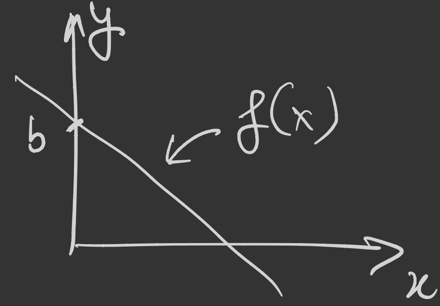
$y, x \qquad f(x) = \underset{\uparrow \text{ slope}}{a}x + \underset{\uparrow \text{ intercept/bias}}{b}$



$x_1 \quad y_1 \quad f(x_1) = ax_1 + b$

$\vdots$

$x_7 \quad y_7 \quad f(x_7) \ldots$

$y - f(x)$: residual

$$L(a,b) = \sum_n (y_n - \underbrace{f(x_n)}_{ax_n + b})^2$$

$$L(a,b) = \sum_n (y_n - ax_n - b)^2$$

$$① \quad \frac{dL}{da} = -2\sum_n (y_n - ax_n - b)\cdot x_n = -2\Big[\underbrace{\sum_n y_n x_n}_{S_{xy}} - a\underbrace{\sum_n x_n^2}_{S_{x^2}} - b\underbrace{\sum_n x_n}_{S_x}\Big]$$

$$② \quad \frac{dL}{db} = -2\sum_n (y_n - ax_n - b) = -2\Big[\underbrace{\sum_n y_n}_{S_y} - a\underbrace{\sum_n x_n}_{S_x} - Nb\Big]$$

$a, b \leftarrow \operatorname{argmin} L(a,b)$

$\frac{dL}{da} = 0 \quad \& \quad \frac{dL}{db} = 0$

$$② \Rightarrow \boxed{b = \frac{S_y - aS_x}{N}} \quad ③$$

$$\text{Sub } ③ \text{ into } ① \rightarrow \boxed{a = \frac{NS_{xy} - S_x S_y}{NS_{x^2} - S_x^2}}$$

$$a = -0.095 \qquad b = 3.0$$

$$\cancel{X} \begin{bmatrix} x_1 \\ \vdots \\ x_7 \end{bmatrix}$$

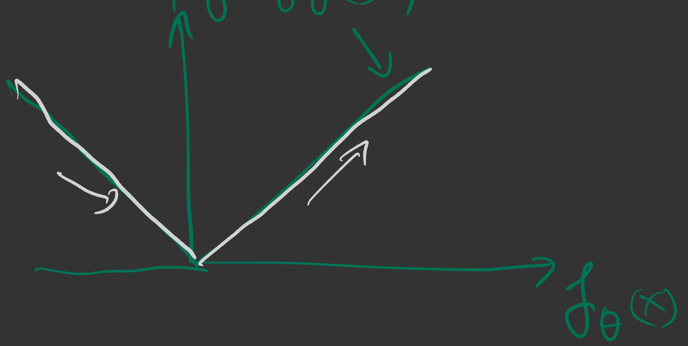
$$\hat{X} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_7 & 1 \end{bmatrix}$$

$$f(x) = ax + b \qquad f(x) = \theta^T \cdot x$$

$$= \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}^T}_{\theta} \cdot \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}}_{\hat{x}}$$

$$|y - f_\theta(x)|$$



y	$f_\theta(x)$	$(y - f_\theta(x))^2$
3	2	$1^2 = 1$
-3	-5	$2^2 = 4$
-10	-5	$5^2 = 25$

$$L(\theta) = \frac{1}{N}(y - X\theta)^T \cdot (y - X\theta) + \lambda \|\theta\|_2^2 \qquad \boxed{\min_{\theta} L(\theta)} \leftarrow$$

$$= \frac{1}{N}\left(y^T y - \boxed{2 y^T X \theta} + \theta^T X^T X \theta\right) + \lambda \|\theta\|_2^2$$

$$\frac{dL(\theta)}{d\theta} = \frac{1}{N}\left(0 - 2 y^T X + 2 \theta^T X^T X\right) + 2\lambda \theta^T$$

$$\frac{dL(\theta)}{d\theta} = 0 \Rightarrow [y^T X]^T = [\theta^T X^T X]^T + \lambda N \cdot \theta$$

$$N\lambda\theta \Rightarrow X^T X \theta = X^T \cdot y$$

$$\underbrace{(X^T X)^{-1} \cdot (X^T X)}_{I} \theta = (X^T X)^{-1} \cdot X^T \cdot y$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

$$\theta = (X^T X + N\lambda I)^{-1} X^T y$$

$$(y - X\theta)^T \cdot (y - X\theta)$$

$$= y^T y - \underbrace{(X\theta)^T y}_{①} - \underbrace{y^T X \theta}_{②} + (X\theta)^T \cdot X\theta$$

$$= y^T y - 2 y^T X \theta + \theta^T X^T X \theta$$

$$(y^T X \theta)^T = \theta^T X^T y$$

Both ① and ② are scalar

$①^T = ② / ②^T = ①$

$\Rightarrow$ same

$$\arg\max_{\theta} \underline{p}(D|\theta)\, p(\theta) \qquad \mathcal{N}(\theta;\ \underline{0},\ \sigma_0^2 \underline{\underline{I_D}})$$

$$\uparrow \qquad \uparrow$$

$$D \times D$$

$$= \arg\max_{\theta} \underline{\log p(D|\theta)} + \log p(\theta)$$

$$+\frac{1}{2\sigma^2}(y - X\theta)^T(y - X\theta) + \frac{\|\theta\|_2^2}{2\sigma_0^2}$$

Reg

$$\frac{1}{N}(y - X\theta)^T(y - X\theta) + \lambda \|\theta\|_2^2$$

$$1 \quad \simeq \frac{1}{N}\left[(y - X\theta)^T(y - X\theta) + N\lambda \|\theta\|_2^2\right]$$

$$2 \quad \simeq \frac{1}{2\sigma^2}\left[(y - X\theta)^T(y - X\theta) + \frac{\sigma^2}{\sigma_0^2}\|\theta\|_2^2\right]$$

$$1 \equiv 2 \qquad \frac{\sigma^2}{\sigma_0^2} = N\lambda$$

$$\boxed{\lambda = \frac{\sigma^2}{\sigma_0^2 N}}$$

$A = X^T X$ rank(A) = D

$D \times D$ positive definite

A is positive semi-definite $\boxed{Xy}$

$$y^T A y = y^T X^T X y = (yX)^T Xy \geq 0$$

$D \times 1$, $y \neq 0$

an eigenvalue of $A = 0$

aka A is not invertible

$B \quad y^T B y > 0$

α : small, > 0

"jitter"

$$B = A + \alpha I$$

$$y^T B y = (yX)^T Xy + \alpha y^T y$$

$\geq 0 \qquad > 0$

> 0

aka B is invertible

$\|\theta\|_2^2$

$$L(\theta) = \sum_{n=1}^{N} l_{n,\theta}(x_n, y_n) + \lambda \, \underline{Reg(\theta)}$$

Regularizer

$$\frac{dL(\theta)}{d\theta} \longrightarrow \underline{GD}$$

$\underline{LR}$: $l_n(x_n, y_n) = (y_n - \theta^T x_n)^2$

When N is large, computing ∇ is expensive

$$L(\theta) \simeq \frac{N}{B} \sum_{b=1}^{B} l_b(x_b, y_b) + \lambda \, Reg$$

$B << N$

minibatching