

회귀

# 목차

- 선형회귀(Linear regression)
- 다항회귀(Polynomial regression)
- 서포트벡터 회귀(Support vector regression)
- 의사결정나무 회귀(Decision tree regression)
- 신경망 회귀(Neural Network regression)

# 선형회귀(Linear regression)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

## ◎ 가정

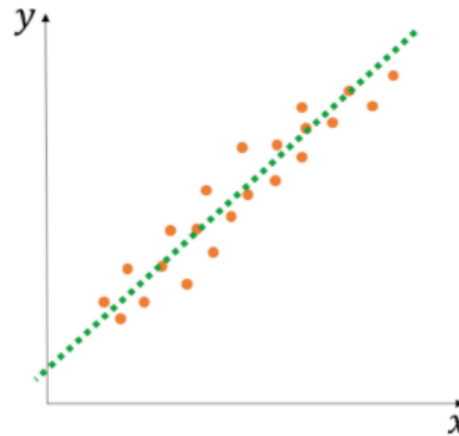
### ○ 모형의 형태에 대한 가정

- 선형성

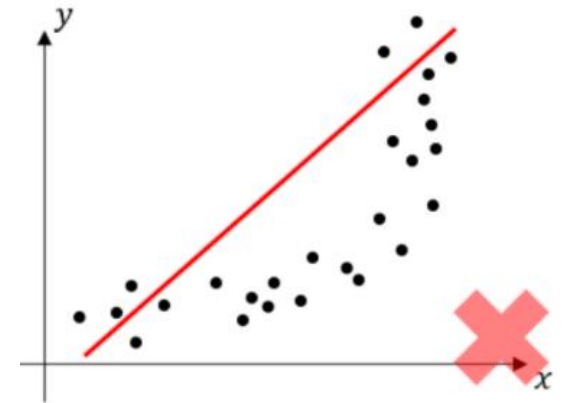
### ○ 오차에 대한 가정

- 독립성
- 등분산성
- 정규성
- 비자기상관적 관계

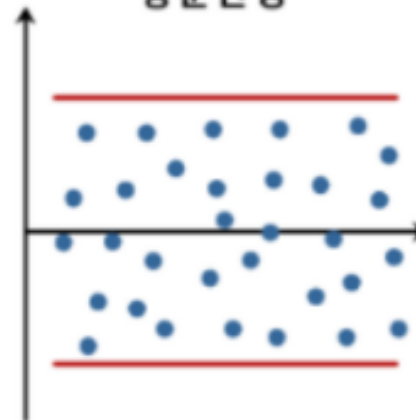
선형성



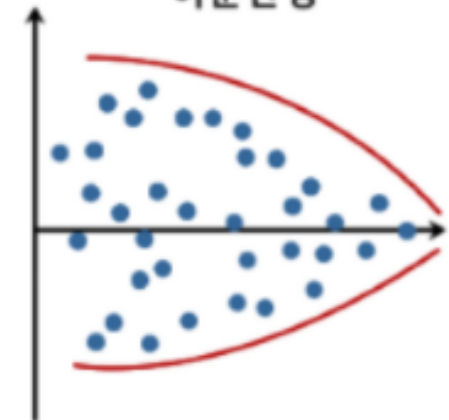
비선형성



등분산성



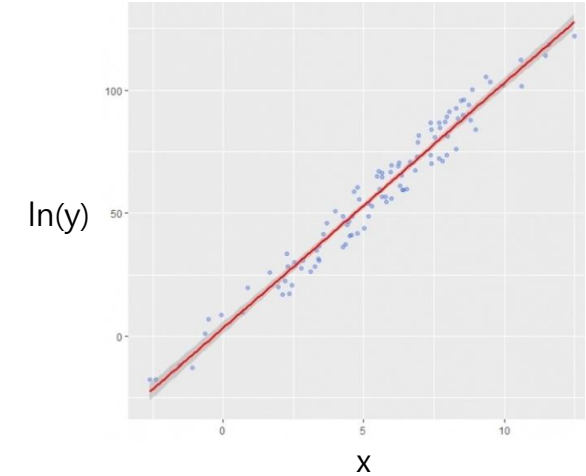
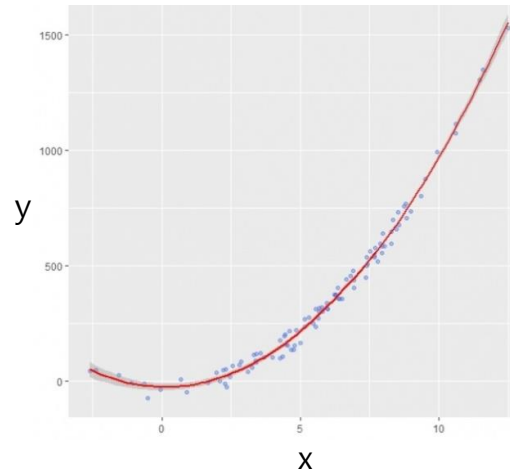
이분산성



