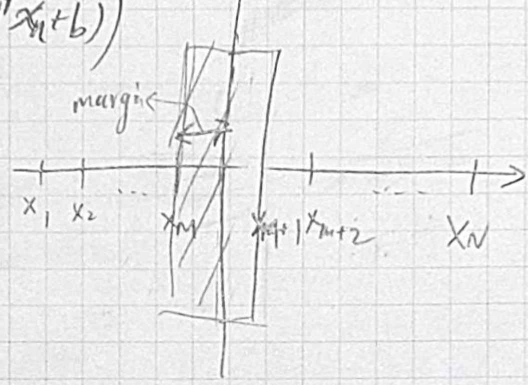
html hw5 B10502166 謝銘倫 电机三 P1

1. $\{(x_n, y_n)\}_{n=1}^{N}$, $x_n \in R$, $\underbrace{x_1 \le x_2 \le \cdots \le x_M}_{y_n=-1} < \underbrace{x_{M+1} \le x_{M+2} \cdots \le x_N}_{y_n=1}$

$$\max_{b,w}\left\{\text{margin}(b,w)\right\} = \frac{1}{\|w\|}\left(y_n(w^T x_n + b)\right)$$



the hyperplane seperate the 1-D data between $[x_M, x_{M+1}]$

$\Rightarrow$ To get biggest margin,

$$\max\{\text{margin}\} = \frac{x_{M+1} - x_M}{2}$$

#

2. $\min_{w,b} \frac{1}{2} w^T w$, subject to $\begin{cases} w^T x_n + b \ge 1 \text{ for } y_n = 1 & \text{—— ①} \\ -(w^T x_n + b) \ge e \text{ for } y_n = -1 & \text{—— ②} \end{cases}$

By ①②,

$$w^T x_n \ge \frac{1-e}{2}$$
$$b \ge \frac{1+e}{2}$$
$$\Rightarrow y_n(w^T x_n + b) \ge \frac{1+e}{2} + y_n\left(\frac{1-e}{2}\right)$$

We build Lagrange function

$$\mathcal{L}(b, w, \alpha) = \frac{1}{2} w^T w + \sum_{n=1}^{N} \alpha_n \left(\frac{1+e}{2} + y_n\left(\frac{1-e}{2}\right) - y_n(w^T x_n + b) \right)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{n=1}^{N} -\alpha_n y_n = 0$$

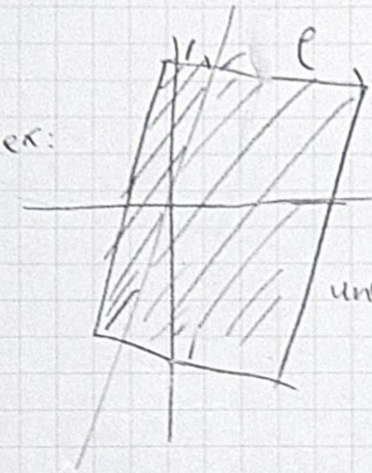
$$\frac{\partial L}{\partial w_i} = 0 \Rightarrow w_i - \sum_n \alpha_n y_n x_{n,i}$$

$$\Rightarrow \max_{\alpha_n \ge 0} \min_{b,w} \left(\mathcal{L}(b, w, \alpha_n) \right)$$

$$= \max_{\substack{\alpha_n \ge 0 \\ \sum \alpha_n y_n = 0}} \left\{ \frac{1}{2} w^T w + \sum \alpha_n \left(\frac{1+e}{2} + y_n \frac{1-e}{2} \right) \right.$$

$$= \max_{\substack{\alpha_n \geq 0 \\ \sum \alpha_n y_n = 0 \\ w = \sum \alpha_n y_n x_n}} \left(\frac{-1}{2} \left\| \sum \alpha_n y_n x_n \right\|^2 + \sum_n \alpha_n \left(\frac{1+\rho}{2} \right) \right)$$

ex:



3.

