Problem 1.

1.

old # dataset : N new # dataset : N

class : 2 # class : K

classifier : 1 # classifier : C_2^K

CPU time = # classifier * time/classifier

$$= C_2^K * a\left(\frac{2N}{K}\right)^3$$

$$= a \, \frac{\cancel{K}\cdot(K-1)}{\cancel{2}} \, \frac{\cancel{8}^{4} N^3}{\cancel{K^3}\, K^2}$$

$$= \frac{4a(K-1)N^3}{K^2}$$

Problem 2.

2.

$$\Phi_Q(x) = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^Q \\ 1 & & & & \vdots \\ \vdots & & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^Q \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_Q \end{bmatrix} = V'$$

consider $Q = N-1$, $V' = V$

$\because$ all $\{x_n\}$ are different

$\therefore$ V 可逆, $V\tilde{w} = y$

$$E_{in}([V\tilde{w} - y]^2)$$

$$= E_{in}([y-y]^2) = 0$$

Problem 3.

3. assume all X_n are different in training set,

$$\Phi(X_n) = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{N \times N}$$

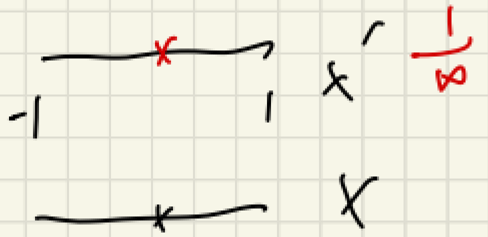
$$E_{in} = E\left[\left(\tilde{w} \cdot \Phi(x) - y\right)^2\right]$$

$$\because \tilde{w} = (\Phi(x)^T \Phi(x))^{-1} \Phi(x) \cdot y = y$$

$$E_{in} = E\left[(y-y)^2\right] = 0 \ \#$$

Assume x', y' 是 testing set 的 data

x 與 x' 相同机率趨近於0



$$\Rightarrow \tilde{w} \cdot \Phi(x') = 0$$

$$E\left[\left(\tilde{w} \cdot \Phi(x') - y'\right)^2\right]$$

$$= E[y']^2$$

$$= E[y']^2 + var(y')$$

$$= E[x+\varepsilon]^2 + var(x+\varepsilon)$$

$$= 0 + \frac{1}{3} + 1 = \frac{4}{3}$$

Problem 4.

4.

$$X_h^T X_h$$

$$= \begin{bmatrix} | & & | & | & & | \\ x_1 & \cdots & x_N & \tilde{x}_1 & \cdots & \tilde{x}_N \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_N & - \\ - & \tilde{x}_1 & - \\ & \vdots & \\ - & \tilde{x}_N & - \end{bmatrix}$$

$$= \sum_{n=1}^{N} \left(X_n X_n^T + \tilde{X}_n \tilde{X}_n^T \right)$$

$$= \sum_{n=1}^{N} x_n x_n^T + \sum_{n=1}^{N} (x_n + \varepsilon_n)(x_n + \varepsilon_n)^T$$

$$= 2 \sum_{n=1}^{N} x_n x_n^T + \sum_{n=1}^{N} \varepsilon_n \varepsilon_n^T + \sum_{n=1}^{N} \varepsilon_n x_n^T$$

$$+ \sum_{n=1}^{N} x_n \varepsilon_n^T$$

令 $X = \begin{bmatrix} - x_1 - \\ \vdots \\ - x_n - \end{bmatrix}, \varepsilon = \begin{bmatrix} - \varepsilon_1 - \\ \vdots \\ - \varepsilon_n - \end{bmatrix}$

$$X_n^T X = 2XX^T + \varepsilon\varepsilon^T + \varepsilon X^T + X\varepsilon^T$$

$$E(X_n^T X)$$

$$= 2E(XX^T) + E(\varepsilon\varepsilon^T) + \cancel{E(\varepsilon X^T)} + \cancel{E(X\varepsilon^T)},$$

