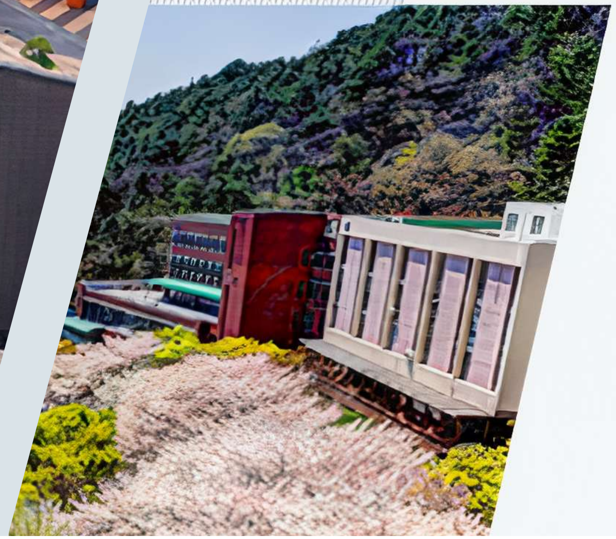
Dong-A Univ. (ISPL)

Linear Regression 이론

컴퓨터AI공학부 컴퓨터공학과
머신러닝



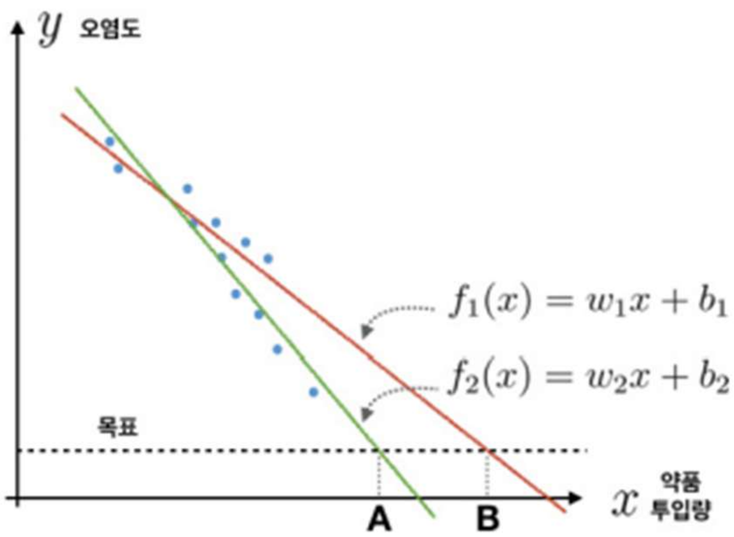
Linear Regression

- **회귀 (Regression)**
  - 관측된 데이터를 통해 변수 사이의 숨어있는 관계를 추정하는 것
  - 변수 y를 오염도, x를 약품 투입량으로 가정했을 때, **기존의 데이터를 이용하여 $y = wx + b$ 수식 생성**

변수 x	변수 y
사람의 키	사람의 몸무게
주택의 크기	주택의 가격
공부 시간	시험 점수
약품 투입량	오염도
…	…

Linear Regression

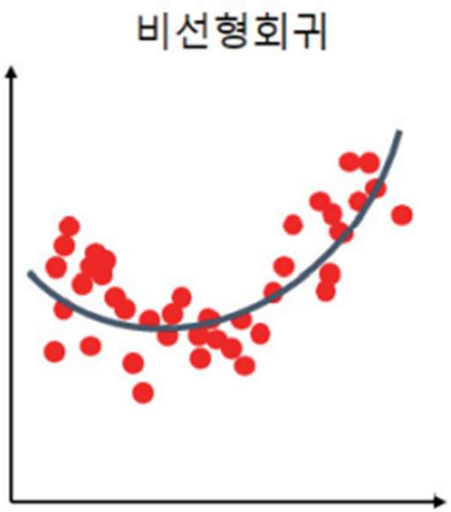
- **회귀 (Regression)**
  - 관측된 데이터를 통해 변수 사이의 숨어있는 관계를 추정하는 것
  - 변수 y를 오염도, x를 약품 투입량으로 가정했을 때, **기존의 데이터를 이용하여** $\boldsymbol{y = wx + b}$ **수식 생성**
  - 이후 신규 데이터를 이용하여 오염도 **예측**



