Lecture 03

How to minimize cost

Hypothesis and Cost

$$H(x) = Wx + b$$

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x^{(i)}) - y^{(i)})^2$$

Simplified hypothesis

$$H(x) = Wx$$

$$cost(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})^2$$

What cost(W) looks like?

$$cost(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

x	Y
1	1
2	2
3	3

$\rightarrow$ $W = 1$, $cost(W) = ?$ $\boxed{0}$

$$\frac{1}{3}(\underbrace{(1*1-1)^2}_{0} + \underbrace{(1*2-2)^2}_{0} + \underbrace{(1*3-3)^2}_{0})$$

What cost(W) looks like?

$$cost(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

X	Y
1	1
2	2
3	3

- W=1, cost(W)=0

$$\frac{1}{3}((1*1-1)^2 + (1*2-2)^2 + (1*3-3)^2)$$

- W=0, cost(W)=4.67

$$\frac{1}{3}((0*1-1)^2 + (0*2-2)^2 + (0*3-3)^2)$$

- W=2, cost(W)=? *4.67*

$$\frac{1}{3}((2*1-1)^2 + (2*2-2)^2 + (2*3-3)^2)$$

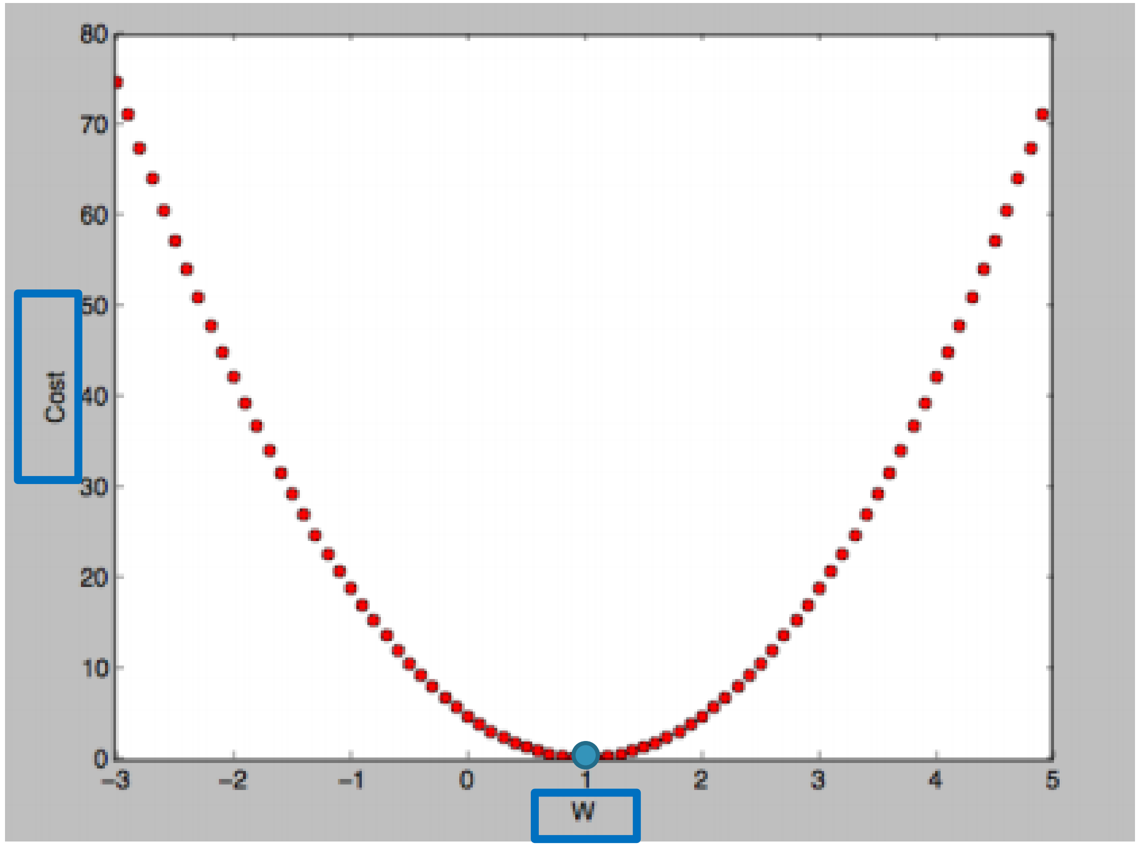
What cost(W) looks like?

