<Lec 01>

ML: 컴퓨터가 스스로 학습할 수 있도록 연구하는 분야.

Supervised learning →
- regression
- binary classification
- multi-label classification

  - data with label
  - ex) cat, dog, ... classification

Unsupervised learning
  - data without label
  - ex) Google news, word clustering.

<Lec 02>

Regression: $\boxed{x} \xrightarrow{\text{예측}} \boxed{y}$

i) data가 주어져 있음.

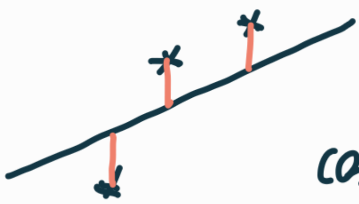
ex)

x	y
1	·
2	·
3	·

ii) hypothesis 세우기.

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

iii) 세운 가설과 실제 데이터의 오차 계산하기.



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

IV) 최적의 가설 찾기

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

→ Goal : minimise Cost(w,b) by choosing optimal w, b

<Lec 03>

문제 단순화 : b=0 이라 하기.

$Cost(W) = cost(W,0) = \frac{1}{m}\sum_{i=1}^{m}\{Wx_i - y_i\}^2$ 을

최소화하자.

⇒ Gradient descent algorithm

- minimise cost function
- works for convex functions
- can be applied to more general functions

  ex) $Cost(W_1, W_2, \cdots, W_n)$

① Cost(W)는 2차 함수다.

→ Cost(W) is convex

② Set start point such as (0,0) = (W, cost(W))

③ 이후의 과정을 위해 $cost(W) = \frac{1}{2m}\sum_{i=1}^{m}\{Wx_i - y_i\}^2$ (2로 나눔)

④ 다음을 극솟값을 찾을 때 까지 반복

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

ㄴ기울기

조정값(step size)

Note. $\frac{\partial}{\partial W} cost(W) = \frac{1}{m}\cdot\sum_{i=1}^{m}\{Wx_i - y_i\}x_i$

<Lec 04>

Multi-variable linear regression.

$$H(x_1, x_2, \cdots, x_n) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

x_1 x_2 $\cdots$ x_n	y
$\vdots$	

라는 data가 주어질 때는 H()를 위와 같이 설정.

With Matrix

$$H(X) = X \cdot W, \quad \begin{cases} X = (x_1, x_2, \cdots, x_n) \\ W = (w_1, w_2, \cdots, w_n)^T \end{cases}$$

* Instance(데이터의 개수)가 많을 때 H()값을 행렬로 한번에

$$\begin{array}{c} I_1 : \\ I_2 : \\ \vdots \\ I_m : \end{array} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} H(I_1) \\ H(I_2) \\ \vdots \\ H(I_m) \end{pmatrix}$$

X : m개의 인스턴스   W   결과