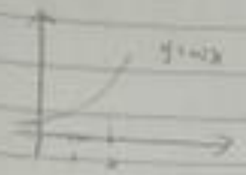
ML Hw4 B06110S2 武敬祥

1. [c]



$$\int_0^2 (e^x - wx)^2\, dx$$

$$= \int_0^2 e^{2x} - 2wxe^x + w^2x^2\, dx$$

$$= \int_0^2 e^{2x}\, dx - 2w\int_0^2 xe^x\, dx + \int_0^2 w^2x^2\, dx \quad \Rightarrow \int u\,dv = uv - \int v\,du$$

$$= \frac{1}{2}e^{2x}\Big|_0^2 - 2w(xe^x - e^x)\Big|_0^2 + w^2\frac{1}{3}x^3\Big|_0^2$$

令 $u = x$ $du = dx$

$v = e^x$ $e^x dx = dv$

$= xe^x - \int e^x dx$

$= xe^x - e^x$

$$= \frac{1}{2}e^4 - \frac{1}{2} - 2w(e^2+1) + w^2\frac{8}{3}$$

$$\frac{d\left(\frac{1}{2}e^4 - \frac{1}{2} - 2w(e^2+1) + w^2\frac{8}{3}\right)}{dw} = 0$$

$$= -2(e^2+1) + 2w\frac{8}{3} = 0$$

$$w = \frac{3}{8}(e^2+1)$$

$$\text{error} = \left| e^x - \left(\frac{3e^2+3}{8}\right)x \right|_x$$

2.

3. [c]

$$X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_N \\ | & & | \end{bmatrix}^T \qquad X_h = \begin{bmatrix} | & & | & | & & | \\ x_1 & \cdots & x_N & \tilde{x}_1 & \cdots & \tilde{x}_N \\ | & & | & | & & | \end{bmatrix}^T$$

$$X_h^T X_h = \underbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & \tilde{x}_N \\ | & & | \end{bmatrix}}_{d+1 \times 2N} \underbrace{\begin{bmatrix} - x_1 - \\ \vdots \\ - \tilde{x}_N - \end{bmatrix}}_{2N \times d+1} \qquad X^T X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_N \\ | & & | \end{bmatrix} \begin{bmatrix} - x_1 - \\ \vdots \\ - x_N - \end{bmatrix}$$

$$X^T X = \begin{bmatrix} \sum_{i=1}^{N} x_i^{1^2} & \cdots & \sum_{i=1}^{N} x_i^{1} x_i^{d+1} \\ \vdots & & \vdots \\ \sum_{i=1}^{N} x_i^{d+1} x_i^{1} & \cdots & \sum_{i=1}^{N} x_i^{d+1^2} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sum_{i=1}^{N} \left(x_i^{1^2} + \tilde{x}_i^{1^2}\right) & \cdots \\ \vdots & \ddots \\ \sum_{i=1}^{N} \left(x_i^{d+1^2} + \tilde{x}_i^{d+1^2}\right) & \cdots \end{bmatrix}}_{d+1 \times d+1}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} x_i^{1^2} & \cdots \\ \vdots & \ddots \\ \sum_{i=1}^{N} x_i^{d+1^2} & \cdots \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{N} \tilde{x}_i^{1^2} & \cdots \\ \vdots & \ddots \\ \sum_{i=1}^{N} \tilde{x}_i^{d+1^2} & \cdots \end{bmatrix}$$

$\tilde{x}_i = x_i + \varepsilon_i$

$\rightarrow \varepsilon_i =$

$$= X^T X + \begin{bmatrix} \sum_{i=1}^{N} \left(x_i^2 + \underline{2 x_i \varepsilon_i + \varepsilon_i^2}\right) & \cdots \\ \sum_{i=1}^{N} x_i^1 x_i^2 + [illegible] & \cdots \\ \vdots & \end{bmatrix} = X^T X + \left(X^T X + \sigma^2 I_{d+1}\right)$$

$$= 2X^T X + \sigma^2 I_{d+1}$$

4. [e]

$$X_h^T y_h = \begin{bmatrix} | & | & & | & & | \\ x_1 & x_2 & \cdots & x_n & \cdots & x_N \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} y_1 \\ y_n \\ y_1 \\ y_n \end{bmatrix}$$

