모델을 통해 파라미터를 배운다

경험을 통해 지식이 성장해 간다

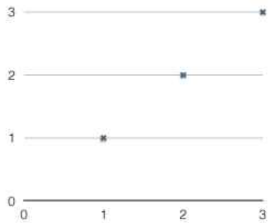
경험: 데이터와 정답(클래스 라벨)

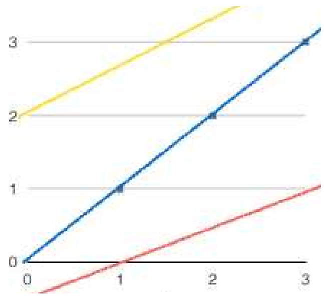
지식 = 작업 목표: 모델(회귀 모델, 분류 모델)

성장 = 성능 지표: 정답에 가까워지는가(평가)

선형 회귀(Linear Regression)

x	y
1	1
2	2
3	3



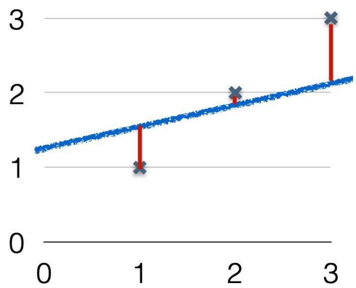


1. Hypothesis(가설 설정)

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

2. Cost Function(정답과 예측값의 차이)

= Loss function, Objective function



Cost: 예측값 - 정답(ground truth) = $H(x) - y$

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

Finding minimum values of cost function

$H(x) = Wx + b$와 $cost = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$을 통해

가장 작은 cost를 갖게 하는 W와 b를 찾는다: W, b = 모델 파라미터

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

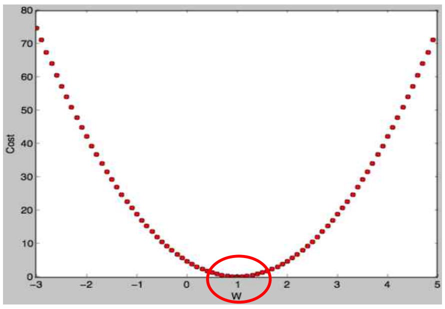
3. 가장 작은 cost 찾기

$$\underset{W,b}{minimize}\, cost(W,b) = \underset{W,b}{min}\, cost(W,b) = \underset{W,b}{argmin}\, cost(W,b)$$

simplify → $H(x) = Wx$ and $cost(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$

경사 하강 알고리즘(GD. Gradient Descent Algorithm)

$$W_{t+1} := W_t - \alpha \frac{\partial}{\partial W_t} cost(W_t)$$



가장 낮은 지점 → 경사 없는 부분(기울기 = 0)

미분을 이용해서 기울기의 크기가 작아지는 방향으로 이동한다

$\frac{\partial}{\partial W_t} cost(W_t) = 0$이 되어 $W_{t+1} := W_t$ 까지

: W가 더는 변하지 않아 cost가 가장 작은 값이 될 때까지 반복해서 갱신

1. 미분의 편의를 위해 cost 함수 변경

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

2. 경사 하강 알고리즘에 대입

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

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

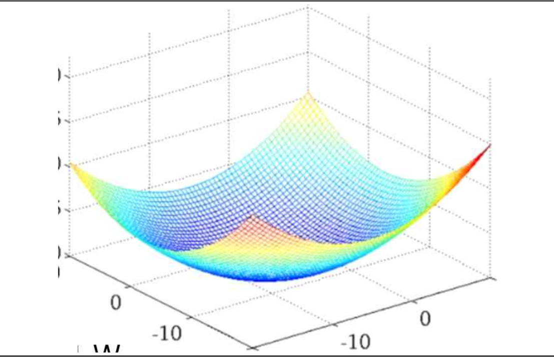
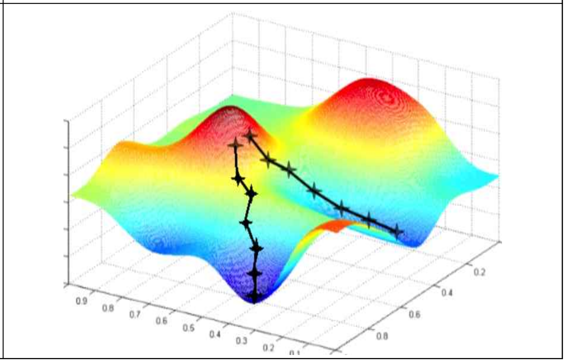
Linear Regression 정리

1. Hypothesis: $H(x) = Wx + b$

2. Cost: $cost(W, b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$ and $\underset{W,b}{min}\, cost(W, b)$

3. Gradient Descent: $W := W - \alpha \frac{\partial}{\partial W} cost(W)$

※ Linear hypothesis 이차방정식은 Convex function으로 Global minimum이 한 개 존재한다
BUT. Non-Convex function은 local minimum을 여러 개 갖는다

Convex function	Non-Convex function
	

다중 선형 회귀(Multivariable Linear Regression)

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

$$H(x_1, x_2, x_3, \cdots, x_n) = w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_{nx_n} + b$$

cost function: $cost(W, b) = \frac{1}{m}\sum_{i=1}^{m}(H(x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}) - y^{(i)})^2$

using Matrix

Hypothesis: $H(X) = XW + b = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} + b$

plus.

$$H(X) = XW + b = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{np} \end{pmatrix} + b$$