* 기계학습 2주차 A4 1Page 정리 2016104109 컴퓨터공학과 김상수

* 선형 모델

$h(x_1, \cdots, x_n) = \theta_0 + x_1\theta_1 + \cdots + x_n\theta_n$

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \quad \Big\} \; \theta^T x = x^T \theta$

+하나의 Training Sample

$h(x)^{(i)} = \theta^T x^{(i)} + \theta_0$

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} = \begin{bmatrix} -x^{(1)T}- \\ -x^{(2)T}- \\ \vdots \\ -x^{(m)T}- \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} + \theta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\underset{(m\times 1)}{h} = \underset{(m\times n)}{X} \cdot \underset{(n\times 1)}{\theta} + \theta_0 \underset{(m\times 1)}{[1]}$$

* Error L2 Norm

$e^{(i)} = h(x^{(i)}) - y^{(i)}$

$= \hat{y}^{(i)} - y^{(i)}$

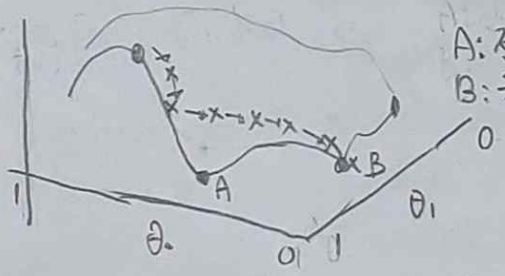
$J = \frac{1}{2m}\|e\|_2^2 = \frac{1}{2m} e^T e$

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

cf) $\|x\|_2 = \sqrt{a^2+b^2}$ → 길이 계산방과 동일!

* Optimization

- Gradient Descent: 국소적 최소점 O 전역적 최소점 X

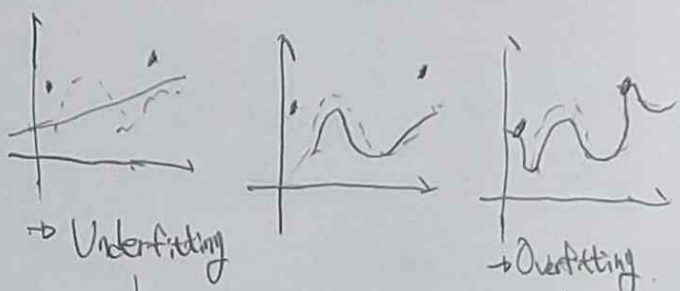


A: 전역적 최소점
B: 국소적 최소점

* MSE (Mean Squared Error)

$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

* Overfitting & Underfitting

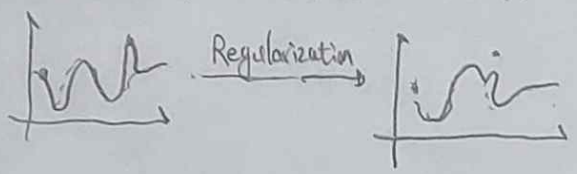


→ Underfitting
↓
Model의 학습량이 부족.
TrainSet에서도 학습 못한 상태

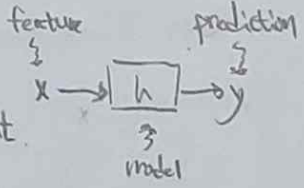
→ Overfitting
↓
Exception까지 학습.
과도하게 학습한 상태
↳ 낮은 일반화능력

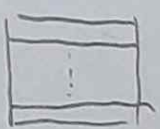
* Regularization

→ Overfitting 방지
→ 규제를 적용하여 스무스하게 모델이 나오도록



* Terminology



- Feature: 특징, input
- Prediction: 예측치, output
- Model: 함수, hypothesis.
- Label: 정답, target
- Classifier: 분류기. (Ex: k-NN, Decision Tree ...)
- Training Sample: ▭ 학습의 단위
- Training Data:  Sample의 모임.

- Loss function = Cost Function = Objective Function
  손실 비용 목적