$\because$ ε, X^T 互相獨立 且 $E(\varepsilon) = 0$

$$= 2XX^T + E(\varepsilon\varepsilon^T)$$

$$E(\varepsilon^T\varepsilon)$$

$$= E\left(\begin{bmatrix} | & & | \\ \varepsilon_1 & \cdots & \varepsilon_n \\ | & & | \end{bmatrix} \begin{bmatrix} - \varepsilon_1 - \\ \vdots \\ - \varepsilon_n - \end{bmatrix}\right)$$

$$= E\left(\begin{bmatrix} - \varepsilon_1' - \\ \vdots \\ - \varepsilon_{d+1}' - \end{bmatrix} \begin{bmatrix} | & & | \\ \varepsilon_1' & \cdots & \varepsilon_{d+1}' \\ | & & | \end{bmatrix}\right)$$

$$= E(0_{ij}), \quad \text{if } 1 \le i,j \le N$$

$$\because E(\varepsilon_i^{T\prime}\varepsilon_j^{\prime}) = \begin{cases} E(\varepsilon_i^{T\prime})E(\varepsilon_j^{\prime})=0, & \text{if } i\neq j \\ E(\varepsilon_i^{T\prime}\ \varepsilon_i^{\prime}), & \text{if } i=j \end{cases}$$

$$\therefore E(\varepsilon^T\varepsilon) = E(\varepsilon_i^{T\prime}\varepsilon_i^{\prime}) = N E(\varepsilon_{ik}^2)$$

$$= N\left[\cancel{E(\varepsilon_{ik})^2} + var(\varepsilon_{ik})\right] = \frac{N\delta^2}{3}$$

$\because E(\varepsilon)=0$ 且 $var(\varepsilon)=\frac{(a-b)^2}{12}$

$$E(X_h^T X) = 2XX^T + \begin{bmatrix} \frac{N\delta^2}{3} & & 0 \\ & \frac{N\delta^2}{3} & \\ 0 & & \ddots & \\ & & & \frac{N\delta^2}{3} \end{bmatrix}$$

#

Problem 5.

5.
$$\nabla E_{aug}(w) = \nabla E_{in}(w) + \frac{2\lambda}{N} w$$

$$\begin{aligned} W_{t+1} &= W_t - \eta \nabla E_{aug}(w) \\ &= W_t - \eta \left(\nabla E_{in}(w) + \frac{2\lambda}{N} W_t \right) \\ &= W_t - \eta \nabla E_{in}(w_t) - \frac{2\lambda}{N} \eta W_t \\ &= \alpha W_t - \alpha \beta \nabla E_{in}(W_t) \end{aligned}$$

$$\begin{cases} \alpha\beta = \eta & -① \\ 1 - \frac{2\lambda}{N}\eta = \alpha & -② \end{cases}$$

② $\alpha = \frac{N - 2\lambda\eta}{N} < 1 \Rightarrow$ decay

由 ① $\beta = \frac{\eta}{\alpha} = \frac{N\eta}{N - 2\lambda\eta}$ #

Problem 6.

6.

$$\min \frac{1}{N}\sum_{n=1}^{N}(wx_n-y_n)^2+\frac{\lambda}{N}w^2$$

對 w 偏微分

$$\frac{2}{N}\sum_{n=1}^{N}(wx_n^2-x_ny_n)+\frac{\lambda}{N}w=0$$

同乘 $\frac{N}{2}$

$$\sum_{n=1}^{N}(wx_n^2-x_ny_n)+\lambda w=0$$

$$w\left(\lambda+\sum_{n=1}^{N}x_n^2\right)=\sum_{n=1}^{N}x_ny_n$$

$$w=\frac{\sum_{n=1}^{N}x_ny_n}{\lambda+\sum_{n=1}^{N}x_n^2}=\frac{x^Ty}{\lambda+x^Tx}$$

$$\left(\frac{x^T y}{\lambda + x^T x} \right)^2 = C$$

$$\frac{(x^T y)^2}{(\lambda + x^T x)^2} = C$$

$$\frac{(x^T y)^2}{C} = (\lambda + x^T x)^2$$

$$\frac{\| x^T y \|}{\sqrt{C}} = \lambda + x^T x$$

$$\frac{\| x^T y \|}{\sqrt{C}} - x^T x = \lambda$$

$$\alpha = \| \sum_{n=1}^{N} x_n y_n \| \quad , \quad \beta = - \sum_{n=1}^{N} x_n^2$$

Problem 7.