# 선형회귀(Linear regression)

## ◎ 선형성 가정 위배

- 로그 변환



## ◎ 등분산성 가정 위배

- 가중최소제곱법(WLS) 적용
- 로그 변환
- 멱변환

### ✓ 최소제곱법(OLS)

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

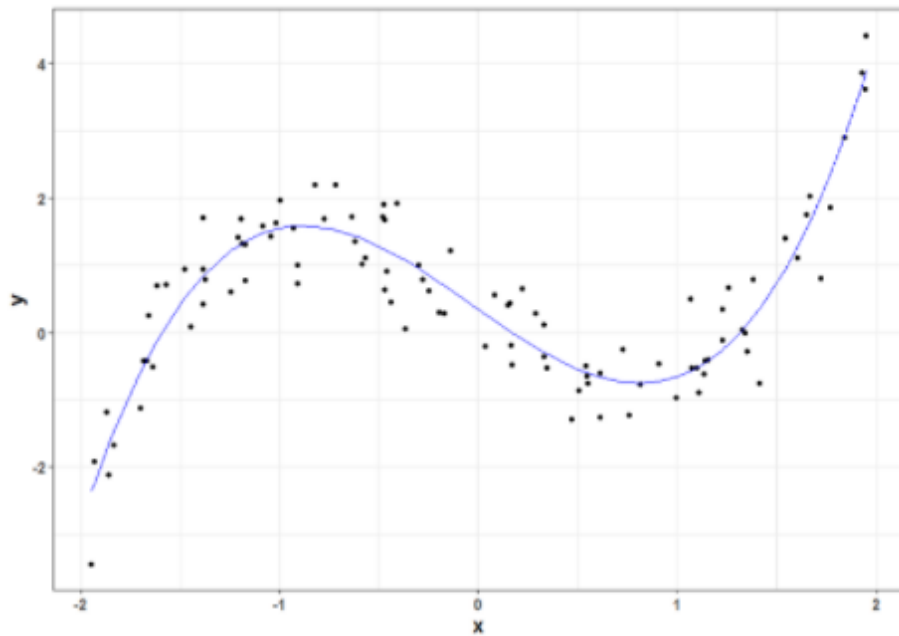
### ✓ 가중최소제곱법(WLS)

$$\sum_{i=1}^n w_i (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

$w_i$  : 분산에 반비례한 가중치  
 $w_i = \frac{1}{\sigma_i^2}$

# 다항회귀(Polynomial regression)

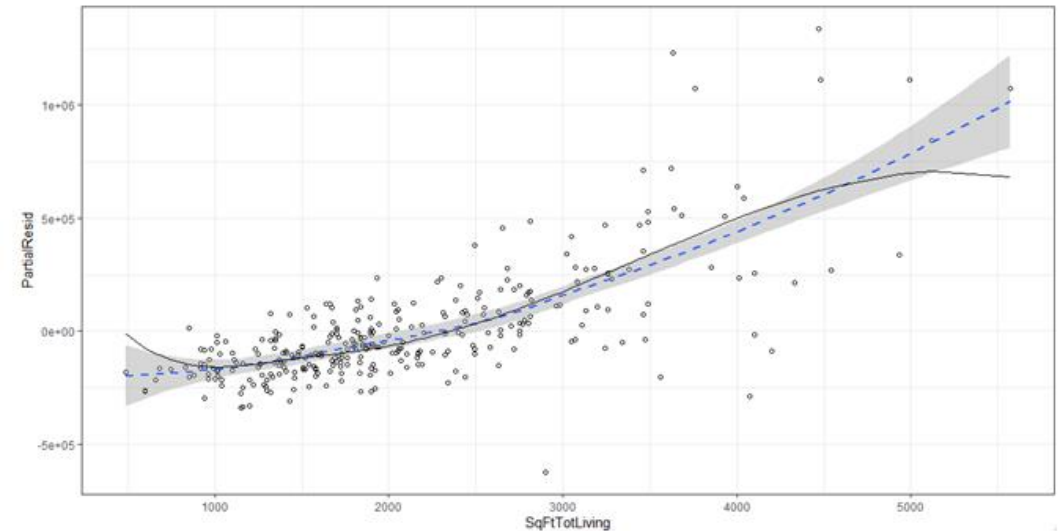
$$Y_i = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p + \varepsilon_i$$