회귀 분석을 통해 두 개의 수식 중
오차가 더 작은 수식 탐색

Linear Regression

- **선형 회귀 (Linear Regression)**
  - 선형회귀: 변수 사이의 숨어있는 관계를 1차원 수식으로 표현 ($\boldsymbol{y = wx + b}$)
  - 비선형회귀: 변수 사이의 숨어있는 관계를 2차원 이상의 수식으로 표현 ($\boldsymbol{y = w_1 x^n + w_2 x^{n-1} \ldots w_n x + b}$)

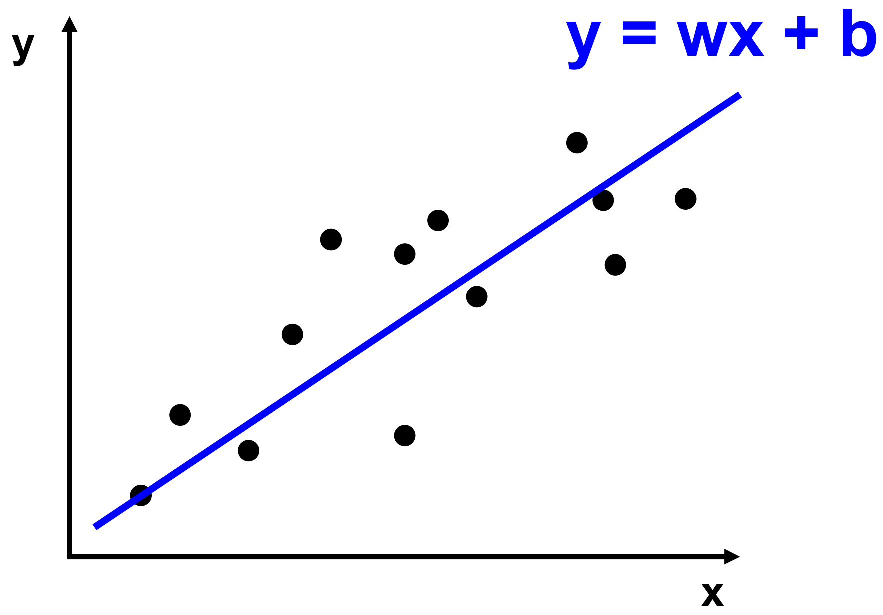


Linear Regression

- **문제 해결방법**
  - Ordinary Least Squares method (OLS)

    = Least Squares Method (LSM)

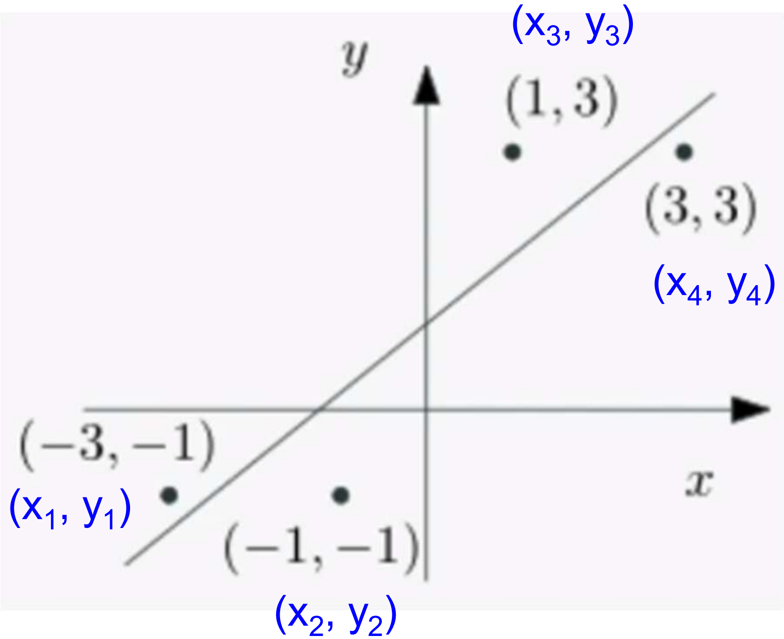
    = Normal Equation
  - Gradient Descent method (GD)



단일 선형회귀 예시

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 θ (w, b) 탐색)