From even case to uneven case:

even ($\rho^* = 1$) $\Rightarrow$ $\text{svm} = \max_{\substack{\alpha_n \geq 0,\ \sum \alpha_n y_n = 0 \\ w = \sum \alpha_n y_n x_n}} \frac{-1}{2} \left\| \sum_n \alpha_n y_n x_n \right\|^2 + \sum_n \alpha_n$ —— ①

uneven ($\rho^* = \rho'$) $\Rightarrow$ $\text{svm} = \max_{\substack{\beta_n \geq 0,\ \sum \beta_n y_n = 0 \\ w = \sum \beta_n y_n x_n}} \frac{-1}{2} \left\| \sum_n \beta_n y_n x_n \right\|^2 + \sum_n \beta_n \left(\frac{1+\rho}{2} \right)$ —— ②

By comparing, we find we can rewrite ① $\Rightarrow$

$$\frac{-1}{2} \left(\frac{1+\rho}{2} \right)^2 \left\| \sum \alpha_n y_n x_n \right\|^2 + \sum \alpha_n \left(\frac{1+\rho}{2} \right)^2$$

$\Rightarrow$ We find the relation between α_n, β_n

$$\beta_n = \left(\frac{1+\rho}{2} \right) \alpha_n$$

By $w = \sum \alpha_n y_n x_n$, We have $w_1^* = \sum \alpha_n y_n x_n$

$w_\rho^* = \sum \beta_n y_n x_n$

$\therefore\ w_\rho^* = \left(\frac{1+\rho}{2} \right) w_1^*$ #

$$b_1^* = y_n - w_1^{*T} x_n \quad \#$$

By result in Problem 2,

$$L(b, w, \alpha) = \frac{1}{2} w^T w + \sum_{n=1}^{1} \alpha_n \left(\frac{1+\rho}{2} + y_n \left(\frac{1-\rho}{2} \right) - y_n \left(w^T x_n + b \right) \right)$$

By primal-inner-condition, we have

$$\frac{1+\rho}{2} + y_n \left(\frac{1-\rho}{2} \right) - y_n \left(w^T x_n + b \right) = 0$$

$$\Rightarrow y_n^2 \left(w^{*T} x_n + b^* \right) = y_n \left(\frac{1+\rho}{2} \right) + y_n^2 \left(\frac{1-\rho}{2} \right)$$

$y_n^2 = 1$

$$\Rightarrow \left(w^{*T} x_n + b^* \right) = y_n \left(\frac{1+\rho}{2} \right) + \left(\frac{1-\rho}{2} \right)$$

$$\therefore b^* = y_n \left(\frac{1+\rho}{2} \right) + \left(\frac{1-\rho}{2} \right) - w^{*T} x_n$$

$$\Rightarrow b_\rho^* = \left(\frac{1+\rho}{2} \right) y_n + \left(\frac{1-\rho}{2} \right) - \underbrace{w_\rho^* x_n}_{\hookrightarrow (= w_1^{*T} x_n)}$$

$$= \left(\frac{1+\rho}{2} \right) y_n + \left(\frac{1-\rho}{2} \right) - \left(\frac{1+\rho}{2} \right) w_1^{*T} x_n$$

$$= \left(\frac{1+\rho}{2} \right) \left(y_n - w_1^{*T} x_n \right) + \left(\frac{1-\rho}{2} \right)$$

$$= \left(\frac{1+\rho}{2} \right) b_n^* + \left(\frac{1-\rho}{2} \right)$$

$\therefore$ By substituting $\rho = 1126$, we get

$$b_{1126}^* = \frac{1127}{2} b_1^* - \frac{1125}{2}$$

$$W_{1126}^* = \frac{1127}{2} W_1^*$$

4. From Problem 3, we derive $\beta_n = \frac{1+e}{2}\alpha_n$

$\Rightarrow$ we can find $\alpha_p^* = \frac{1+e}{2}\alpha_1^*$ #

5. $K_1(x, x') = \phi_1(x)^T \phi_1(x')$

$K_2(x, x') = \phi_2(x)^T \phi_2(x')$

Consider kernel function $K(x, x') = K_1(x, x') \cdot K_2(x, x')$

$$K(x, x') = \phi_1(x)^T \phi_1(x') \cdot \phi_2(x)^T \phi_2(x')$$

$$= \phi(x)^T \phi(x)$$

$$K(x, x') = \sum_{i=1}^{n} \phi_1(x_i)\, \phi_1(x_i') \cdot \sum_{j=1}^{n} \phi_2(x_j)\, \phi_2(x_j')$$

