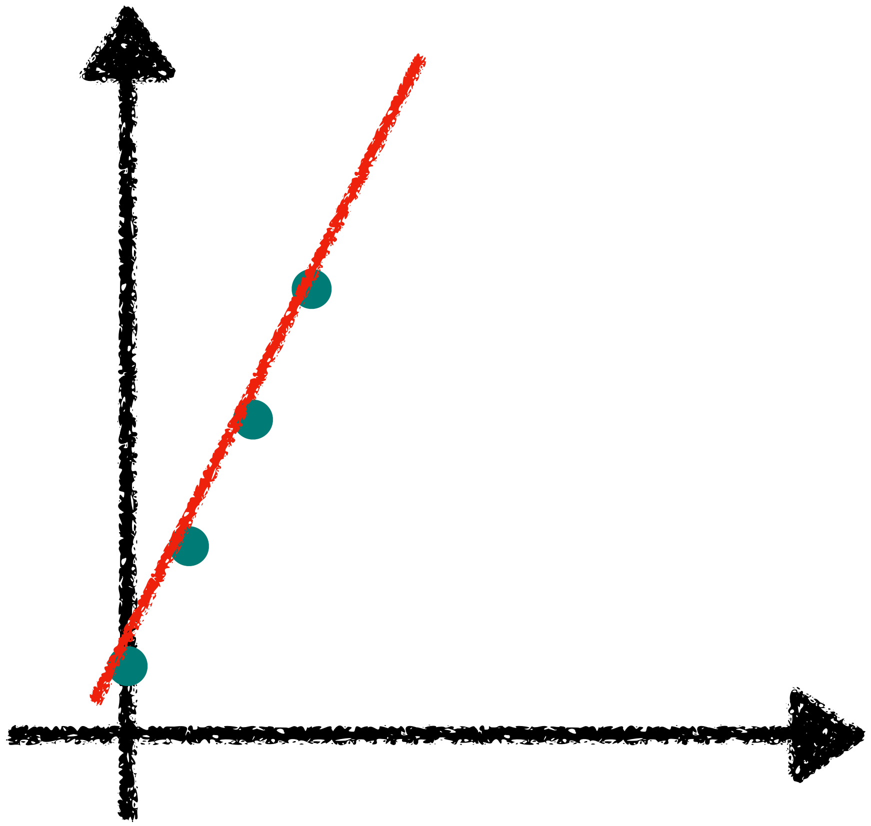
WEEK-08 | Supervised Learning - Regression