## ✓ 스플라인 회귀

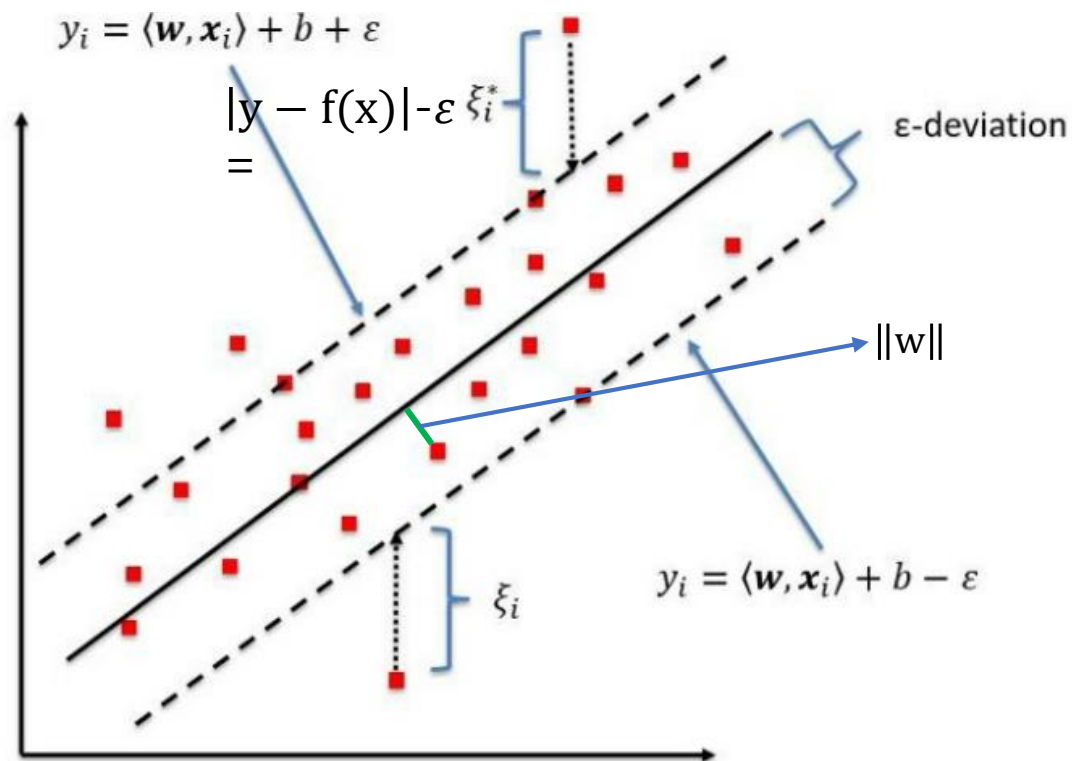
스플라인: 고정된 점들 사이를 부드럽게 보간하는 방법

매듭: 스플라인 구간을 구분하는 값



- 실선: 스플라인 회귀
- 점선: 평활 곡선

# 서포트 벡터 회귀(SVR, Support Vector Regression)



- 회귀식

$$y = f(x) = \langle w, x \rangle + b = \sum_{j=1}^M w_j x_j + b, \quad y, b \in \mathbb{R}, x, w \in \mathbb{R}^M$$

