KU KONKUK UNIVERSITY

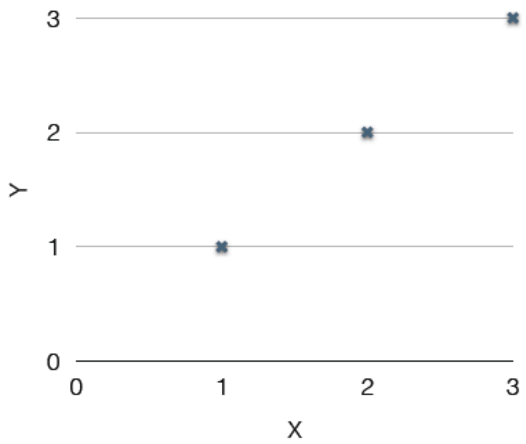
Concept of Machine Learning

자연어처리연구실

본 챕터는 김성훈 박사님의 "모두를 위한 머신러닝/딥러닝 강의 (https://hunkim.github.io/ml/)" 자료를 참고하여 편집되었습니다.

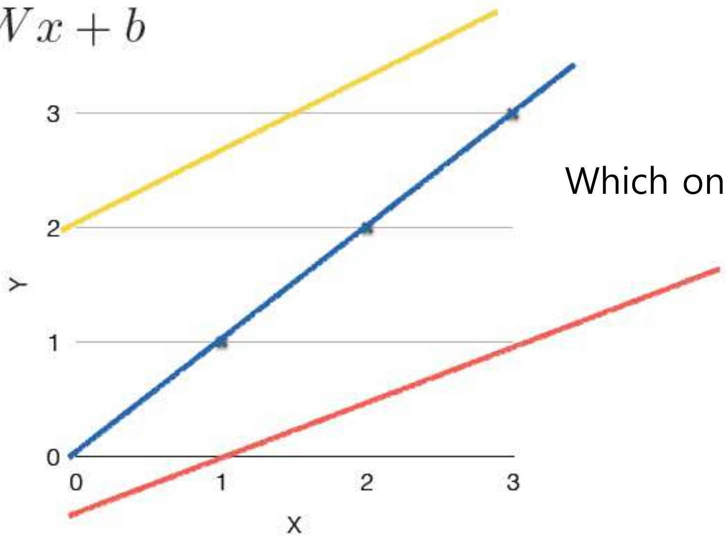
What is Learning?

X	Y
1	1
2	2
3	3



Linear Hypothesis

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



Which one is better?

Multi-variable

$$H(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$

$$H(x_1, x_2, x_3, ..., x_n) = w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n + b$$

Matrix representation

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n$$

$$\begin{bmatrix} w1 & w2 & w3 \end{bmatrix} \times \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} w1 \times x1 + w2 \times x2 + w3 \times x3 \end{bmatrix}$$

Matrix Representation

$$\begin{bmatrix} w1 & w2 & w3 \end{bmatrix} \times \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} w1 \times x1 + w2 \times x2 + w3 \times x3 \end{bmatrix}$$

$$H(X) = WX + b$$

With b vector

$$\begin{bmatrix} b & w1 & w2 & w3 \end{bmatrix} \times \begin{bmatrix} 1 \\ x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} b \times 1 + w1 \times x1 + w2 \times x2 + w3 \times x3 \end{bmatrix}$$

$$H(X) = WX$$

Without b vector

$$H(X) = W^T X$$

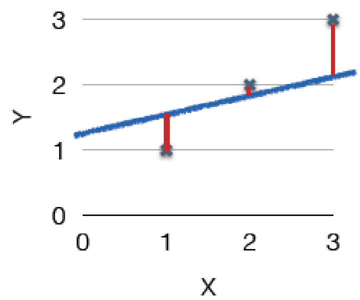
Transpose representation

Which hypothesis is better?

- How fit the line to our (training) data

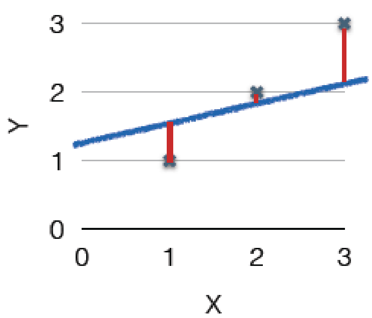
$$H(x) - y$$

$H(x) = Wx + b$



Cost Function

$H(x) = Wx + b$



$$\frac{(H(x^{(1)}) - y^{(1)})^2 + (H(x^{(2)}) - y^{(2)})^2 + (H(x^{(3)}) - y^{(3)})^2}{3}$$

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

Our goal? $\underset{W,b}{\text{minimize}}\, cost(W, b)$

Cost function을 최소로 하는hypothesis가 무엇일까?