Yonsei University
Jungwon Seo

기본 원리

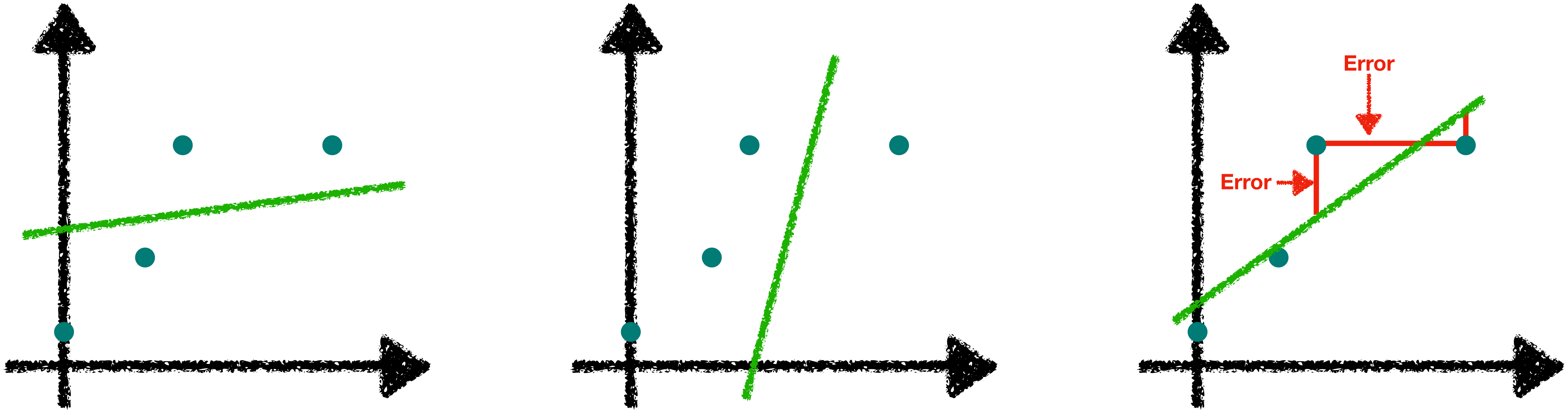
$$y = 3x+1$$

x	y
1	4
2	7
3	10
4	?



Regression

- Error를 최소화 할 수 있는 방정식(Equation)을 찾기



Linear Regression

- 단순 선형회귀
  - $y = wx + b$

- 다중 선형회귀 (n개의 feature)
  - $y = w_1x_1 + w_2x_2 + \cdots \ + w_nx_n + b$

- 가설 (Hypothesis)
  - 변수간의 관계를 유추하기 위해 수학적으로 나타낸 식
  - $H(x) = wx + b$

Loss function (손실함수)

- 실제값과 예측값과의 차이를 측정 할 수 있는 함수

- Regression Loss Functions
  - Mean Squared Error Loss