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + b \quad x, w \in \mathbb{R}^{M+1}$$

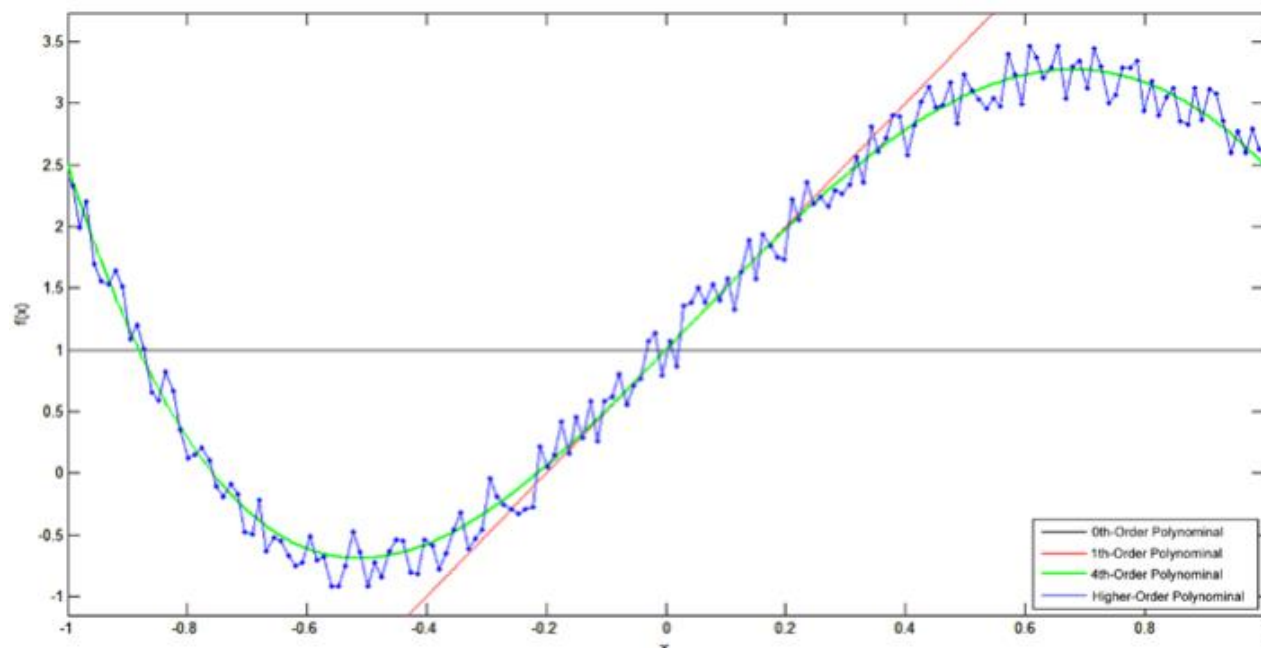
M : 회귀식 차수

회귀식 주변의 가장 좁은 폭의 튜브를 찾는 것이  
최적화 문제

$$\min_w \frac{1}{2} \|w\|^2$$

$\|w\|$  : 회귀식에 대한 법선 벡터 크기

# 서포트 벡터 회귀(SVR, Support Vector Regression)



$$y = f(x) = \langle w, x \rangle + b = \sum_{j=1}^M w_j x_j + b, \quad y, b \in \mathbb{R}, x, w \in \mathbb{R}^M$$

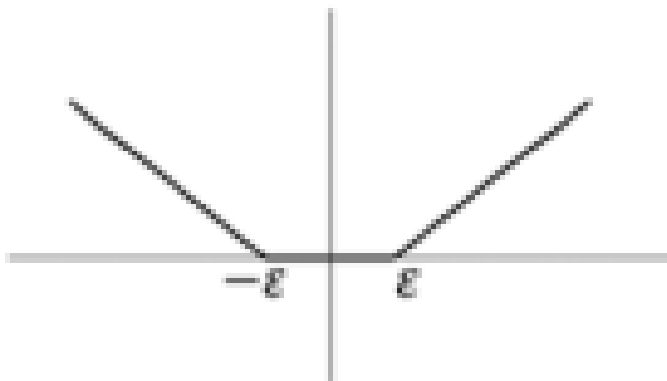
$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + b \quad x, w \in \mathbb{R}^{M+1}$$