$$X^T y = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_N \\ | & | & & | \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i^1 y_i & \cdots \\ \vdots & \\ \sum_{i=1}^{N} x_i^{d+1} y_i & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} (x_i^1 y_i + \tilde{x}_i^1 y_i) & \cdots \\ \vdots & \\ \sum_{i=1}^{N} (x_i^{d+1} y_i + \tilde{x}_i^{d+1} y_i) & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} x_i^1 y_i & \cdots \\ \vdots & \\ \sum_{i=1}^{N} x_i^{d+1} y_i & \cdots \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{N} \tilde{x}_i^1 y_i & \cdots \\ \vdots & \\ \sum_{i=1}^{N} \tilde{x}_i^{d+1} y_i & \cdots \end{bmatrix}$$

$$= X^T y + \begin{bmatrix} \sum_{i=1}^{N} (x_i + \epsilon_i) y_i & \cdots \\ \vdots & \\ \sum_{i=1}^{N} (x_i^{d+1} + \epsilon_i^{d+1}) y_i & \cdots \end{bmatrix} = X^T y + X^T y + \begin{bmatrix} \sum_{i=1}^{N} \epsilon_i y_i & \cdots \\ \vdots & \\ \sum_{i=1}^{N} \epsilon_i^{d+1} y_i & \cdots \end{bmatrix}$$

$$= 2X^T y + \begin{bmatrix} \epsilon_1 y_1 + \epsilon_2 y_2 + \cdots + \epsilon_N y_N \\ \vdots \end{bmatrix}$$

5.

$$X^T X = Q \Gamma Q^T$$

$$z_n = \Phi(x_n) = Q^T x_n$$

$$Z = Q^T X$$

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \| Zw - y \|^2 + \frac{\lambda}{N} w^T w$$

6. [a]

$$\min_{w \in R} \frac{1}{N}\sum_{n=1}^{N}(w_{REG}x_n - y_n)^2 + \frac{\lambda}{N} w_{REG}^2$$

$$\Updownarrow$$

$$\min_{w \in R} \frac{1}{N}\sum_{n=1}^{N}(w \cdot x_n - y_n)^2 \quad \text{with } w^2 \le C$$

$$\frac{2}{N}\sum_{n=1}^{N}(w_{REG}x_n - y_n)x_n + \frac{2\lambda}{N} w_{REG} = 0$$

$$\downarrow$$

$$\cancel{\frac{2}{N}} w_{REG}\sum_{n=1}^{N} x_n^2 - \cancel{\frac{2}{N}}\sum_{n=1}^{N} y_n x_n + \frac{\cancel{2}\lambda}{\cancel{N}} w_{REG} = 0$$

$$w_{REG}\left(\sum_{n=1}^{N} x_n^2 + \lambda\right) = \sum_{n=1}^{N} y_n x_n$$

$$w_{REG} = \frac{\sum_{n=1}^{N} y_n x_n}{\sum_{n=1}^{N} x_n^2 + \lambda}$$

$$w_{REG}^2 = w^{*2} = C.$$

$$C = \left(\frac{\sum_{n=1}^{N} y_n x_n}{\sum_{n=1}^{N} x_n^2 + \lambda}\right)^2$$

2. [d]

$$\frac{\left(\sum_{n=1}^{N} y_n\right) + K}{N + 2K} \rightarrow \text{optimal solution}$$

$$\min_{y \in R} \frac{1}{N}\sum_{n=1}^{N}(y-y_n)^2 + \frac{2K}{N}\Omega(y)$$

$$\frac{\partial \frac{1}{N}\sum_{n=1}^{N}(y-y_n)^2 + \frac{2K}{N}\Omega(y)}{\partial y} = 0$$

$$\frac{2}{N}\sum_{n=1}^{N}(y-y_n) + \frac{2K}{N}\Omega'(y) = 0$$

$$\sum_{n=1}^{N}(y-y_n) + K\Omega'(y) = 0$$

$$\Omega'(y) = -\frac{1}{K}\sum_{n=1}^{N}(y-y_n) = -\frac{1}{K}\sum_{n=1}^{N} y + \frac{1}{K}\sum_{n=1}^{N} y_n$$