  - Mean Squared Logarithmic Error Loss
  - Mean Absolute Error Loss

- Binary Classification Loss Functions
  - Binary Cross-Entropy
  - Hinge Loss
  - Squared Hinge Loss

- Multi-Class Classification Loss Functions
  - Multi-Class Cross-Entropy Loss
  - Sparse Multiclass Cross-Entropy Loss
  - Kullback Leibler Divergence Loss

Optimizer

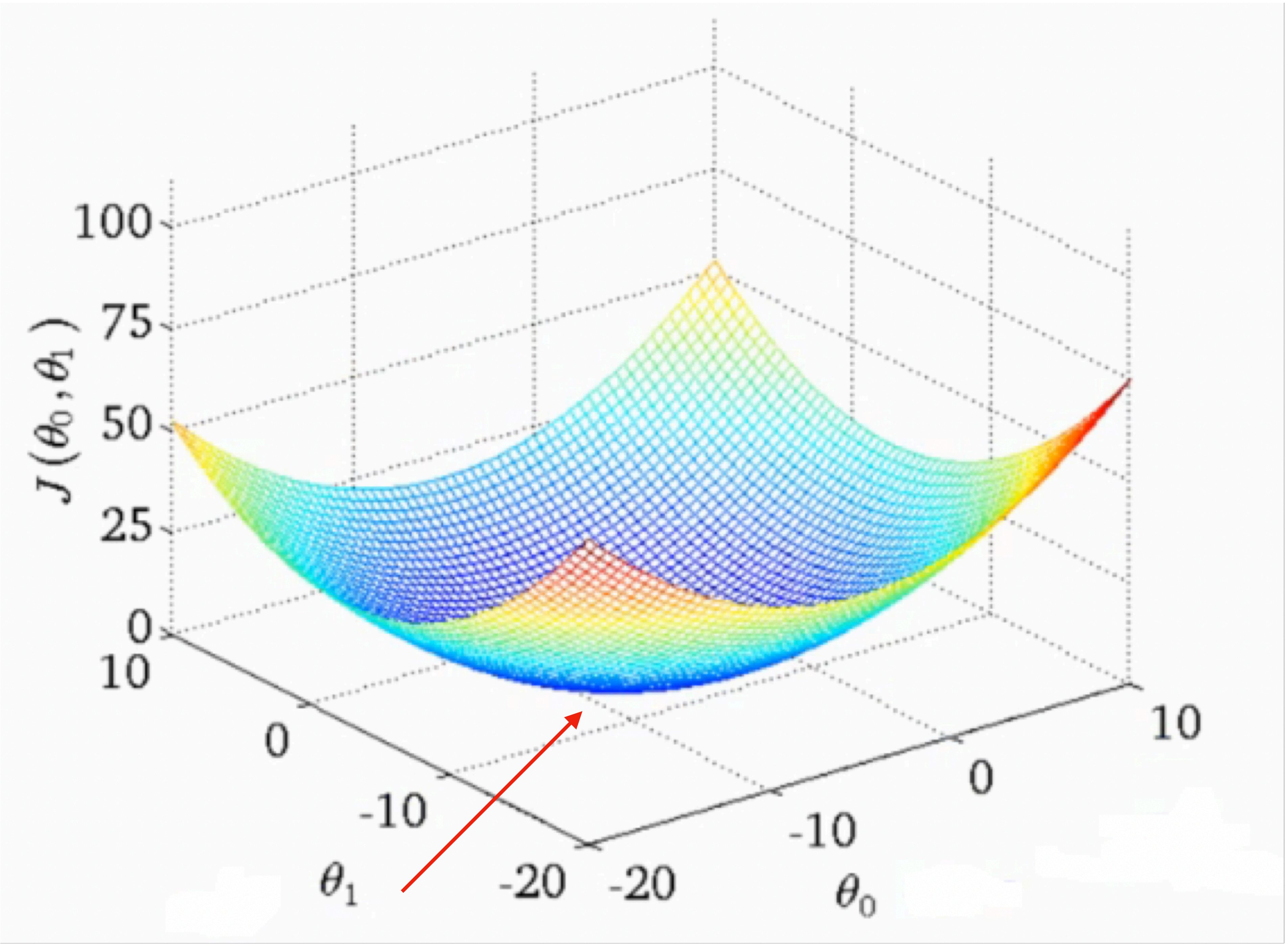
- Loss function을 이용해서 error 값을 측정을한다면, 가장 최소의 error를 빠르게 찾는 알고리즘은 없을까?