# 서포트 벡터 회귀(SVR, Support Vector Regression)

## ✓ 손실함수

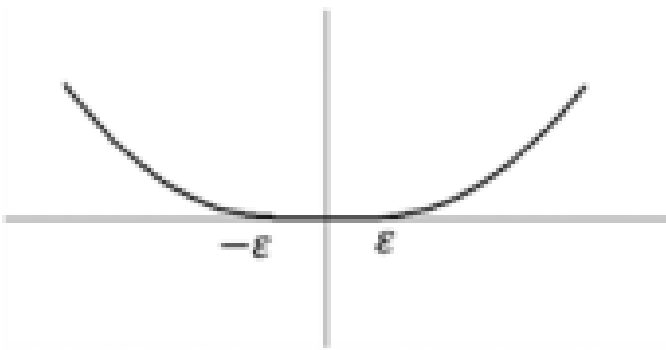
- 선형

$$L_{\varepsilon}(y, f(x, w)) = \begin{cases} 0 & |y - f(x, w)| \leq \varepsilon; \\ |y - f(x, w)| - \varepsilon & \text{otherwise,} \end{cases}$$



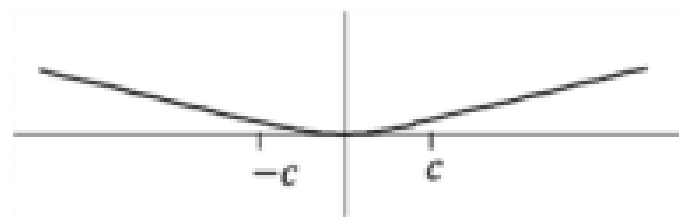
- 2차

$$L_{\varepsilon}(y, f(x, w)) = \begin{cases} 0 & |y - f(x, w)| \leq \varepsilon; \\ (|y - f(x, w)| - \varepsilon)^2 & \text{otherwise,} \end{cases}$$