$$= \frac{1}{K}\sum_{n=1}^{N} y_n - \frac{N}{K}y$$

$$y = \frac{\sum_{n=1}^{N} y_n + K}{N+2K} \Rightarrow y(N+2K) - K = \sum_{n=1}^{N} y_n$$

$$\Omega'(y) = \frac{1}{K}\left(y(N+2K) - K\right) - \frac{N}{K}y$$

$$= \frac{N+2K}{K}y - 1 - \frac{N}{K}y$$

$$= \cancel{\frac{Ny}{K}} + 2y - 1 - \cancel{\frac{N}{K}y}$$

$$= 2y - 1$$

$$\Omega(y) = (y-0.5)^2$$

8. [b]

$$\min_{\tilde{w} \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (\tilde{w}^T \Phi(x_n) - y_n)^2 + \frac{\lambda}{N} (\tilde{w}^T \tilde{w})$$

$$= \min_{\tilde{w} \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (\tilde{w}^T \Gamma^{-1} x_n - y_n)^2 + \frac{\lambda}{N} (\tilde{w}^T \tilde{w})$$

$$= \min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 + \frac{\lambda}{N} \Omega(w)$$

$\tilde{w}^T \cancel{\Gamma} = w^T \Gamma \Rightarrow \tilde{w}^T = w^T \Gamma$

$\tilde{w}^T \tilde{w} = \Omega(w)$

$$\Omega(w) = w^T \Gamma (w^T \Gamma)^T = w^T \underbrace{\Gamma}_{\text{diagonal}} \Gamma^T w = w^T \Gamma^2 w$$

9. [b]

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 + \frac{\lambda}{N} \sum_{i=0}^{d} \beta_i w_i^2$$

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N+K} \left(\sum_{n=1}^{N} (w^T x_n - y_n)^2 + \sum_{k=1}^{K} (w^T \tilde{x}_k - \tilde{y}_k)^2 \right)$$

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 + \frac{\lambda}{N} (w^T B w)$$

$$\frac{\partial \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 + \frac{\lambda}{N} (w^T B w)}{\partial w} = \frac{2}{N} (X^T X w - X^T y) + \frac{\lambda}{N} (2 B w)$$

$$\frac{\partial \frac{1}{N+K} \left(\sum_{n=1}^{N} (w^T x_n - y_n)^2 + \sum_{k=1}^{K} (w^T \tilde{x}_k - \tilde{y}_k)^2 \right)}{\partial w} = \frac{2}{N+K} (X^T X w - X^T y) + \frac{2}{N+K} (\tilde{X}^T \tilde{X} w - \tilde{X}^T \tilde{y})$$

$\frac{2}{N} (X^T X w - X^T y) + \frac{\lambda}{N} (2 B w) = 0$

$X^T y = (X^T X + \lambda B) w$

$w = (X^T X + \lambda B)^{-1} X^T y$

$X^T X w - X^T y + \tilde{X}^T \tilde{X} w - \tilde{X}^T \tilde{y} = 0$

$-\lambda B w + \tilde{X}^T \tilde{X} w - \tilde{X}^T \tilde{y} = 0$

$\tilde{X} = \sqrt{\lambda} \cdot \sqrt{B}, \quad \tilde{y} = 0$

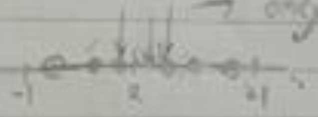
$\cancel{-\lambda B w} + \cancel{\lambda B w} - (\sqrt{\lambda} \sqrt{B})^T \tilde{y} = 0$

10. [e]

if we choose o as validation sample
we predict x as "correct"
Error = 1.
if we choose x as validation sample
we predict o as "correct"
Error = 1.