- Hypotheis 재정의
  - $h_\theta(x) = \theta_0 + \theta_1 x$

- Cost Function 재정의
  - $J(\theta_1, \theta_2) = \dfrac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$

- $J(\theta_1, \theta_2)$ 의 최솟값 찾기

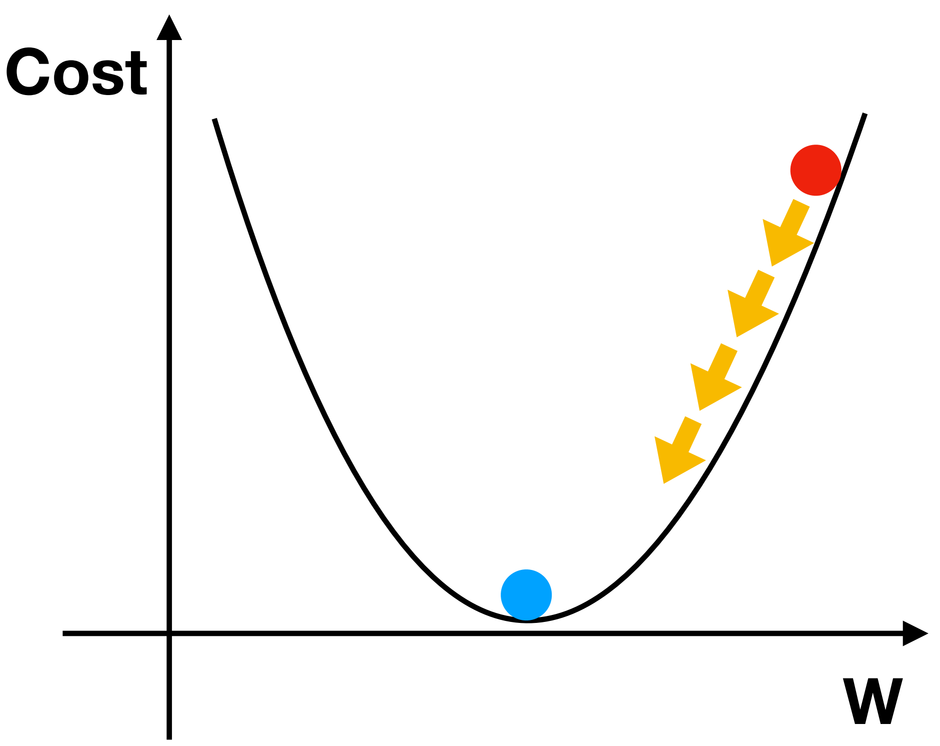
$$J(\theta_1, \theta_2)$$

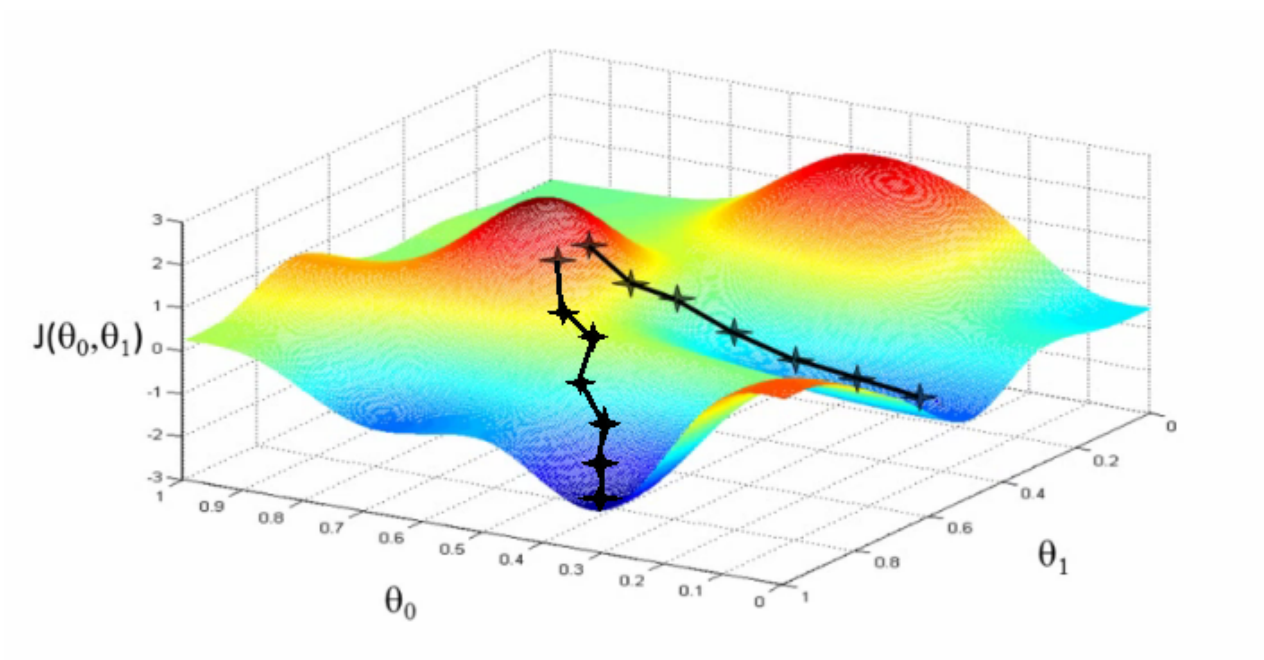


Gradient Descent

- 경사하강법

- 함수를 미분해서, 기울기가 음수인 곳을 계속 찾아간다면 언젠간 그 함수의 최솟값에 도달 하지 않을까?