$$= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(\phi_1(x_i)\, \phi_2(x_j)\right)\left(\phi_1(x_i')\, \phi_2(x_j')\right)$$

$$\overset{n_1 \times n_2 = N}{=} \sum_{n=1}^{N} \phi(x)_n\, \phi(x')_n$$

$$\therefore \phi(x) = \begin{bmatrix} \phi_1(x_1)\phi_2(x_1) \\ \phi_1(x_1)\phi_2(x_2) \\ \vdots \\ \phi_1(x_1)\phi_2(x_n) \\ \phi_1(x_2)\phi_2(x_1) \\ \phi_1(x_2)\phi_2(x_2) \\ \vdots \\ \phi_1(x_n)\phi_2(x_1) \\ \phi_1(x_n)\phi_2(x_2) \\ \vdots \\ \phi_1(x_n)\phi_2(x_n) \end{bmatrix}$$

#

6. x and x', $\|\phi(x)-\phi(x')\|$, max? min?

$$K_2(x,x')=(1+x^Tx')^2=1+2(x^Tx')+(x^Tx')^2$$

$$\text{distance}=\|\phi(x)-\phi(x')\|=\sqrt{\|\phi(x)-\phi(x')\|^2}$$

$$=\sqrt{\phi^2(x)-2\phi(x)\phi(x')+\phi^2(x')}$$

$$=\sqrt{K(x,x)+K(x',x')-2K(x,x')}$$

x, x' are unit vectors,

$$\min\,(1+x^Tx')^2=0 \quad (\text{when } x^Tx'=-1)$$

$$\max\,(1+x^Tx')^2=4 \quad (\text{when } x^Tx'=1)$$

For Distance $\|\phi(x)-\phi(x')\|$,

$$\max=\|\phi(x)-\phi(x')\|_{max}=\sqrt{4+4-2\cdot\underset{(x^Tx'=-1)}{0}}=2\sqrt{2}$$ #

$$\min=\|\phi(x)-\phi(x')\|_{min}=\sqrt{4+4-2\cdot\underset{(x^Tx'=1)}{4}}=0$$ #

7. $$\|\tilde{\phi}(x)\|^2=\left\|\left(1,\sqrt{\tfrac{2}{1!}}x,\sqrt{\tfrac{2^2}{2!}}x^2,\cdots\right)\right\|^2$$

$$=\left(1+\frac{2}{1!}x^2+\frac{2^2}{2!}x^4+\cdots\right)$$

$$\left(e^{(-x^2)}\right)^2=e^{-2x^2}=1+\frac{(-2x^2)}{1!}+\frac{(-2x^2)^2}{2!}+\cdots$$

$$\|\tilde{\phi}(x)\|^2\left(e^{-x^2}\right)^2=\left(1+\frac{2x^2}{1!}+\frac{2^2}{2!}x^4+\cdots\right)\left(1-\frac{2x^2}{1!}+\frac{2x^4}{2!}+\cdots\right)$$

$$=1$$

$$\|\tilde{\phi}(x)\|^2 (e^{-x^2})^2 = 1 \Rightarrow (e^{-x^2})^2 = \frac{1}{\|\tilde{\phi}(x)\|^2}$$

$$\Rightarrow \exp(-x^2) = \frac{1}{\|\tilde{\phi}(x)\|}$$

#

8. $\cos(x, x') = \dfrac{x^T x'}{\|x\| \cdot \|x'\|} = \dfrac{x^T}{\|x\|} \cdot \dfrac{x'}{\|x'\|}$