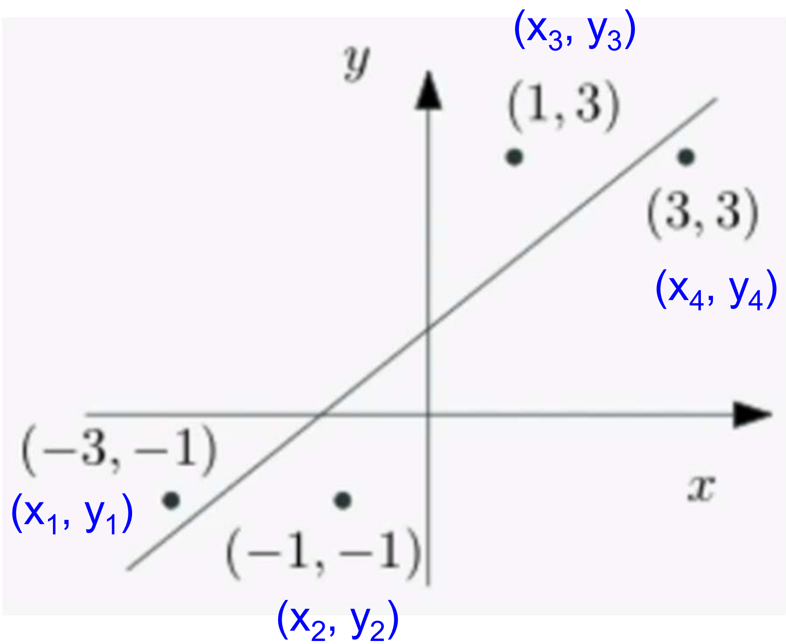
$$X = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \end{bmatrix} \qquad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \qquad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x1 4x1 2x1

동아대학교 DONG-A UNIVERSITY

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 $\theta\ (w, b)$ 탐색)



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

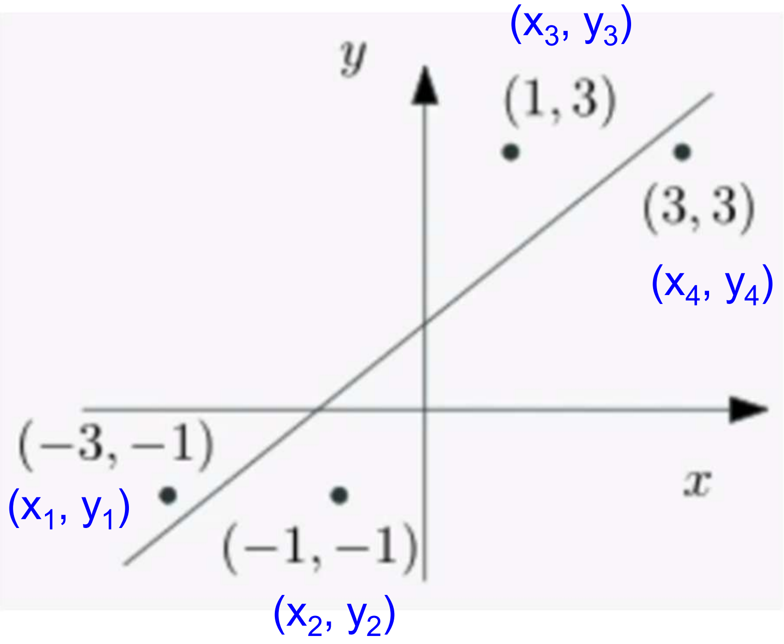
4x2 4x1 2x1

$$X\theta = \hat{Y}$$

❖ *최종목표: $Y = \hat{Y}$ 가 되는 θ 탐색*

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 $\theta\ (w, b)$ 탐색)



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x2 4x1 2x1

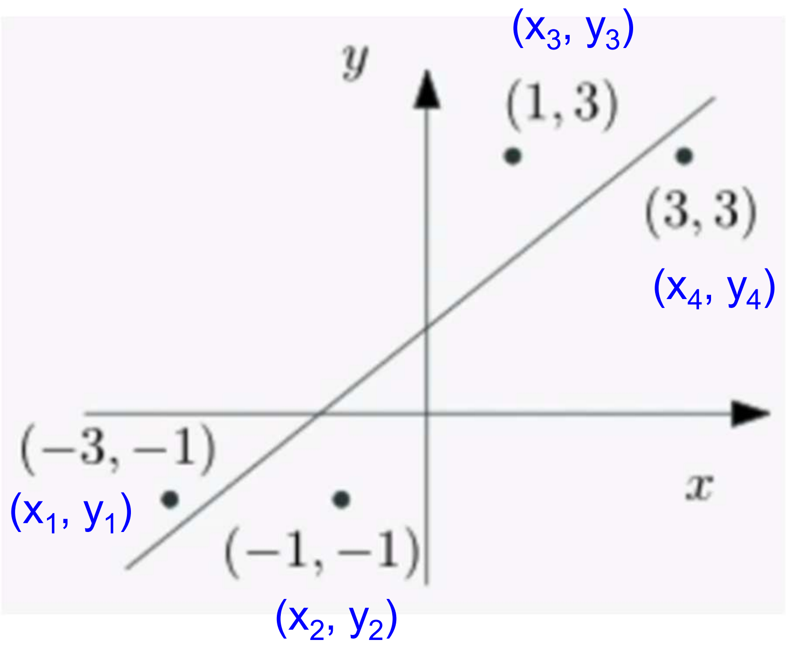
$X\theta = \hat{Y}$ ❖ *최종목표: $Y = \hat{Y}$ 가 되는 θ 탐색*