Hypothesis and Cost

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

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

Simplifying without b vector

$$H(x) = Wx$$

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

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

- W=1, cost(W)=0

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

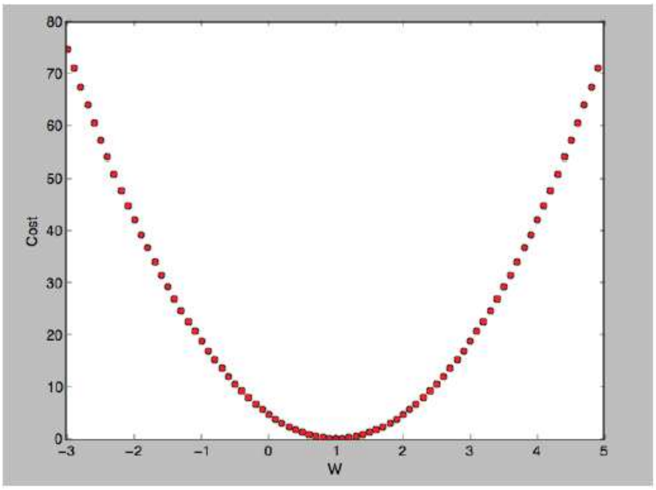
- W=0, cost(W)=4.67

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

- W=2, cost(W)=? $\boxed{?}$

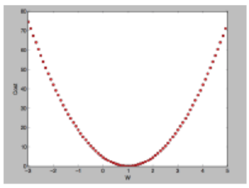
What cost(W) looks like?

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



How to Minimize Cost?

= How to find the lowest point?



- 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

Formal Definition of Gradient Decent

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

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

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

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

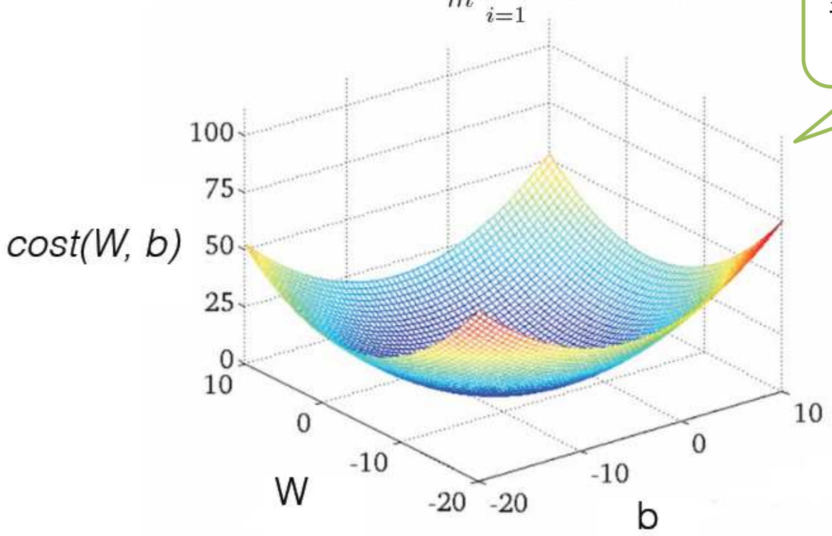
$$W := W - \alpha\frac{1}{2m}\sum_{i=1}^{m}2(Wx^{(i)} - y^{(i)})x^{(i)}$$

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

Convex Function

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

Learning → Cost를 최소화하는 W와 b를 찾자!

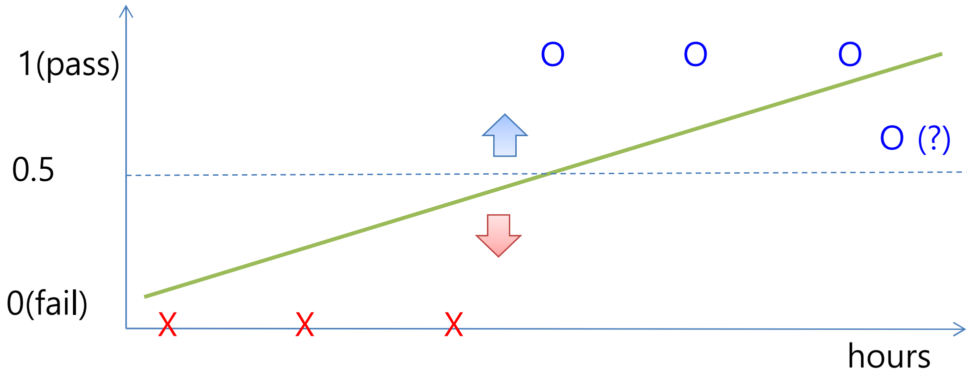


Regression to Classification

Classification problems

- Spam Detection: Spam (1) or Ham (0)
- Facebook feed: show(1) or hide(0)
- Credit Card Fraudulent Transaction detection: legitimate(0) or fraud (1)