7.

$$\tilde{w}^T = \left(\Phi(x)^T \Phi(x)\right)^{-1} \Phi(x)^T y$$

$$= \left(x^T v^T v x\right)^{-1} v x y$$

$$= \left(v^2 x^T x\right)^{-1} v x y$$

$$= \begin{bmatrix} \frac{1}{v_1^2} & & & \\ & \frac{1}{v_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{v_N^2} \end{bmatrix} v (x^T x)^{-1} x y$$

$$= \begin{bmatrix} \frac{1}{v_1} & & & \\ & \frac{1}{v_2} & & \\ & & \ddots & \\ & & & \frac{1}{v_N} \end{bmatrix} w$$

$$= v^{-1} w$$

$$\min \frac{1}{N} \sum_{n=1}^{N} \left[\tilde{w}^T \Phi(x_n) - y_n \right]^2 + \frac{\lambda}{N} \| \tilde{w} \|$$

$$= \min \frac{1}{N} \sum_{n=1}^{N} \left[\tilde{w}^T V x_n - y_n \right]^2 + \frac{\lambda}{N} \| \tilde{w} \|$$

$$= \min \frac{1}{N} \sum_{n=1}^{N} \left[(V^{-1} w)^T V x_n - y_n \right]^2 + \frac{\lambda}{N} \| \tilde{w} \|$$

$$= \min \frac{1}{N} \sum_{n=1}^{N} \left(w^T x_n - y_n \right)^2 + \frac{\lambda}{N} \left(\| V^{-1} w \| \right)$$

$$\begin{cases} \Omega(w) = \| V^{-1} w \| \\ \tilde{w} = V^{-1} w \end{cases}$$

#

Problem 8.

8.

$$E_{loocv}(g) = \frac{1}{N}\sum_{n=1}^{N} err(g_n(x_n), y_n)$$

抽到 y_n 有兩种可能，$y_n=1$ or $y_n=-1$

而抽到 $y_n=1$ 時，postive data 變少，

$g_n(x_n)$ 永遠輸出 1；反之，永遠輸出 -1

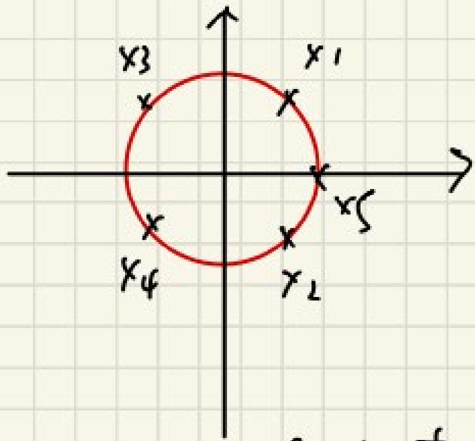
$$g_n(x_n) = \begin{cases} 1, & \text{if } y_n = 1 \\ -1, & \text{if } y_n = -1 \end{cases}$$

∴ 永遠會預測對

$\Rightarrow E_{loocv}(g) = 0$ #

Problem 9.

9.



紅線代表不可完全切開

x_3	x_1	x_5	x_2	x_4
o	o	o	o	o
o	o	o	o	x
o	o	o	x	o
o	o	o	x	x
o	o	x	o	o
~~o~~	~~o~~	~~x~~	~~o~~	~~x~~
o	o	x	x	o

err: $\frac{1}{5}$

x_3	x_1	x_5	x_2	x_4
x	x	x	x	x
x	x	x	x	o
x	x	x	o	x
x	x	x	o	o
x	x	o	x	x
~~x~~	~~x~~	~~o~~	~~x~~	~~o~~
x	x	o	o	x