$\theta = X^{-1} \cdot Y$ → X가 정방행렬이 아니기 때문에 불가능

동아대학교 DONG-A UNIVERSITY

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 θ (w, b) 탐색)



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x2 4x1 2x1

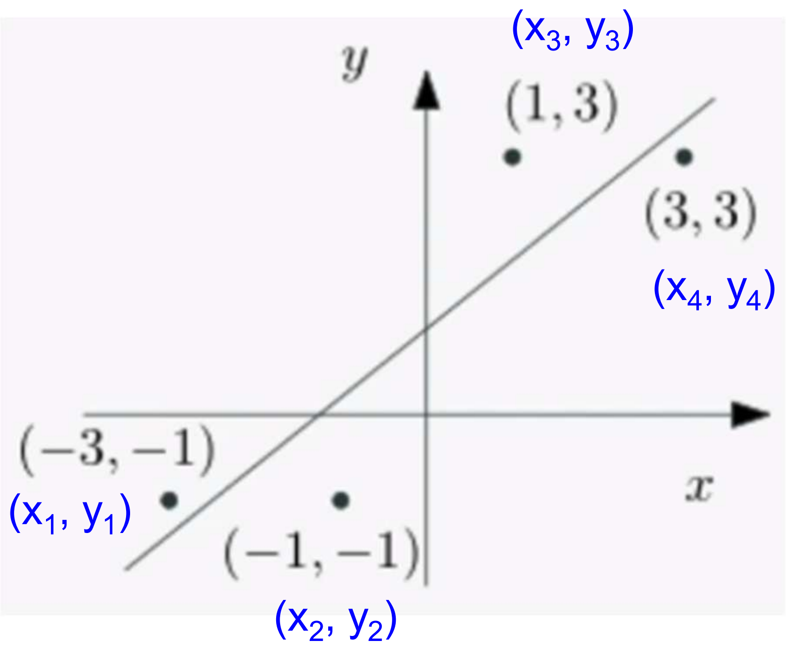
$$X\theta = \hat{Y}$$

❖ *최종목표: $Y = \hat{Y}$ 가 되는 θ 탐색*

$$(X^T \cdot X)\theta = X^T \cdot Y$$

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 θ (w, b) 탐색)



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x2 4x1 2x1

$$X\theta = \hat{Y}$$

❖ *최종목표: $Y = \hat{Y}$ 가 되는 θ 탐색*

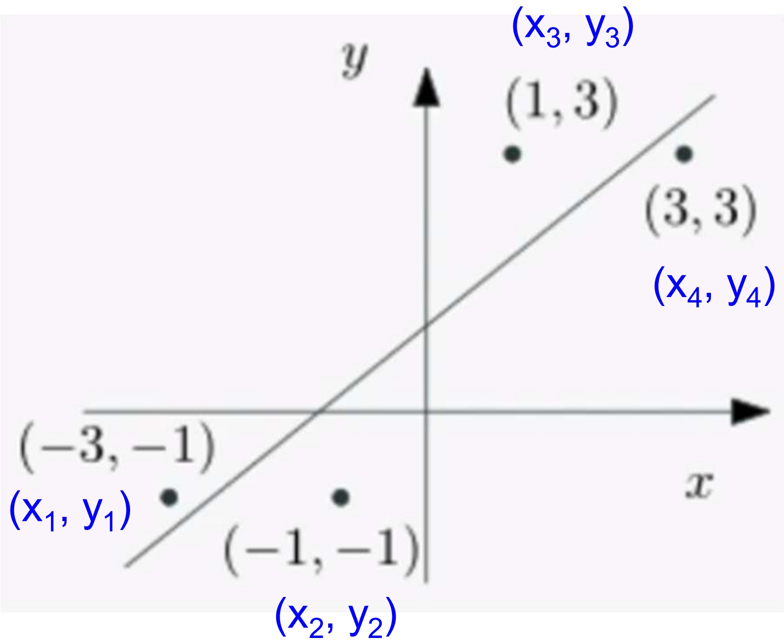
$$(X^T \cdot X)\theta = X^T \cdot Y$$

$$\cancel{(X^T \cdot X)^{-1} \cdot (X^T \cdot X)}\theta = (X^T \cdot X)^{-1} \cdot (X^T \cdot Y)$$