GD Step by Step

- Cost function 정의

$$J(\theta_1, \theta_2) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2$$

- 임의의 점에서 시작
  - 예: (0,0), (10,3) 등

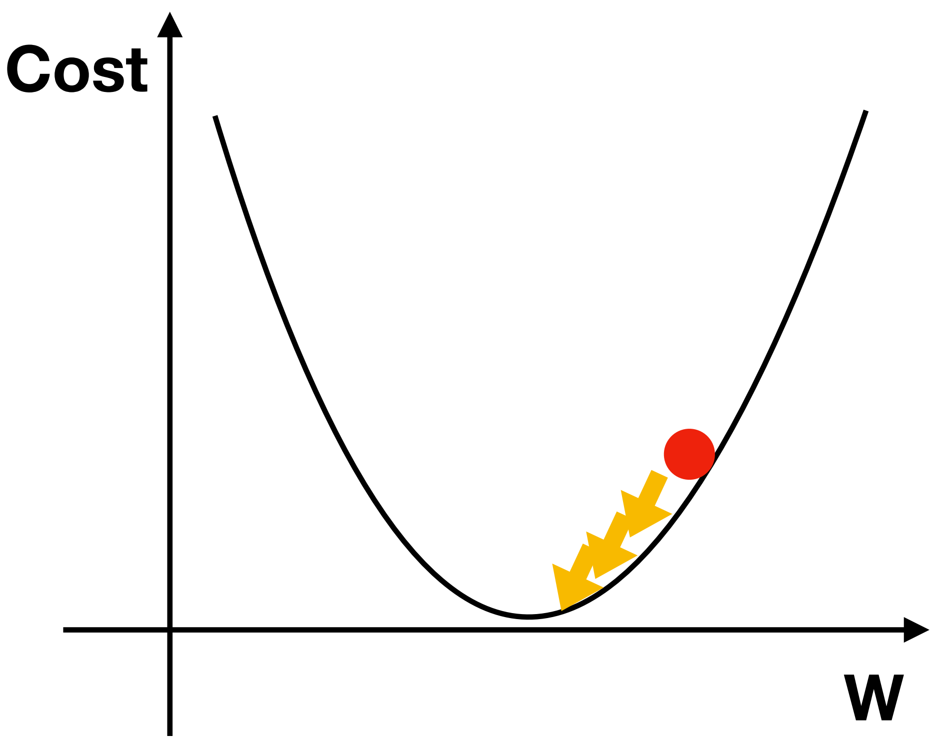
- J를 줄이는 θ_1,θ_2 로 업데이트

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

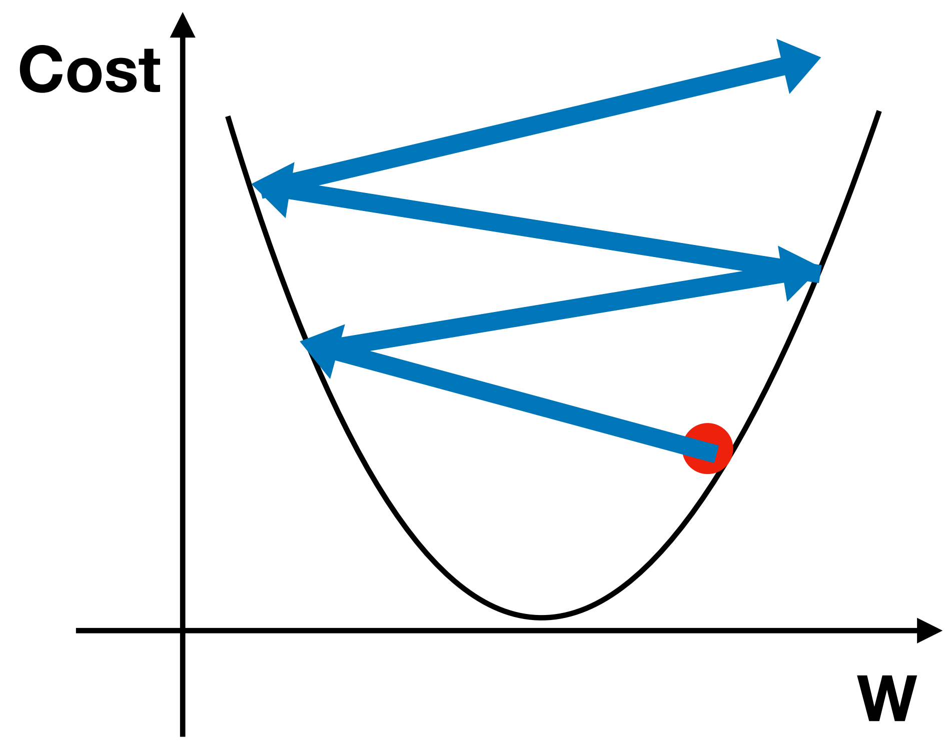
α : learning rate

- 수렴 할 때까지 반복!

Learning Rate



$\alpha = 0.0001$



$\alpha = 0.1$

α는 작으면 작을 수록 좋을까?

E.O.D