err: $\frac{1}{5}$

o o   x x X

o X   o o o

~~o X   o o X~~ err: $\frac{1}{5}$

~~o X   o X o~~ err: $\frac{1}{5}$

~~o X   o X X~~ err: $\frac{1}{5}$

o X   X o o

~~o X   X o X~~ err: $\frac{1}{5}$

o X   X X o

o X   X X X

X X o o o

X o X X X

~~X o X X o~~ err: 1

~~X o X o X~~ err: 1

~~X o X o o~~ err: 1

X o o X X

~~X o o X o~~ err: 1

X o o o X

X o o o o

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}[[wx_n \neq y_n]]$$

有 10 個組合 $E_{in}(w) = \frac{1}{5} * 1 = \frac{1}{5}$

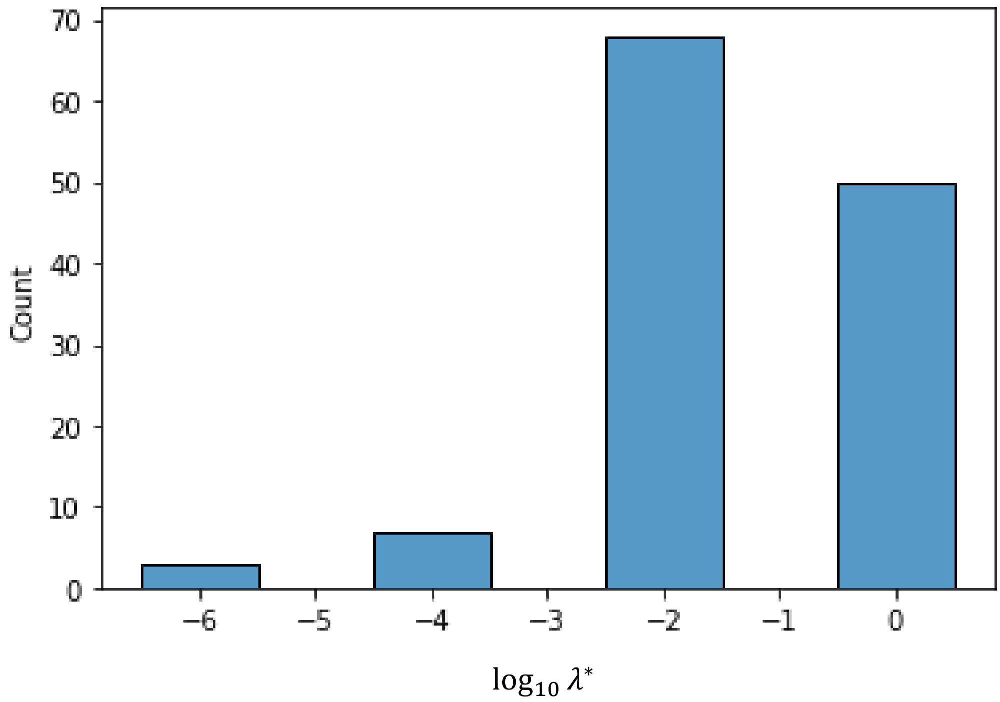
有 22 個組合 $E_{in}(w) = 0$

$$E[E_{in}(w)] = \frac{10}{32} * \frac{1}{5} = \frac{1}{16}$$ ✗

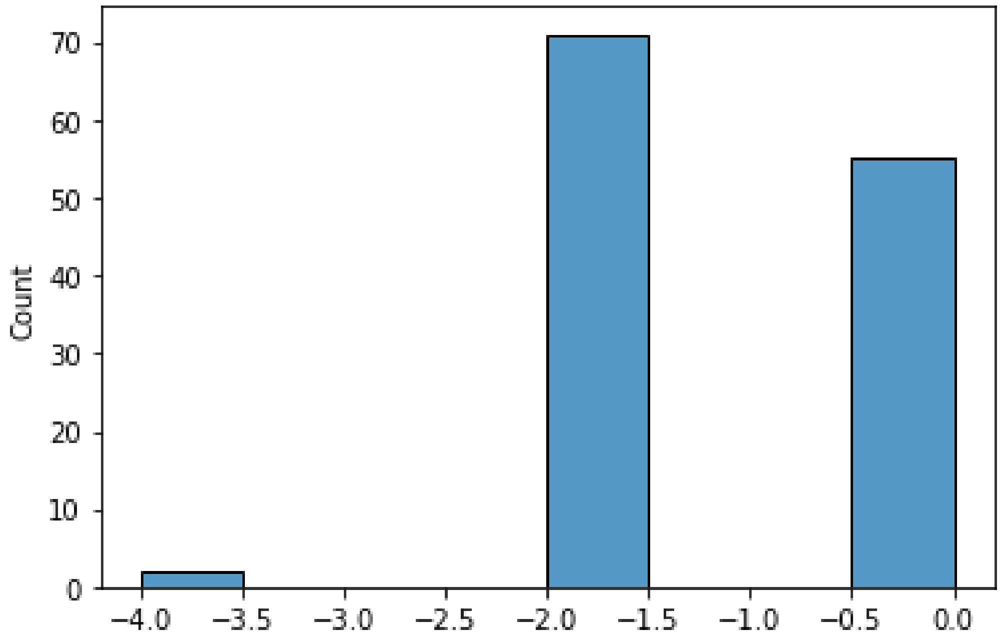
Problem 10.

The best $\log_{10} \lambda^*$ is -6

Problem 11.



Problem 12.



$\log_{10}\lambda^*$

Describe your findings : 我們發現 $\log_{10}\lambda^* = -6$ 不存在，整個分布更向右傾，可推測出當 $\log_{10}\lambda^*$ 落在 -2 與 0 所給予的限制最適當。