→ $(X^T \cdot X)^{-1}$ 를 양변에 곱하여 θ 계산

Linear Regression

- **Ordinary Least Squares method (OLS)**
  - 아래 예시 data를 이용하여 선형회귀 진행 (최적의 θ (w, b) 탐색)



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

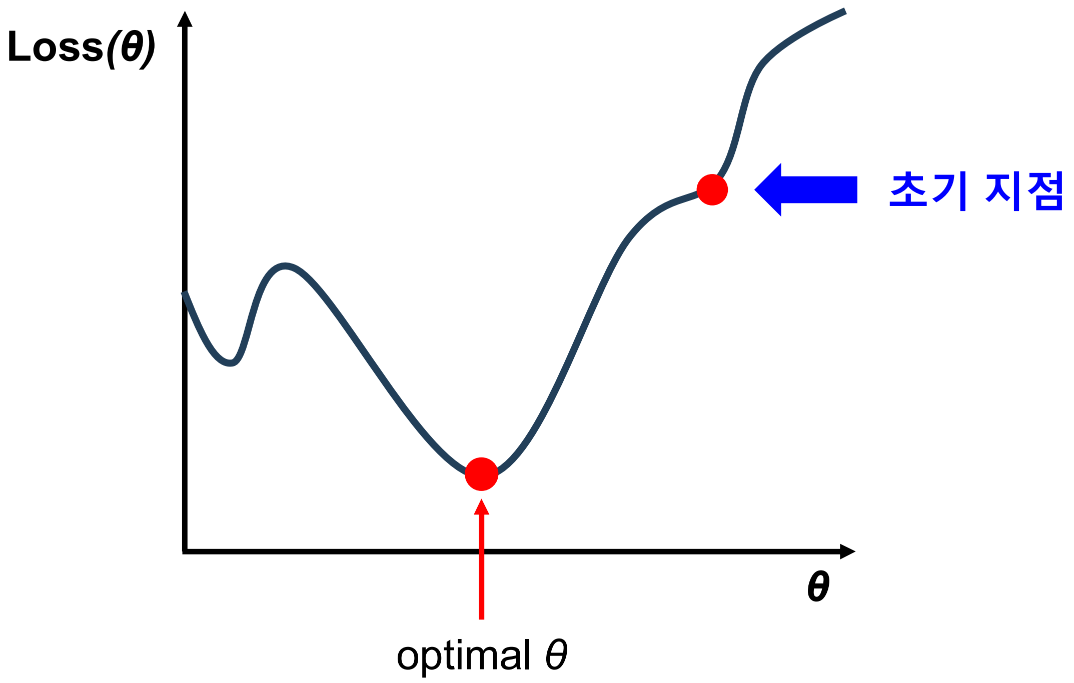
4x2 4x1 2x1

$$\theta = (X^T \cdot X)^{-1} \cdot (X^T \cdot Y)$$

$$\theta = \begin{bmatrix} 0.8 \\ 1 \end{bmatrix}$$

Linear Regression

- **Gradient Descent method (GD)**
  - 예측 값($\hat{Y}$) 과 정답 값(Y) 간의 차이를 이용하여 θ 를 업데이트하며 최적의 θ 를 탐색하는 방법



