* 기계학습 5주차 A4 1Page 정리 2016/04/09 컴퓨터공학과 김성수

* MSE : Mean Squared Error

- $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$
- $J(\theta) = \frac{1}{2m}\|X\theta - y\|_2^2$

$$= \frac{1}{2m}(X\theta - y)^T(X\theta - y)$$

* Gradients of MSE. p.8

- $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$
- $\nabla J(\theta) = \frac{1}{m}X^T(X\theta - y)$

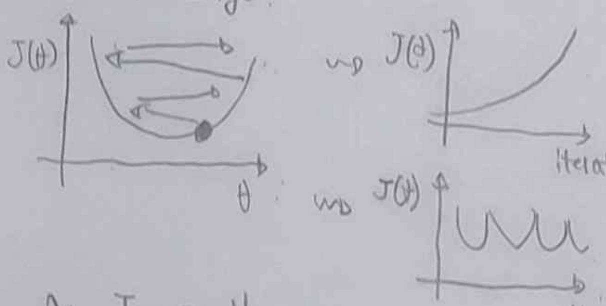
* Normal Equation. P.10.

$$\nabla J(\theta) = \frac{1}{m}X^T(X\theta - y) = 0$$

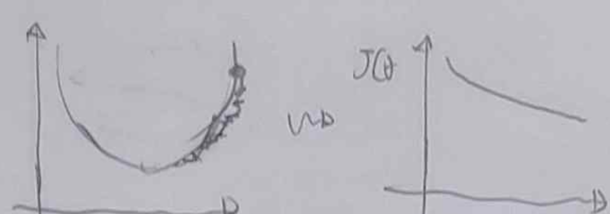
$$\Rightarrow \theta^* = (X^TX)^{-1}X^Ty$$

* Learning Rate (α)

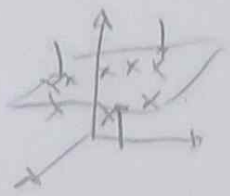
$\alpha \to$ Too Large



$\alpha \to$ Too Small



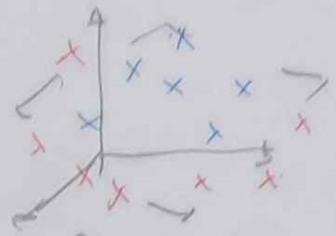
* Logistic Regression. (Classification)



Regression

$y = h_\theta(x)$

$= \theta^T x$



Classification

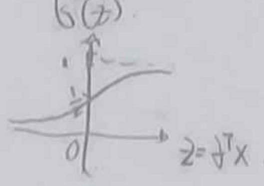
$y = h_\theta(x)$

$= h_\theta\left(\frac{1}{1+e^{-\theta^T x}}\right)$

$= h_\theta(z)$

$z = \sigma(z) = \frac{1}{1+e^{-z}}$

+Logistic Sigmoid

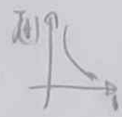


* Cost function.

- Cross Entropy Loss.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}Cost\left(h_\theta(x^{(i)}), y^{(i)}\right)$$

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



$Cost(h_\theta(x), y) = -\log(1-h_\theta(x))$ if) $y=0$

$\to Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$

* Gradient Descent

$$\theta_j := \theta_j - \frac{\alpha}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} =$$

$$\theta := \theta - \alpha\frac{1}{m}X^T(\sigma(X\theta) - y).$$

* Derivation of Gradients of Cross-Entropy Loss.

$$h_\theta(x) = \sigma(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}Cost\left(h_\theta(x^{(i)}), y^{(i)}\right)$$

$$Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

$$\frac{\partial Cost}{\partial \theta_j} = -yx_j + h_\theta(x)\cdot x_j = (h_\theta(x) - y)\cdot x_j$$