11. [c]

→ only this two may possibly generate error



$E_{loocv} = \frac{2}{N}$

1-[e]

$E_{loocv} = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}(e_1+e_2+e_3)$

$e_1 = 1^2 \quad e_2 = 1^2 \quad e_3 = 2^2$

$E_{loocv_1} = \frac{1}{3}(1+1+4) = 2$

$E_{loocv_2} = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}(e_1+e_2+e_3)$

$\frac{-2}{3+\rho}$ $\quad -2x-(\rho+3)y=-6$

$2x-(\rho+3)y=-6$

$\frac{2}{\rho+3}x + \frac{6}{\rho+3}$ $\quad \frac{2x+6}{(\rho+3)} = y$

$h(x) = w_0 + w_1 x \quad w_0 = \frac{6}{\rho+3}, \quad w_1 = \frac{2}{\rho+3}$

$e_1 = \left(\frac{6}{\rho+3} + \frac{6}{\rho+3} - 0\right)^2 = \left(\frac{12}{\rho+3}\right)^2$ $\qquad \frac{2}{\rho+3}x + \frac{6}{\rho+3} = y$

$e_2 = (2-0)^2 = 4$

$e_3 = \left(\frac{6}{3-\rho} + \frac{2}{3-\rho}\times(-3) - 0\right)^2 = \left(\frac{12}{3-\rho}\right)^2$ $\qquad \frac{6}{3-\rho} - \frac{2}{3-\rho}x = y$

$\left(\frac{12}{3+\rho}\right)^2 + \left(\frac{12}{3-\rho}\right)^2 + \cancel{4} = \cancel{6}\ 2$

$\frac{\cancel{144}\ 72}{(3+\rho)^2} + \frac{\cancel{144}\ 72}{(3-\rho)^2} = 1$

$72(3-\rho)^2 + 72(3+\rho)^2 = (3+\rho)^2(3-\rho)^2$

$72(9-6\rho+\rho^2) + 72(9+6\rho+\rho^2) = (9-\rho^2)^2$

$\rho^4 - 162\rho^2 - 1215 = 0 \qquad \rho^2 = x$

$x^2 - 162x - 1215 = 0$ $\qquad 1215 = 9\times 9\times 5\times 3$

$x = \frac{162 \pm \sqrt{162^2 + 4\times 1215}}{2}$

$= 81 \pm \frac{\sqrt{4\times 9^4 + 4\times 9^2\times 15}}{2}$ $\qquad \frac{162 \pm \sqrt{162^2 - 4\times 1215}}{2}$

$= 81 \pm \frac{\sqrt{4\times 9^2(9^2+15)}}{2}$

$= 81 \pm 9\sqrt{96}$ $\qquad \rho^2 = 81 + 36\sqrt{6}$

$= 81 \pm 36\sqrt{6}$ $\qquad \rho = \sqrt{81 + 36\sqrt{6}}$

$$\underset{\mathcal{D}_{val} \sim P^K}{\mathrm{Variance}}[E_{val}(h)] = \square \cdot \underset{(x,y)\sim P}{\mathrm{Variance}}[err(h(x), y)]$$

$$\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

$$= \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

$$\mathrm{Var}(E_{val}(h)) = \sum_{i=1}^{K} p_i (x_i - \mu)^2 = \sum_{i=1}^{K} (p_i \cdot x_i^2) - \mu^2$$

$$= \frac{1}{K} \sum_{i=1}^{K} x_i^2 - \mu^2$$

$$\mathrm{Var}(err(h(x), y)) = p_i (x_i - \mu)^2 = (p_i \cdot x_i^2) - \mu^2$$

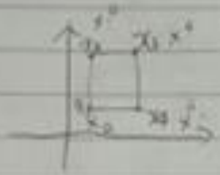
$$= \frac{1}{K} x_i^2 - \mu^2$$

$$\frac{1}{K} \sum_{i=1}^{K} x_i^2 - \mu^2 = \square \cdot \frac{1}{K} x_i^2 - \mu^2$$

$\downarrow$ for normal distribution $\Rightarrow \mu^2 = 0$

$$\cancel{\frac{1}{K}} \sum_{i=1}^{K} x_i^2 = \cancel{\frac{1}{K}} x_i^2$$

14 [d]



$X = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$

$y_n \overset{iid}{\sim} [-1, +1]$

$E_{y_1, y_2, y_3, y_4}\left(\min_{w \in R^{d+1}} E_{in}(w)\right)$

$E(E_{in}(w)) = \dfrac{4}{4 \times 16} = \dfrac{4}{64}$

15.

$P(y = +1) = p$

$E_{out}(g_c) = 1 - p$