bias

$$X = \begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x2 4x1 2x1

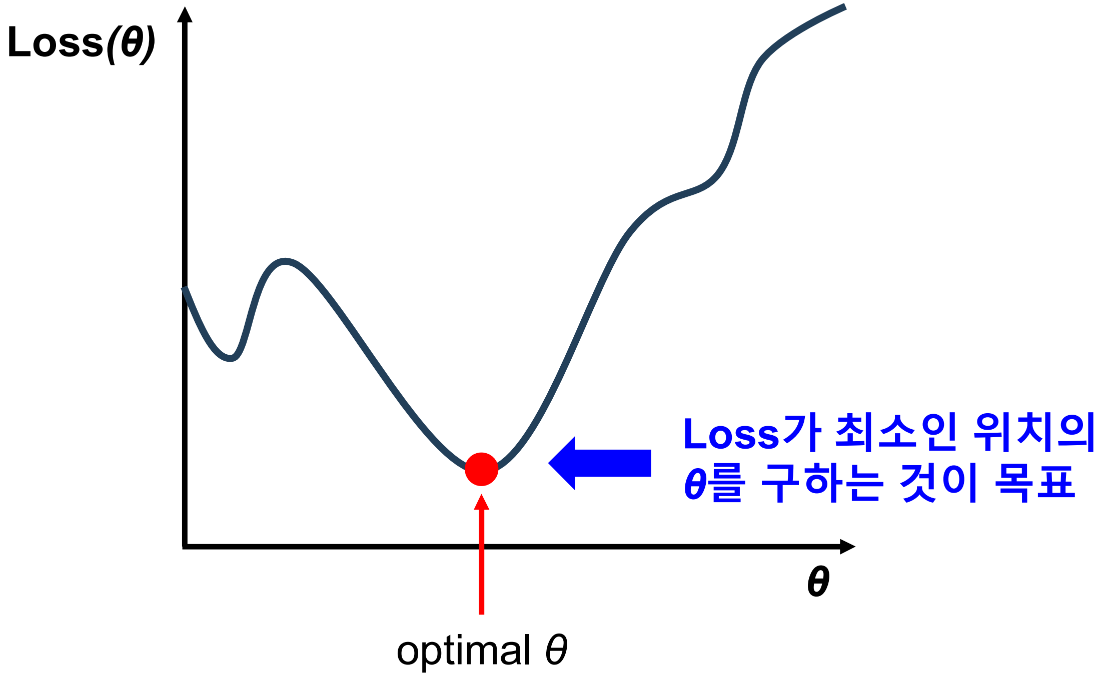
$$X\theta = \hat{Y}$$

$$Loss = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2$$

❖ *평균제곱오차 (Mean Square Error (MSE))*

동아대학교 DONG-A UNIVERSITY

- **Gradient Descent method (GD)**
  - 예측 값($\hat{Y}$) 과 정답 값(Y) 간의 차이를 이용하여 θ 를 업데이트하며 최적의 θ 를 탐색하는 방법



$$X = \begin{array}{cc} & \text{bias} \\ \multicolumn{2}{c}{\begin{bmatrix} -3 & 1 \\ -1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}} \end{array} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} w \\ b \end{bmatrix}$$