- W=1, cost(W)=0
- W=0, cost(W)=4.67
- W=2, cost(W)=4.67

What cost(W) looks like?

$$cost(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})^2$$

- W=1, cost(W)=0
- W=0, cost(W)=4.67
- W=2, cost(W)=4.67

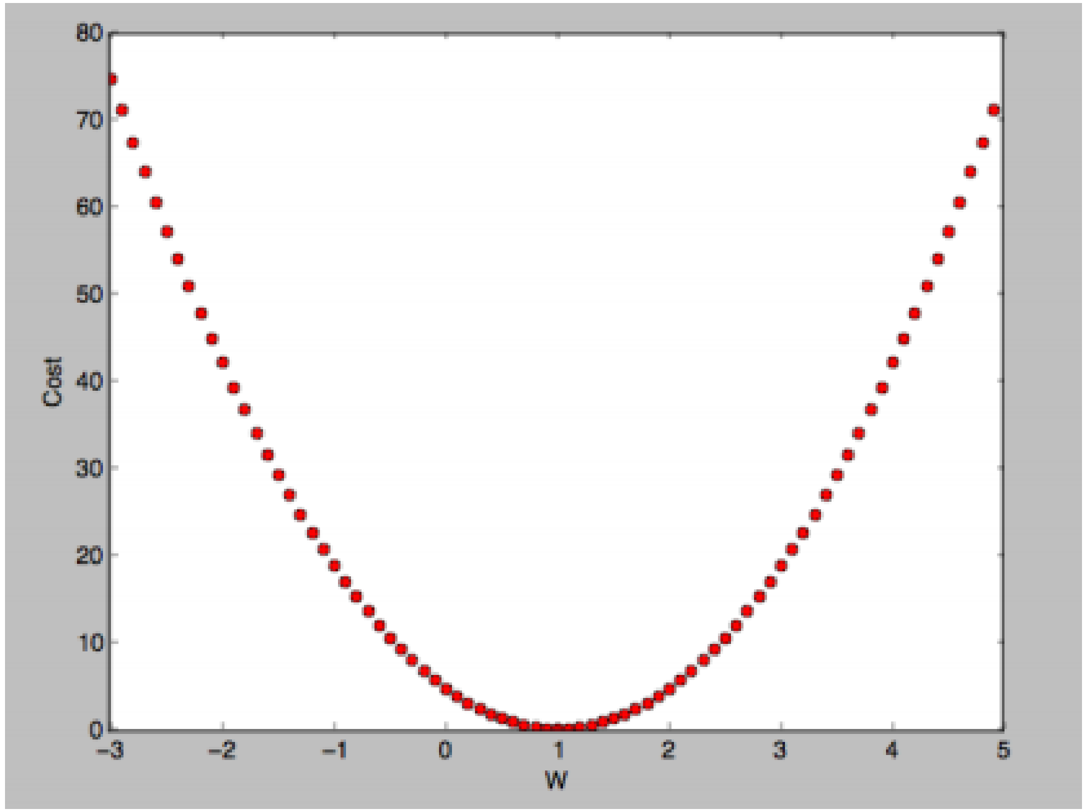


Gradient descent algorithm

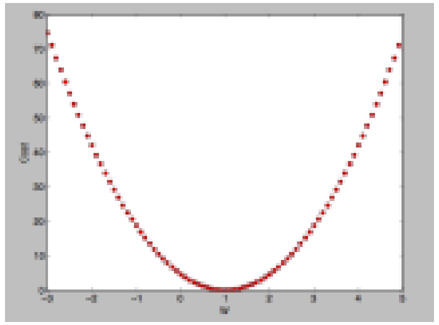
- Minimize cost function
- **Gradient descent** is used many **minimization problems**
- For a given **cost function**, **cost (W, b)**, it will find W, b to minimize cost
- It can be applied to more general function: **cost (w1, w2, w3 …)**

How it works?

- How would you find the lowest point?



How it works?



- Start with initial guesses
  - Start at **0,0** (or **any other value**)
  - Keeping changing W and b a little bit to try and reduce cost(W, b)
- **Each time you change the parameters**, you **select the gradient** which reduces cost(W, b) the most possible
- Repeat
- Do so until you converge to a local minimum
- Has an interesting property
  - Where you start can determine which minimum you end up

gradient는 미분을
이용하여 구할 수 있다.

Formal definition

$$cost(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

$$\Downarrow$$

$$cost(W) = \frac{1}{2m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

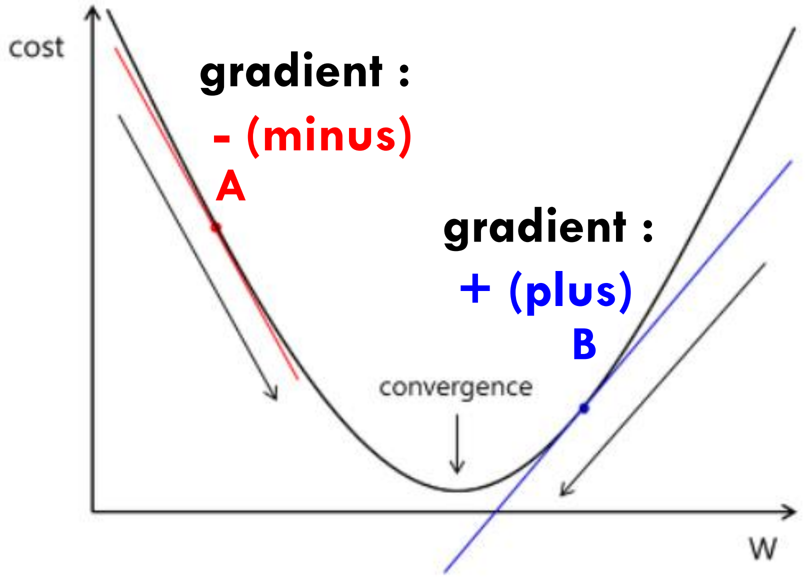
미분을 적용하기 쉽도록
분모에 2를 곱한다.

