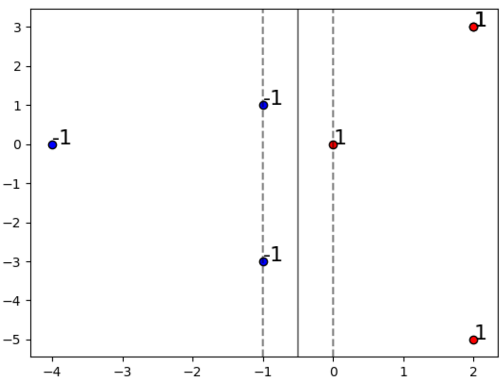
1.

x = [[1,0],[0,1],[0,-1],[-1,0],[0,2],[0,-2],[-2,0]]

y = [-1,-1,-1,1,1,1,1]

z = [[-4,0],[-1,-3],[-1,1],[0,0],[2,-5],[2,3],[2,3]]

plot of z:



It's easy to observe that the optimal separating "hyperplane" in Z space is $x_1 = -0.5$

2.

I used sklearn package. Below is my code:

```
import matplotlib.pyplot as plt
import numpy as np

x = np.array([[1,0],[0,1],[0,-1],[-1,0],[0,2],[0,-2],[-2,0]])
y = np.array([-1,-1,-1,1,1,1,1])

def mykernel(x1,x2):
    tmp = (1 + np.dot(x1,x2.T)) ** 2
    return tmp

from sklearn import svm
clf = svm.SVC(kernel = mykernel, C = 1e10, degree = 2, gamma = 1, coef0 = 1)
clf.fit(x,y)

print('Indices of SV:', clf.support_)
print('alpha_n:', clf.dual_coef_)
```

The optimal α are $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) \approx (0, 0.5964, 0.8106, 0.8887, 0.2056, 0.3127, 0)$

and the support vectors are (0,1), (0,-1), (-1,0), (0,2), (0,-2)

3.

◆ date/ ◆ page/

When α is optimal, $b = \sum_{n=1}^{N} \alpha_n y_n k(x_n, x_s) = -1.666$

$W^T x + b = \sum_{n=1}^{N} \alpha_n y_n k(x_n, x) + b$

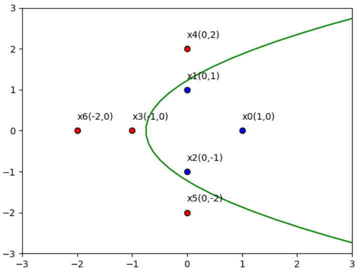
$= \alpha_2 y_2 (1+x_2)^2 + \alpha_3 y_3 (1-x_2)^2 + \alpha_4 y_4 (1-x_1)^2 + \alpha_5 y_5 (1+2x_2)^2 + \alpha_6 y_6 (1-2x_2)^2$

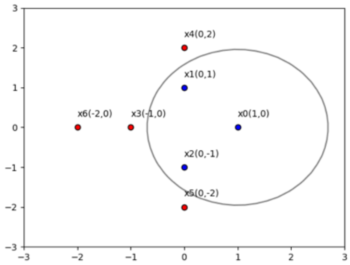
$\doteq -1.666$

$\Rightarrow -\alpha_2 (1+x_2)^2 - \alpha_3 (1-x_2)^2 + \alpha_4 (1-x_1)^2 + \alpha_5 (1+2x_2)^2 + \alpha_6 (1-2x_2)^2 - 1.666 = 0$

where $\alpha_2 = 0.5964$, $\alpha_3 = 0.8106$, $\alpha_4 = 0.8887$, $\alpha_5 = 0.2056$, $\alpha_6 = 0.3127$

4.





They dont have to be the same, because different kernel function (different nonlinear transformation) will lead to different optimal separating nonlinear curve

5.

$$(P_1')\ \min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n$$

$$\text{s.t. } y_n(w^T x_n + b) \geq p_n - \xi_n$$

$$\xi_n \geq 0$$

$$L(b,w,\xi,\alpha,\beta) = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \alpha_n (p_n - \xi_n - y_n(w^T x_n + b)) + \sum_{n=1}^{N} \beta_n (-\xi_n)$$

6.

$$L(b,w,\xi,\alpha,\beta) = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \alpha_n (p_n - \xi_n - y_n(w^T x_n + b)) + \sum_{n=1}^{N} \beta_n (-\xi_n)$$

By KKT conditions:

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

$$\frac{\partial L(\cdots)}{\partial w_i} = 0 = w_i - \sum_{n=1}^{N} \alpha_n y_n x_{n,i}$$

$$\Rightarrow w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

$$\frac{\partial L(\cdots)}{\partial \xi_i} = 0 = C - \alpha_i - \beta_i$$

$$\Rightarrow \alpha_n = C - \beta_n$$

$$L(\cdots) = \frac{1}{2} w^T w + \sum_{n=1}^{N} \alpha_n (p_n - y_n(w^T x_n + b)) + \sum_{n=1}^{N} \overbrace{\cancel{(C - \alpha_n - \beta_n)}}^{0} \xi_n$$

$$= \frac{1}{2} w^T w + \sum_{n=1}^{N} \alpha_n p_n - \underbrace{\sum_{n=1}^{N} \alpha_n y_n (w^T x_n)}_{w^T w} - \cancel{\sum_{n=1}^{N} \alpha_n y_n b}^{\,0}$$

$$= -\frac{1}{2} w^T w + \sum_{n=1}^{N} \alpha_n p_n$$

Thus (P_1') turns into

$$\max_{\alpha_n \geq 0, \beta_n \geq 0} -\frac{1}{2}w^Tw + \sum_{n=1}^{N} \alpha_n p_n \Leftrightarrow \min \frac{1}{2}w^Tw - \sum_{n=1}^{N} \alpha_n p_n$$

$$s.t. \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

$$\beta_n = C - \alpha_n \text{ for } n = 1, 2, \dots, N$$

$$\alpha_n \geq 0, \beta_n \geq 0, \Rightarrow 0 \leq \alpha_n \leq C$$

7.

(P_1')

$$\min_{w,b,\varepsilon} \frac{1}{2}w^Tw + C\sum_{n=1}^{N} \varepsilon_n$$

$$s.t. \ y_n(w^Tx_n + b) \geq \frac{1}{2} - \varepsilon_n$$

$$\varepsilon_n \geq 0$$

$$\Rightarrow s.t. \ y_n(2w^Tx_n + 2b) \geq 1 - 2\varepsilon_n$$

$$2\varepsilon_n \geq 0$$

(P_1)

$$\min_{w,b,\varepsilon} \frac{1}{2}w^Tw + C\sum_{n=1}^{N} \varepsilon_n$$

$$s.t. \ y_n(w^Tx_n + b) \geq 1 - \varepsilon_n$$

w^*, b^* is the optimal solution of (P_1') with $p_n = 0.5$

rewrite (P_1'), let $w' = 2w$, $b' = 2b$, $\varepsilon_n' = 2\varepsilon_n$

$$\min_{w,b,\varepsilon} \frac{1}{2}w'^Tw' + C\sum_{n=1} \varepsilon_n'$$

$$s.t. \ y_n(w'^Tx_n + b') \geq 1 - \varepsilon_n'$$

now, we can know the optimal solution of P_1 is $(2w^*, 2b^*)$

8.

The difference between hard-margin SVM and soft-margin SVM is that in soft-margin SVM there is an upper bound that constraints α. So, if C is greater the optimal α (which means that α_n in both problems will stay in the range), i.e. $C \geq max_{(1 \leq n \leq N)} \alpha_n^*$, than the two problems have the same solution.

9.

◆ date/ ◆ page/

Valid kernel $\Rightarrow$ positive semidefinite matrix $\Rightarrow$ non-negative eigenvalue
$\Rightarrow$ symmetric

Denote $K = K_1(X, X')$, set $K = 0.5I = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$, $eigen(K) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

a. $eigen((1-K)^{-1}) = eigen(\begin{bmatrix} 0.5 & 1 \\ 1 & 0.5 \end{bmatrix}) = \begin{bmatrix} 2.5 \\ -1.5 \end{bmatrix}$, not valid kernel

b. $eigen((1-K)^0) = eigen(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ valid kernel

$K'(X, X') = (1 - K_1(X, X'))^0 = 1$, $\forall X, X'$, we got $K'(X, X') = K'(X', X)$

lemma 1

if $K_1(X, X')$ and $K_2(X, X')$ are valid kernel, then $K(X, X') = K_1(X, X') K_2(X, X')$ is valid.

lemma 2

if $\forall i \in N$, $K_i(x, x')$ is valid kernel, and $\sum_{i=1}^{\infty} K_i(X, X')$ exists, then $\sum_{i=1}^{\infty} K_i(X, X')$ is valid.

proof:

(1) positive semidefinite is closed under multiplication

$K'(X, X')$ is positive semedefinite.