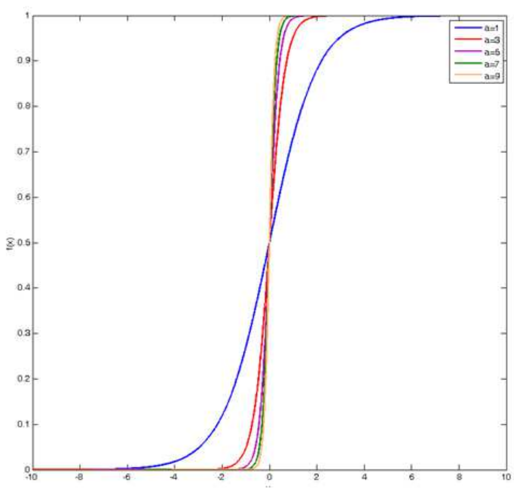
Pass/Fail based on study hours?



Logistic Hypothesis

$$H(x) = Wx + b \Rightarrow g(z) = \frac{1}{(1 + e^{-z})}$$

WHY?
0과 1 사이 값으로 변환



Logistic Hypothesis & Cost Function

$$H(X) = \frac{1}{1 + e^{-W^T X}}$$

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

New Cost Function

$$Cost(W) = \frac{1}{m}\sum c(H(x), y)$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

H(x)=1일 때 C값은? ?

Cost Function

$$Cost(W) = \frac{1}{m}\sum c(H(x), y)$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

⇩

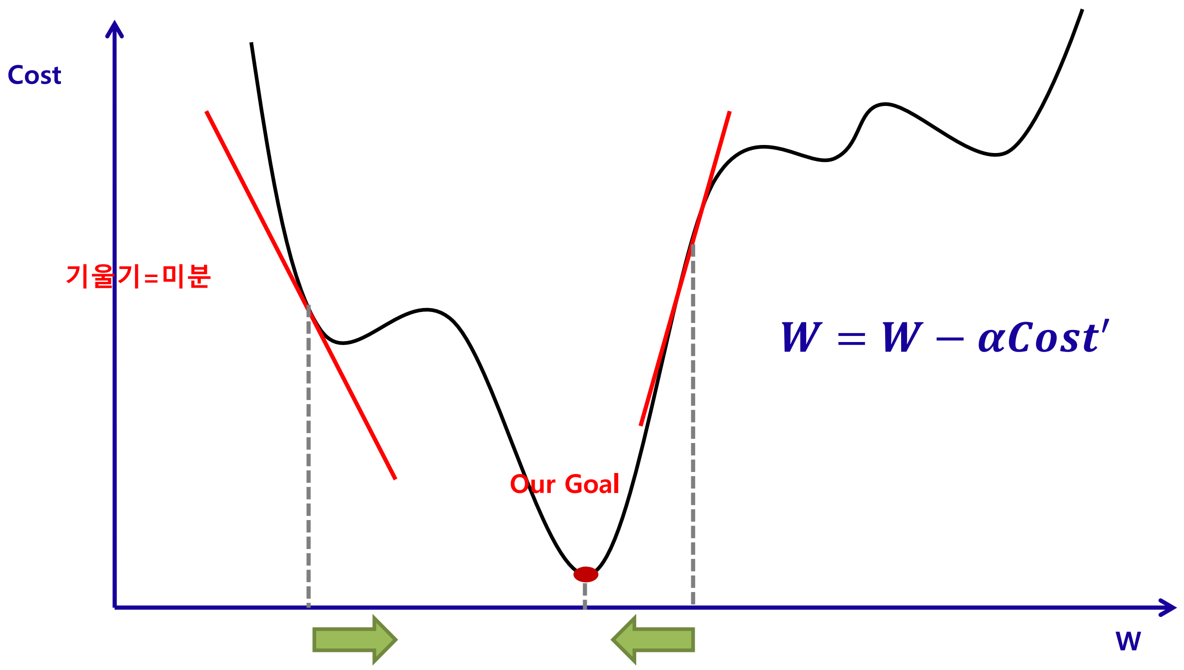
$$c(H(x), y) = -ylog(H(x)) - (1 - y)log(1 - H(x))$$

Minimize Cost → Gradient decent algorithm

$$Cost(w) = -\frac{1}{m}\sum ylog(H(x)) + (1 - y)log(1 - H(x))$$

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

Goal of ML Models



확인 문제

- 다음 학습 데이터와 비용 함수(cost function)가 주어지고, 초기 W값이 2이고 학습률이 0.1일 때, gradient decent 알고리즘에 의해 1회 학습 후 수정된 W 값을 구하시오.

[학습 데이터]

X (입력)	Y (출력)
1	1
2	3
3	5

[비용 함수]

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

확인 문제

?

질의응답

Q&A

Homepage: http://nlp.konkuk.ac.kr
E-mail: nlpdrkim@konkuk.ac.kr

Edited by Harksoo Kim