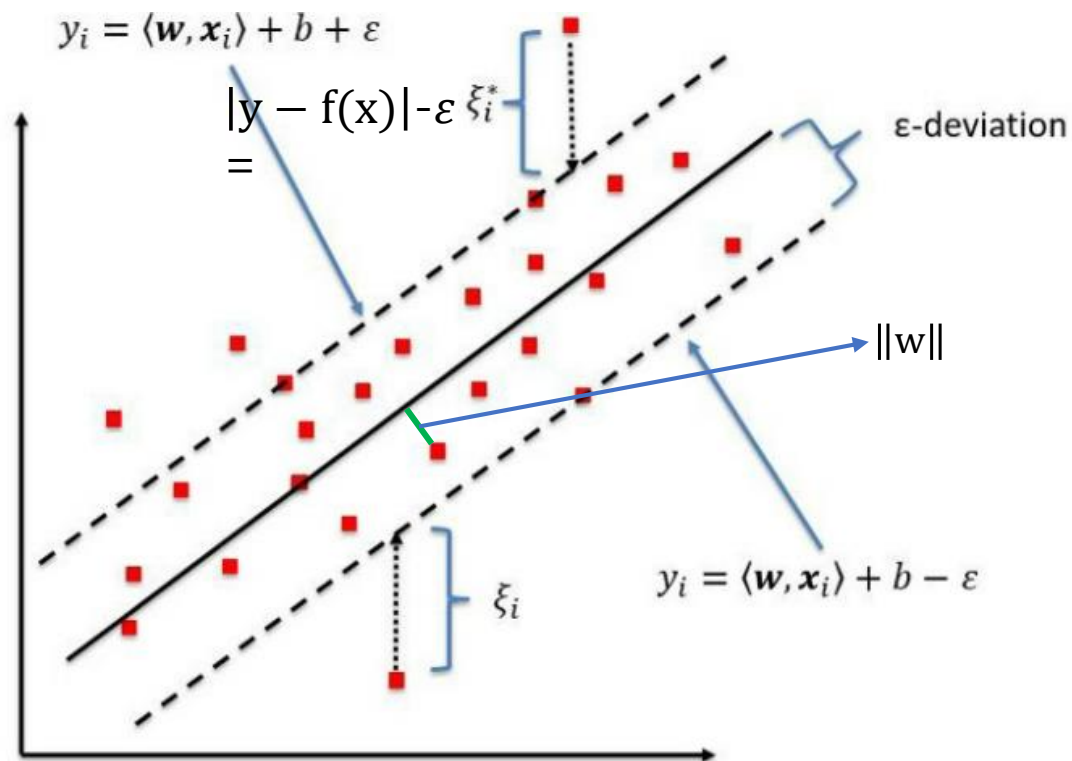
- Huber

$$L(y, f(x, w)) = \begin{cases} c|y - f(x, w)| - \frac{c^2}{2} & |y - f(x, w)| > c \\ \frac{1}{2}|y - f(x, w)|^2 & |y - f(x, w)| \leq c \end{cases}$$





# 서포트 벡터 회귀(SVR, Support Vector Regression)



- 최적화 문제

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i + \xi_i^*$$

- 라그랑지안

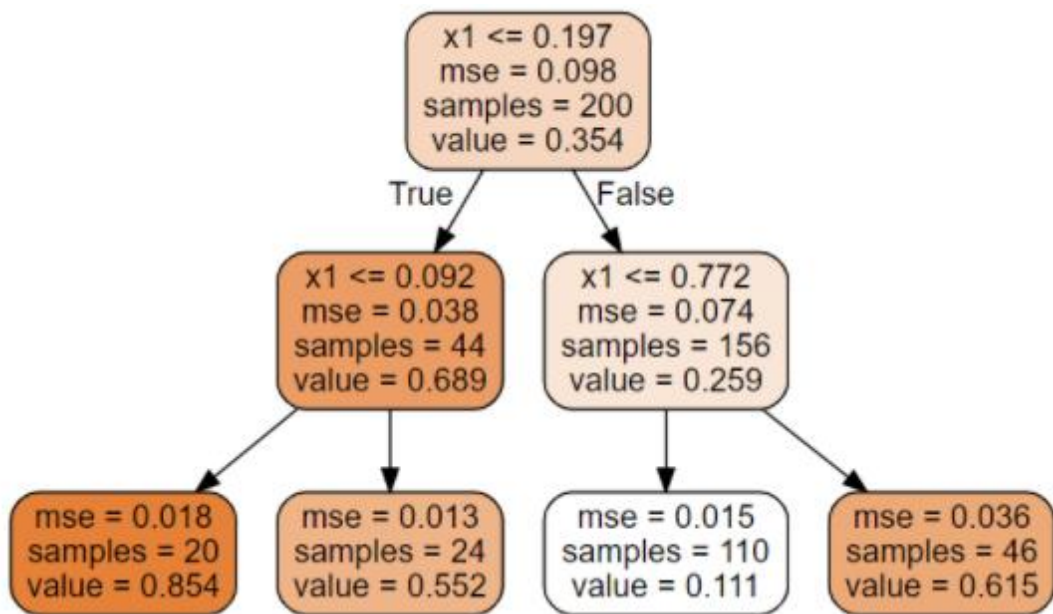
$$\begin{aligned} \mathcal{L}(w, \xi, \xi^*, \lambda, \lambda^*, \alpha, \alpha^*) = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i + \xi_i^* + \sum_{i=1}^N \alpha_i^* (y_i - w^T x_i - \epsilon - \xi_i^*) \\ & + \sum_{i=1}^N \alpha_i (-y_i + w^T x_i - \epsilon - \xi_i) - \sum_{i=1}^N \lambda_i \xi_i + \lambda_i^* \xi_i^* \end{aligned}$$