To prove it's valid kernel

Let $\phi(x) = \dfrac{x}{\|x\|} \Rightarrow \cos(x, x') = \phi^T(x)\phi(x')$

symmetric: $\cos(x', x) = \cos(x, x')$ by definition

positive-semi-definite, $\cos(x, x') = \phi^T \phi(x') = z^T z \geq 0$

$\therefore$ $u^T z^T z u = (u^T z^T)(zu) = (zu)^T (zu) \geq 0$

$\Rightarrow$ positive semi-definite

By Mercer's condition,

$\cos(x, x')$ is a valid kernel on $\mathcal{X} = R^d - \{0\}$

#

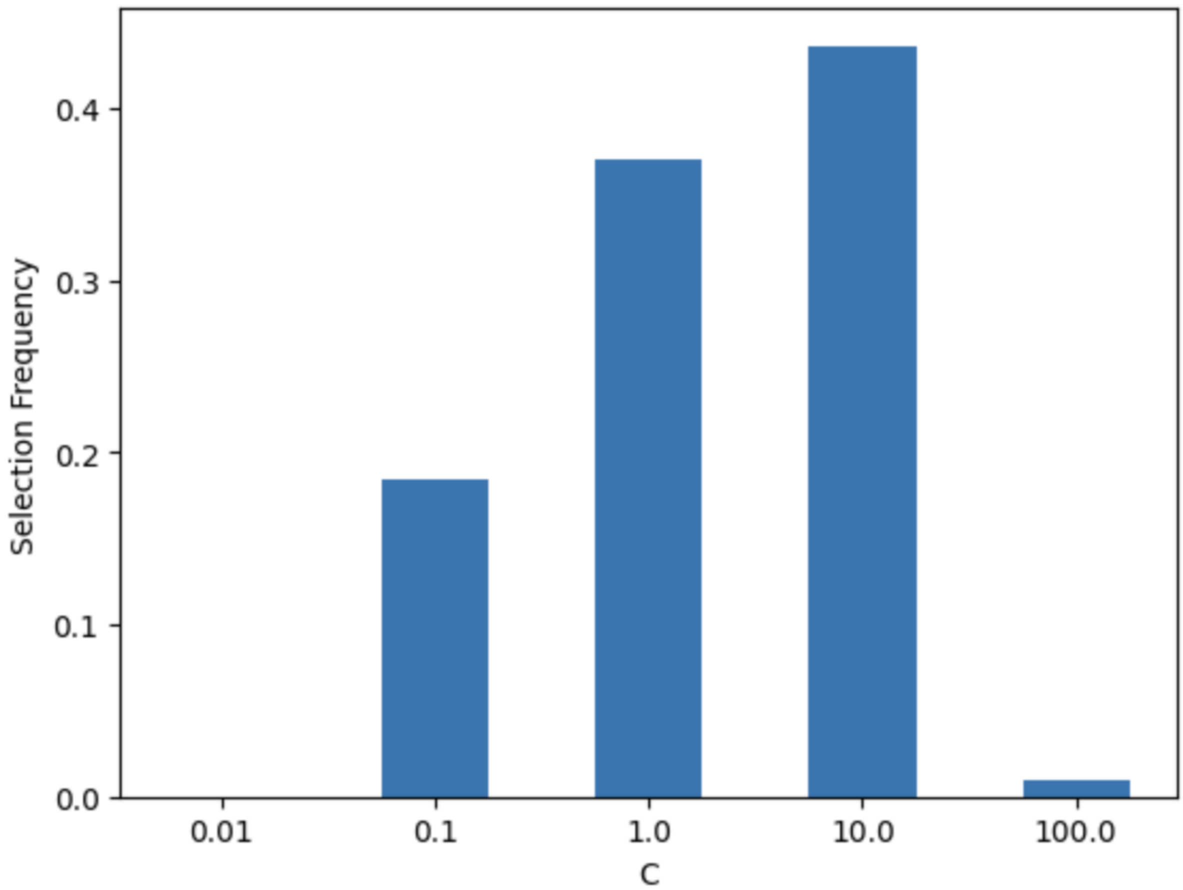
HTML hw5 solution

姓名: 謝銘倫, 系級:電機三, 學號:B10502166

9.
When C = 10.0, Q = 4, we have the smallest number of support vectors =629.