Formal definition

$$cost(W) = \frac{1}{2m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

learning_rate

$$W := W - \alpha\frac{\partial}{\partial W}cost(W)$$



Formal definition

$$cost(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

$$cost(W) = \frac{1}{2m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

$$W := W - \alpha\frac{\partial}{\partial W}cost(W)$$

$$W := W - \alpha\frac{\partial}{\partial W}\frac{1}{2m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

$$W := W - \alpha\frac{1}{2m}\sum_{i=1}^{m}2(Wx^{(i)} - y^{(i)})x^{(i)}$$

$$\boxed{W := W - \alpha\frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})x^{(i)}}$$

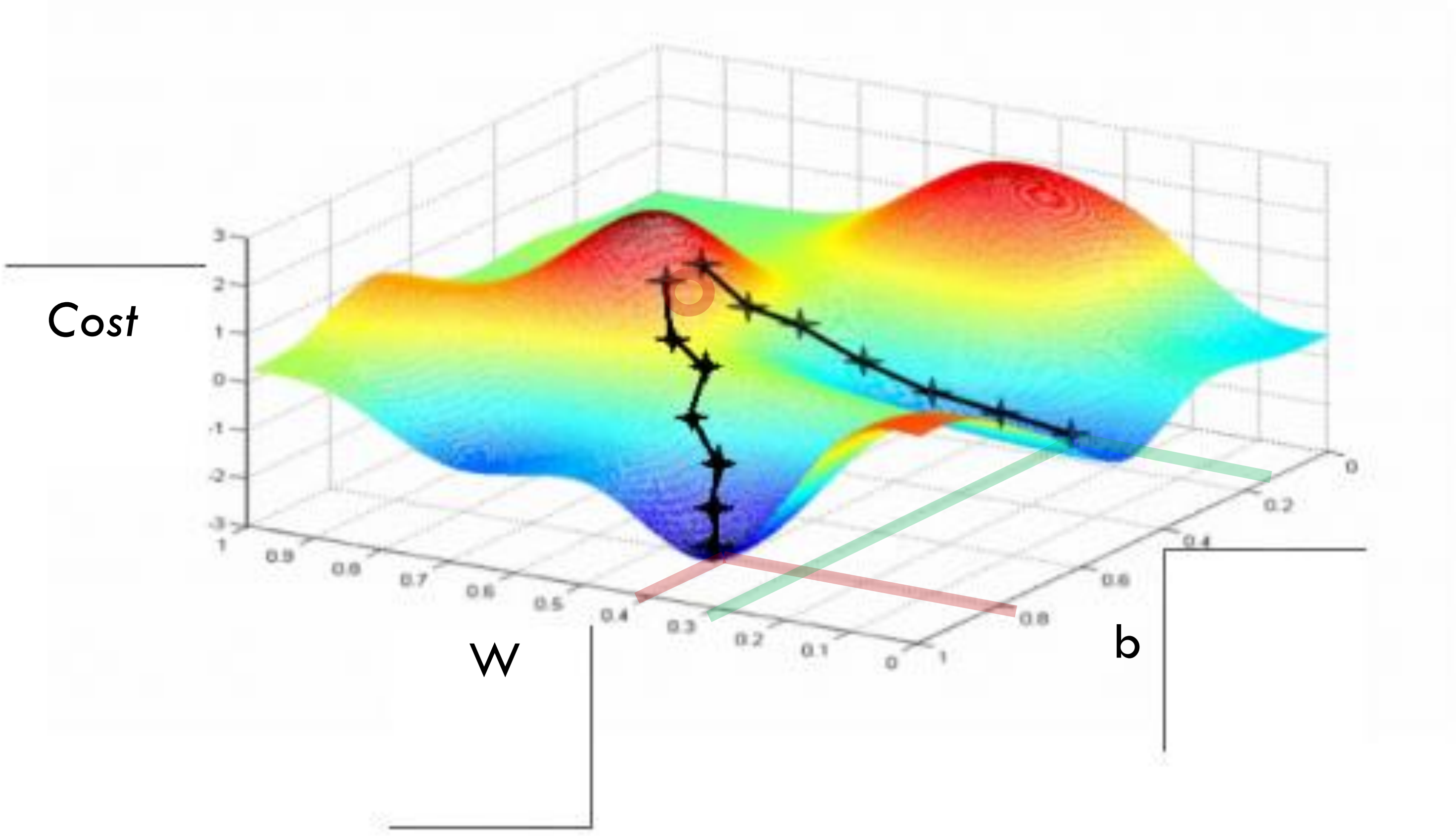
Derivative Calculator

https://www.derivative-calculator.net/

Gradient descent algorithm

$$W := W - \alpha \frac{1}{m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)}) x^{(i)}$$

Convex function



Convex function

$$cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$$