# 의사결정나무 회귀(Decision Tree regression)

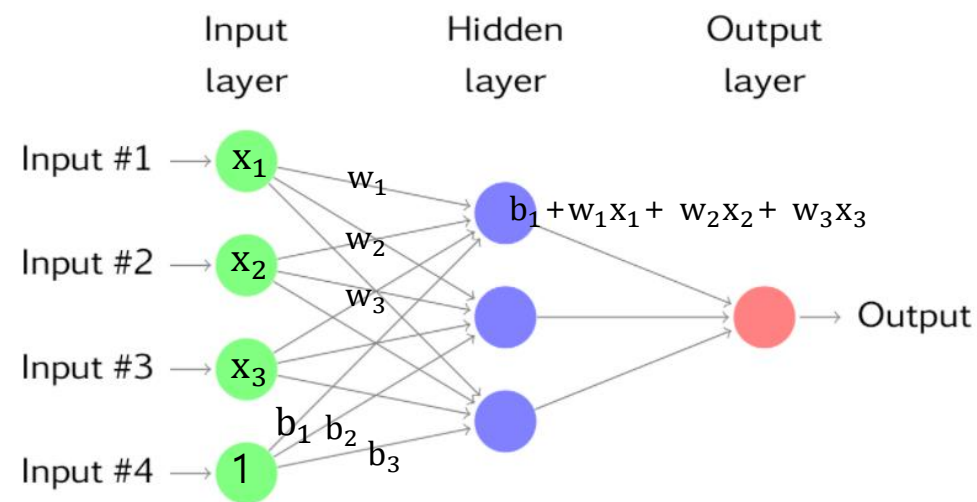
평균제곱오차(MSE)를 최소화하는 방향으로 모델 설계

◎ 규제 항목(sklearn 0.19 기준)

- max\_depth
- min\_samples\_split: 분할되기 위해 노드가 가져야 할 최소 샘플 수
- min\_samples\_leaf: 리프 노드가 가지고 있어야 할 최소 샘플 수
- min\_weight\_fraction\_leaf: min\_sample\_leaf와 동일하지만 가중치 부여된 전체 샘플 수에서의 비율
- max\_leaf\_nodes: 리프 노드의 최대 수
- max\_features: 각 노드에서 분할에 사용할 특성 최대 수
- min\_impurity\_decrease: 분할 대상이 되기 위해 필요한 최소한의 불순도



# 신경망 회귀



하이퍼파라미터	값
입력 뉴런 수	예측변수 개수
출력 뉴런 수	1
은닉층의 활성화 함수	ReLU(또는 SELU)
출력층의 활성화함수	없음 (출력이 양수) ReLU/ softplus (출력 범위 제한) logistic/ tanh
손실함수	MSE (이상치 존재 시) MAE/ Huber

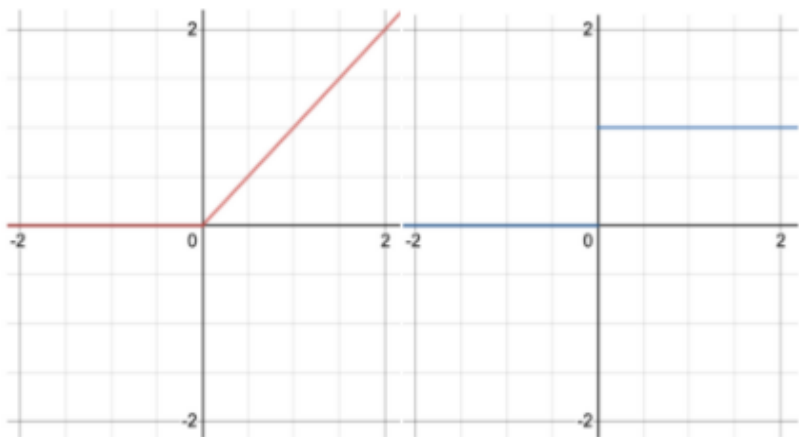
# 신경망 회귀

## ✓ 활성화함수

- SELU

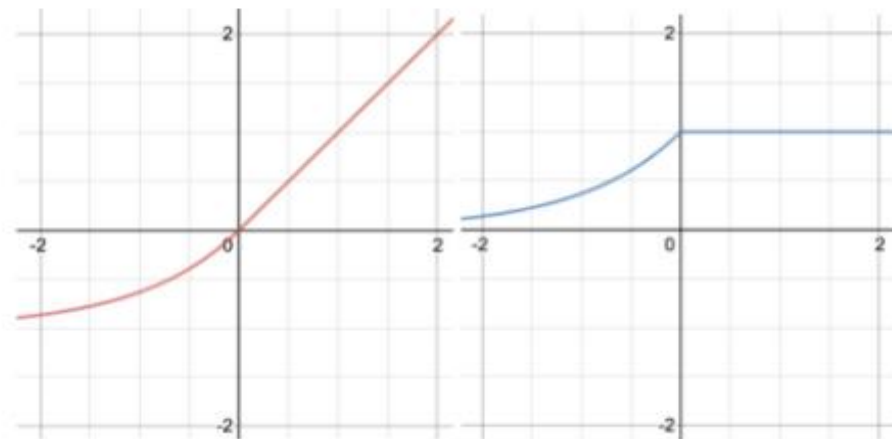
- RELU

$$f(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad f'(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