10.
When C = 1.0, we have min E_out: 0.005

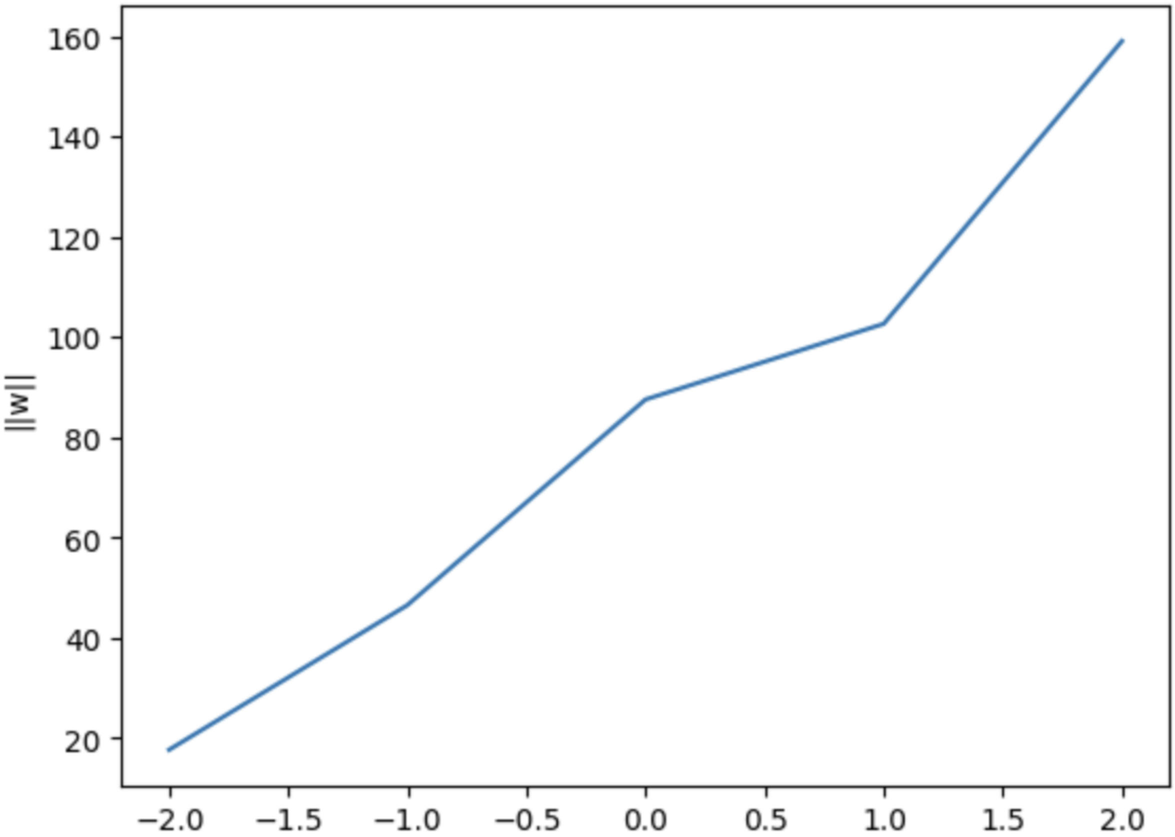
11.



Findings:
After running 1000 steps, we gain the statistic of best C: C_ans = [0. 184. 370. 436. 10.], where the C distribute more in c = [0.1, 1.0, 10.0], and very few in c= [0.01, 100.0]

Compared to problem 10, we find we have a little more smallest E_val in C=10, instead of C =1 in problem 10.

12.



Findings:
the array of ||w||: [17.5913755   46.44587232 87.4856177   102.58834219
159.00656295]
v.s. log(C)

We occur when the value of C increases, the value of ||w|| increases.