4x2   4x1   2x1

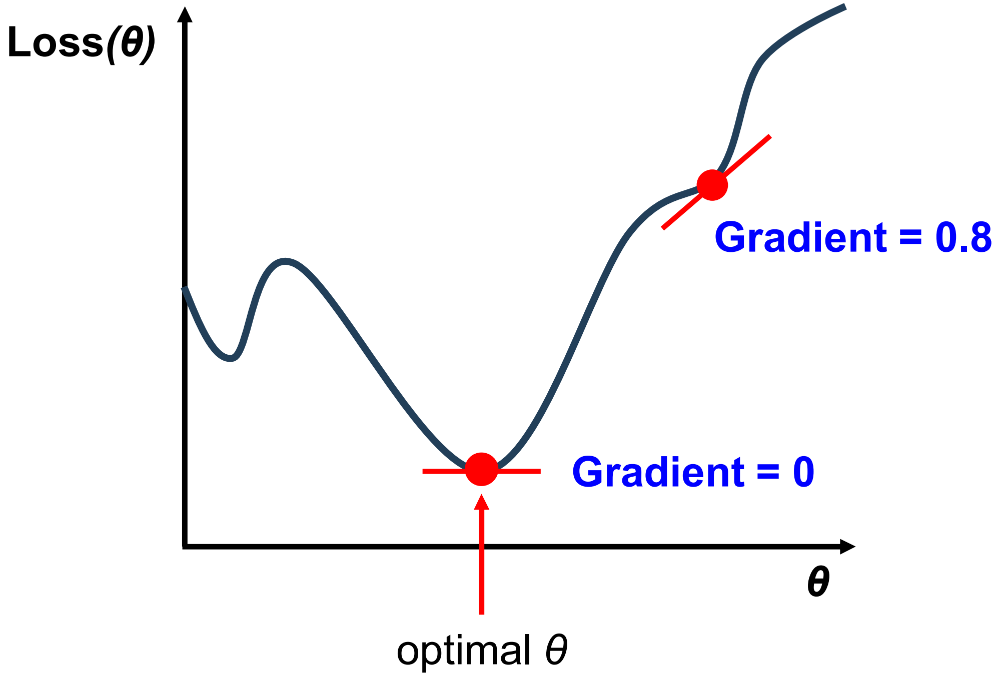
$$X\theta = \hat{Y}$$

$$Loss = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2$$

❖ *평균제곱오차 (Mean Square Error (MSE))*

Linear Regression

- **Gradient Descent method (GD)**
  - 현재 지점에서 Loss 값을 θ 에 대한 편미분을 통해 gradient 계산
  - Gradient에 learning rate를 곱하고 반대방향으로 weight 업데이트

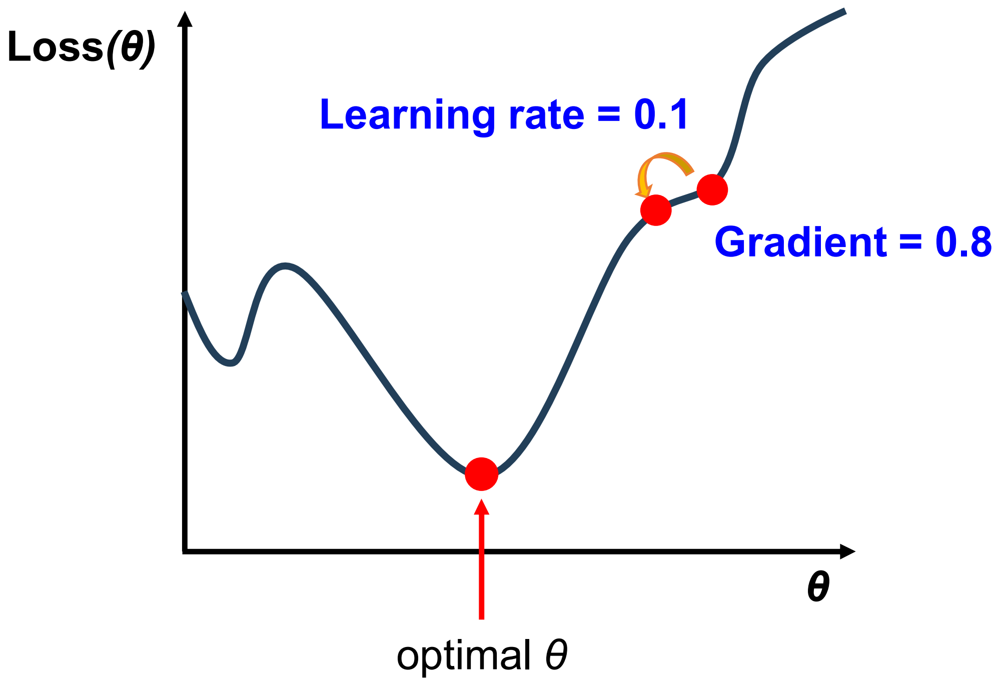


$$\theta_{t+1} = \theta_t - \underbrace{\boxed{\alpha}}_{\text{Learning rate}} \underbrace{\boxed{\frac{\partial L}{\partial \theta_t}}}_{\text{Gradient}}$$

동아대학교 DONG-A UNIVERSITY

Linear Regression

- **Gradient Descent method (GD)**
  - 현재 지점에서 Loss 값을 θ 에 대한 편미분을 통해 gradient 계산
  - Gradient에 learning rate를 곱하고 반대방향으로 weight 업데이트

