Cost function

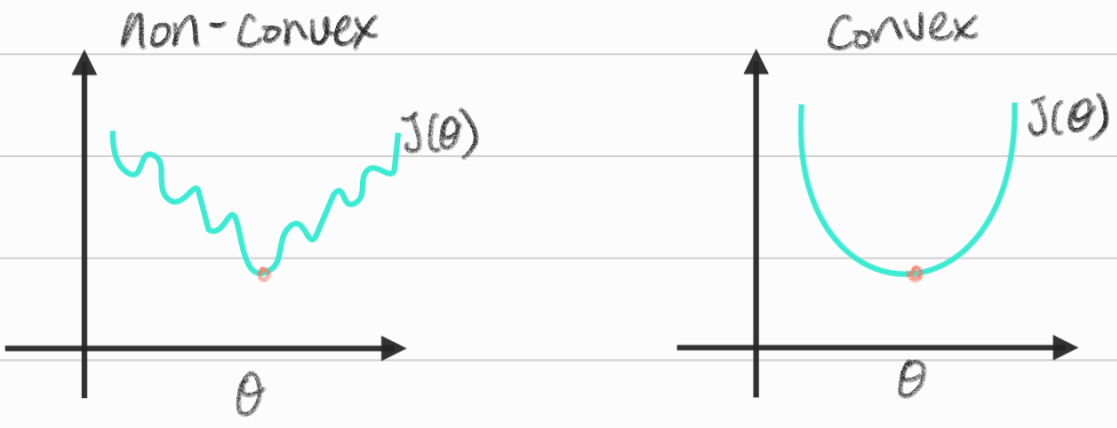
of Linear regression : $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

$\text{Cost}\left(h_\theta(x^{(i)}) - y^{(i)}\right) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$ (오차^2 / 2)

예상값 실제라벨

∴ Cost function 비용 함수는 훈련집합의 총합

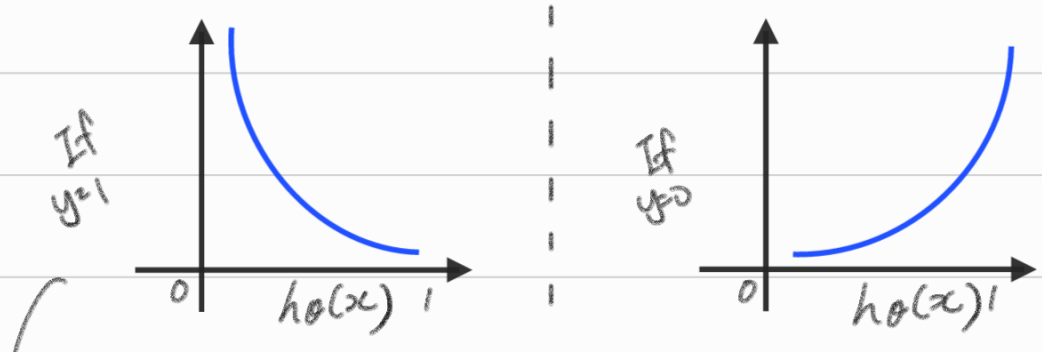
Logistin function은 "non-convex" 해서 위식 적용 적당 x



다른 Costfunction 필요!

Logistic Regression cost function

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$



"Cost=0"

if $y=1$, $h_\theta(x)=1$

실제레이블 예측값

But, as $h_\theta(x) \to 0$

Cost $\to \infty$

"Cost=0"

if $y=0$, $h_\theta(x)=0$

But, as $h_\theta(x) \to 1$

Cost $\to \infty$

Simplified Cost Function and Gradient Descent

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \quad -① \\ -\log(1-h_\theta(x)) & \text{if } y=0 \quad -② \end{cases}$$

$$= -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x)) \quad \text{"}\smile\text{"}$$

$y=1$이면 ①, $y=0$이면 ②

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]$$

최대우도 측정법

came from MLE (Maximum Likelihood Estimation)

=> 서로 다른 모델에 대한 매개 변수 효율적으로 찾는 방법

To fit parameters θ:

$\min_\theta J(\theta)$

Repeat $\{\theta_j := \theta_j - \alpha \cdot \frac{\partial}{\partial \theta_j} J(\theta)\}$ (계속바꿔줘)

학습률 미분항 $= \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$

Advanced Optimization

Not only Gradient Descent,

But also Conjugate gradient, BFGS, L-BFGS ···

(α 직접 필요 X, often faster but more comblex)

이와 Octave library 소개.