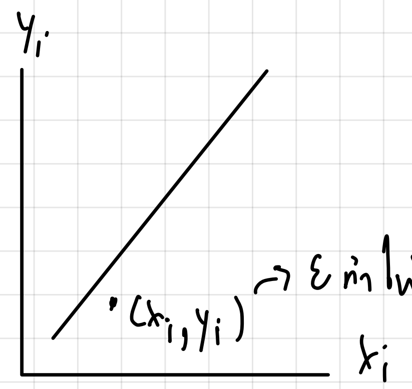
Linear Regression (LR)



$\to$ ε ทำให้ไม่อยู่กับเส้นเพราะมี noise

- Assume $y \in \mathbb{R}$
- LR's assumption $y_1 = \underbrace{w^T x_i}_{\text{Line}} + \underbrace{\varepsilon_i}_{\text{noise}}$

where $\varepsilon_i \sim N(\mu = 0, \sigma)$ (guasian distribution)

- Equivalent assumption: $y_i \sim N(w^T x_i, \sigma)$

$$\Rightarrow P(y_1 | x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w^T x_i - y_i)^2}{2\sigma^2}}$$

Goal: estimate w that define the linear function

Estimate W (หลาย Dimension)

ในกรณี Dimension เดียว

$y_i = mx_i + b + \varepsilon_i$

- $$W_{MLE} = \underset{w}{\text{argmax}} \prod_{i=1}^{n} P(y_i | x_i ; w)$$

$$= \underset{w}{\text{argmax}} \sum_{i=1}^{n} \log_e \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w^T x_i - y_i)^2}{2\sigma^2}} \right)$$

ไม่เกี่ยว W

$$= \underset{w}{\text{argmax}} \sum_{i=1}^{n} \cancel{-\log\sqrt{2\pi\sigma^2}} + \frac{-(w^T x_i - y_i)^2}{2\sigma^2}$$

way to find W_{MLE}: ① "Gradient Descent" on $\nabla l(w)$

② Derive closed-form solution to W_{MLE}

$$= \underset{w}{\text{argmin}} \sum_{i=1}^{n} \left(\frac{(w^T x_i - y_i)^2}{2\sigma^2} \right)$$

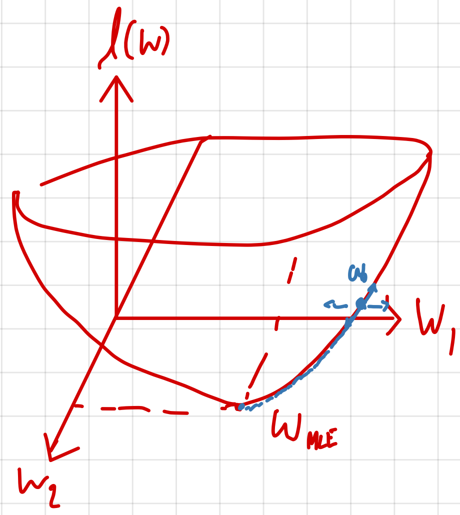
เอาค่าคงที่ออก

$$W_{MLE} = \underset{w}{\text{argmin}} \sum_{i=1}^{n} \underbrace{(w^T x_i - y_i)^2}_{h(x) - y \Rightarrow \text{square loss } l(w)}$$

$$= \boxed{\underset{w}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} (w^T x_i - y_i)^2}$$

วิธีที่ 1 gradient Descent

$$W_{t+1} = W_t - \alpha \nabla l(W)$$



$$l(W) = \frac{1}{n}\sum_{i=1}^{n}(W^T x_i - y_i)^2$$

$$\nabla l(W) = \begin{bmatrix} \frac{\partial l}{\partial w_1} \rightarrow \frac{dl}{dw_1} \\ \frac{\partial l}{\partial w_2} \rightarrow \frac{dl}{dw_2} \end{bmatrix}$$ eg. $W \in \mathbb{R}^2$ $W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$$\frac{\partial l}{\partial w_1} = \frac{\partial}{\partial w_1}\left(\frac{1}{n}\sum_{i=1}^{n}(W^T x_i - y_i)^2\right) = \frac{1}{n}\frac{\partial}{\partial w_1}\left(\sum_{i=1}^{n}(W^T x_i - y_i)^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial w_1}\left[(W^T x_i - y_i)(W^T x_i - y_i)\right]$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial w_1}(W^T x_i - y_i)\right](W^T x_i - y_i)$$

$$W^T x_i - y_i = w_1 \overset{\text{① coordinate}}{x_i^1} + w_2 x_i^2 - y_i$$

$$\frac{\partial (W^T x_i - y_i)}{\partial w_1} = x_i^1$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial w_1}(w_1 x_i^1 + w_2 x_i^2 - y_i)\right](W^T x_i - y_i)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left[x_i^1\right](W^T x_i - y_i)$$