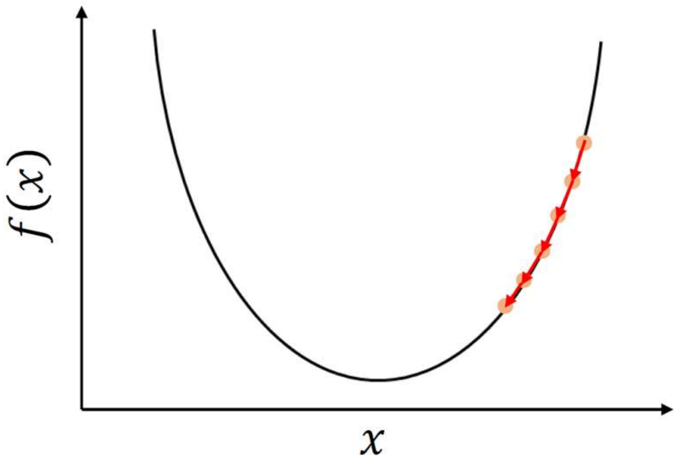


$$\theta_{t+1} = \theta_t - \alpha \frac{\partial L}{\partial \theta_t}$$

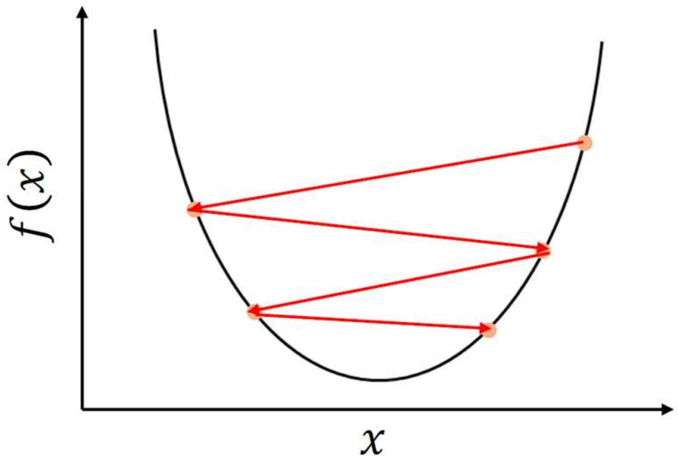
$$\boxed{= \theta_t - 0.08}$$

Linear Regression

- **Gradient Descent method (GD)**
  - 현재 지점에서 Loss 값을 θ 에 대한 편미분을 통해 gradient 계산
  - Gradient에 learning rate를 곱하고 반대방향으로 weight 업데이트
    - Learning rate: 파라미터를 얼마나 업데이트할 지 정하는 **하이퍼파라미터**



α가 너무 작은 경우



α가 너무 큰 경우

❖ α: Learning rate

Linear Regression

- **Gradient Descent method (GD)**

$$\hat{Y} = wx + b$$

$$Loss = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2$$

$$= \frac{1}{N}\sum(Y - wx_i - b)^2$$

→ $\hat{Y} = wx + b$ 대입

Linear Regression

- **Gradient Descent method (GD)**

$$\hat{Y} = wx + b$$

$$Loss = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2$$

$$= \frac{1}{N}\sum(Y - wx_i - b)^2$$

$$\frac{\partial L}{\partial w} = \frac{1}{N} \times 2 \times \sum(Y - wx_i - b) \times -x_i$$

$$\approx \frac{2}{N}\sum\left(Y - \hat{Y}\right) \times -X$$

→ w에 대한 편미분

Linear Regression

- **Gradient Descent method (GD)**

$$\hat{Y} = wx + b$$

$$Loss = \frac{1}{N}\sum(Y - \hat{Y})^2$$

$$= \frac{1}{N}\sum(Y - wx_i - b)^2$$

$$\frac{\partial L}{\partial w} = \frac{1}{N} \times 2 \times \sum(Y - wx_i - b) \times -x_i$$

$$\approx \frac{2}{N}\sum(Y - \hat{Y}) \times -X$$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \times 2 \times \sum(Y - wx_i - b) \times -1$$

$$\approx \frac{2}{N}\sum(Y - \hat{Y}) \times -1$$

→ b에 대한 편미분

Linear Regression

- **Gradient Descent method (GD)**

$$\hat{Y} = wx + b$$

$$Loss = \frac{1}{N}\sum(Y - \hat{Y})^2$$

$$= \frac{1}{N}\sum(Y - wx_i - b)^2$$

$$w_{t+1} = w_t - \alpha \times \frac{\partial \mathrm{L}}{\partial \mathrm{w}}$$

$$b_{t+1} = b_t - \alpha \times \frac{\partial \mathrm{L}}{\partial \mathrm{b}}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N} \times 2 \times \sum(Y - wx_i - b) \times -x_i$$

$$\approx \frac{2}{N}\sum(Y - \hat{Y}) \times -X$$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \times 2 \times \sum(Y - wx_i - b) \times -1$$

$$\approx \frac{2}{N}\sum(Y - \hat{Y}) \times -1$$

Questions & Answers

Dongsan Jun (dsjun@dau.ac.kr)
Image Signal Processing Laboratory (www.donga-ispl.kr)
Dept. of Computer Engineering
Dong-A University, Busan, Rep. of Korea