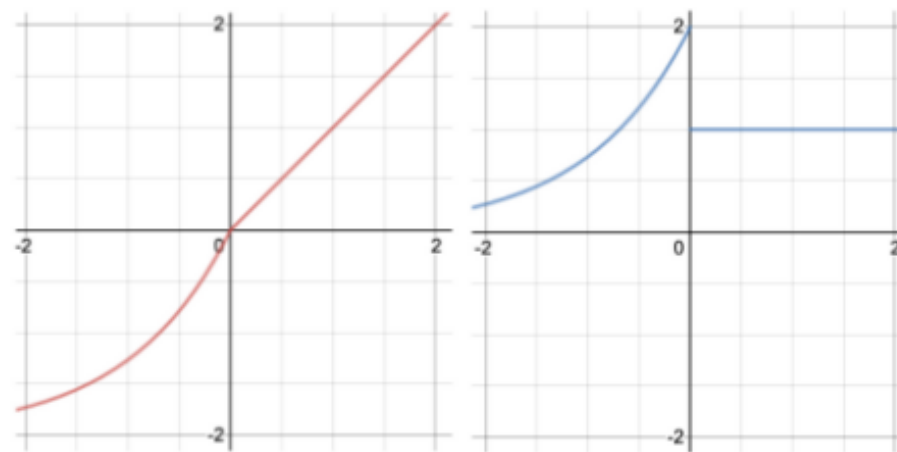


- ELU

$$f(\alpha, x) = \begin{cases} x & (x > 0) \\ \alpha(e^x - 1) & (x \leq 0) \end{cases} \quad f'(\alpha, x) = \begin{cases} 1 & (x > 0) \\ f(\alpha, x) + \alpha & (x \leq 0) \end{cases}$$



- SELU

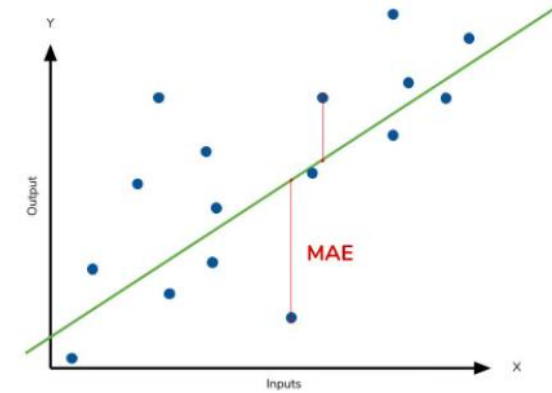


# 신경망 회귀

## ✓ 손실함수

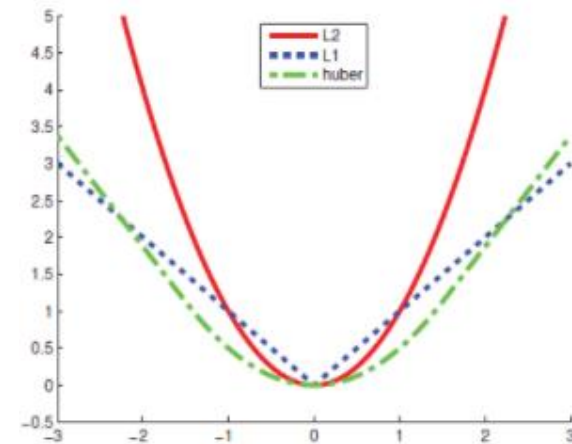
- 평균절대오차(MAE, Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$



- Huber

$$L_H(\gamma, \delta) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{if } |y - f(x)| \leq \delta \\ \delta|y - f(x)| - \delta^2/2 & \text{otherwise} \end{cases}$$



# 정리