$$\frac{\partial l}{\partial w_2} = \frac{2}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial w_2}(W^T x_i - y_i)\right](W^T x_i - y_i)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left[x_2^1\right](W^T x_i - y_i)$$

$$\nabla l(W) = \begin{bmatrix} \frac{\partial l}{\partial w_1} \\ \frac{\partial l}{\partial w_2} \end{bmatrix} = \begin{bmatrix} \frac{2}{n}\sum_{i=1}^{n} x_i^1 (W^T x_i - y_i) \\ \frac{2}{n}\sum_{i=1}^{n} x_i^2 (W^T x_i - y_i) \end{bmatrix}$$

$$\boxed{W_{t+1} = W_t - \alpha \begin{bmatrix} \frac{2}{n}\sum_{i=1}^{n} x_i^1 (W^T x_i - y_i) \\ \frac{2}{n}\sum_{i=1}^{n} x_i^2 (W^T x_i - y_i) \end{bmatrix}}$$

//

วิธีที่ 2 Closed-form W_{MLE} :

claim : $W_{MLE} = (X^TX)^{-1}X^T\vec{y}$ (X คือ matrix) $W = (X^TX)^{-1}Xy^T$

$$X = \begin{bmatrix} \longleftarrow x_1^T \longrightarrow \\ \longleftarrow x_2^T \longrightarrow \\ \vdots \\ \longleftarrow x_n^T \longrightarrow \end{bmatrix} \text{ and } \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Note $X\vec{w} - \vec{y} = \underbrace{\begin{bmatrix} \longleftarrow x_1^T \longrightarrow \\ \longleftarrow x_2^T \longrightarrow \\ \vdots \\ \longleftarrow x_n^T \longrightarrow \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

$$\begin{bmatrix} w^Tx_1 \\ w^Tx_2 \\ \vdots \\ w^Tx_n \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w^Tx_1 - y_1 \\ w^Tx_2 - y_2 \\ \vdots \\ w^Tx_n - y_n \end{bmatrix}$$

$$(Xw - y)^T(Xw - y) = \begin{bmatrix} w^Tx_1 - y_1 & w^Tx_2 - y_2 & \cdots & w^Tx_n - y_n \end{bmatrix} \begin{bmatrix} w^Tx_1 - y_1 \\ w^Tx_2 - y_2 \\ w^Tx_n - y_n \end{bmatrix}$$

$$= (w^Tx_1 - y_1)^2 + (w^Tx_2 - y_2)^2 + \ldots + (w^Tx_n - y_n)^2$$

$$= \sum_{i=1}^{n}(w^Tx_i - y_i)^2 = l(w) \quad ; \quad \sum_{i=1}^{n}(X_i)^2 = (X_1)^2 + (X_2)^2 + \ldots + (X_n)^2$$

$$\nabla_w l(w) = \nabla_w\left((Xw - y)^T(Xw - y)\right)$$

$$\nabla_w\left((X^Tw^T - y^T)(Xw - y)\right)$$

$$\nabla_w\left(\underbrace{X^Tw^TXw}_{w^2X^TX} - y^TXw - X^Tw^Ty + \cancel{y^Ty}^{\,0}\right) \qquad \text{assume } w^Tw = w^2$$

$$= \left(2X^TXw - y^TX - X^Ty\right)$$

$$= 2X^TXw - 2y^TX$$

To find W_{MLE}; $\nabla l(W) = 0$

$$\vec{0} = 2(X^T X W - Y^T X)$$

$$X^T X W = Y^T X$$

$$\boxed{W = (X^T X)^{-1} X Y^T}$$

Note: $X \in \mathbb{R}^{n \times d} \Rightarrow X^T X \in \mathbb{R}^{d \times d}$

- compute inverse of $X^T X \Rightarrow O(d^3)$
- If d is large, then the closed-form solution is not possible to compute
- closed from ดีถ้า d เล็ก

Logistic Regression

- $W_{MLE} = \underset{W}{\arg\min} \sum_{i=1}^{n} \log(1 + e^{-y_i W^T x_i})$
- $W_{map} = ?$
- $W_{MLE} = \underset{W}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (W^T x_i - y_i)^2$

W_{map} (Ridge regression)

Recall: $P(w|D) \propto P(D|W) \cdot P(w)$

Assume $w \sim N(\mu = 0, \sigma)$

$$W_{map} = \underset{w}{\arg\max}\ P(w|D)$$

$$= \underset{w}{\arg\max}\ P(D|W) \times P(w)$$

$$= \underset{w}{\arg\max}\ \log_e P(D|W) + \log_e P(w)$$

$$= \underset{w}{\arg\min} \sum_{i=1}^{n} \log(1 + e^{y_i W^T x_i}) + \lambda w^T w$$

where $\frac{1}{2\sigma^2} = \lambda$ ← กำหนดเอง

$$= \boxed{\underset{W}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (W^T x_i - y_i)^2 + \frac{1}{2\sigma^2} w^T w}$$

(Gradient Desent (closed form)