$K'(X, X') = K_1(X, X') K_2(X, X') = K_1(X', X) K_2(X', X) = K'(X', X)$

$K'(X, X')$ is symmetric. $\Rightarrow$ K' is a valid kernel

(2) positive semidefinite is closed under addition

$K'(X, X') = \sum_{i=1}^{\infty} K_i(X, X')$ is positive semidefinite.

$\because K_i(X, X')$ $\forall i \in N$ is valid $\Rightarrow$ symmetric

$\therefore K'(X, X') = \sum_{i=1}^{\infty} K_i(X, X') = \sum_{i=1}^{\infty} K_i(X', X) = K'(X', X)$ symmetric

$\Rightarrow K'(X, X')$ is a valid kernel.

c. $\because 0 < K(X, X') < 1$ $\therefore$ $K'(X, X') = (1 - K(X, X'))^{-1} = \sum_{i=1}^{\infty} K(X, X')^i$

by lemma 1, we know $\forall i \in N$, $K_1(X, X')^i$ is valid

by lemma 2, we know $K'(X, X') = \sum_{i=1}^{\infty} K(X, X')^i$ is valid

d. $\because K'(X, X') = (1 - K(X, X'))^{-2} = (1 - K(X, X')^{-1})(1 - K(X, X')^{-1})$

by (c) we know $1 - K(X, X')$ is valid

by lemma 1, we know $K'(X, X') = (1 - K(X, X'))^2$ is valid.

10.

$$\tilde{K}(x,x') = pK(x,x') \quad , p>0$$

$$b = y_s - \sum_{\text{sv indices } n}^{N} \alpha_n y_n K(x_n, x_s) \text{ on bounded SV } (x_s, y_s)$$

$$g_{svm} = \text{sign}\left(\left(\sum_{\text{sv indice } n}^{N} \alpha_n y_n K(x_n, x)\right) + b\right)$$

$$= \text{sign}\left(\sum_{\text{sv indices } n}^{N} \alpha_n y_n \left(K(x_n, x) - K(x_n, x_s)\right) + y_s\right)$$

$$\tilde{b} = y_s - \sum_{\text{sv indices } n}^{N} \tilde{\alpha}_n y_n \tilde{K}(x_n, x_s) = y_s - \sum_{\text{sv indices } n}^{N} \tilde{\alpha}_n y_n p K(x_n, x_s) \text{ on bounded SV } (x_s, y_s)$$

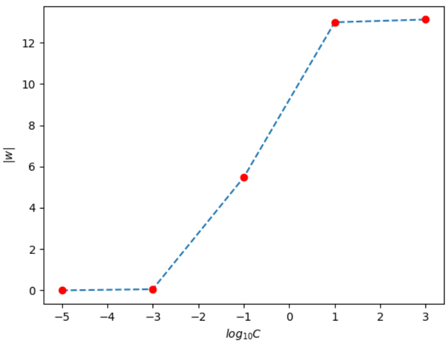
$$\tilde{g}_{svm} = \text{sign}\left(\left(\sum_{\text{sv indices } n}^{N} \tilde{\alpha}_n y_n \tilde{K}(x_n, x)\right) + b\right)$$

$$= \text{sign}\left(\left(\sum_{\text{sv indices } n}^{N} p\tilde{\alpha}_n y_n \left(K(x_n, x) - K(x_n, x_s)\right)\right) + y_s\right)$$

if $g_{svm} = \hat{g}_{svm}$

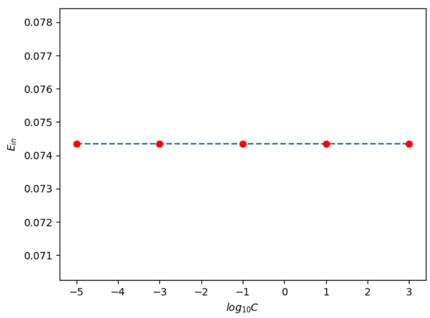
then $\tilde{\alpha}_n = \frac{1}{p}\alpha_n \Rightarrow \tilde{C} = \frac{1}{p}C$

11.



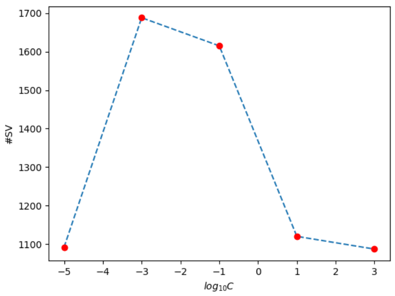
C 越大表示:在邊界周圍能容忍的錯誤較少,同時圖形為了滿足這個條件,也會變複雜,複雜的曲線其|W|也會變大。

12.



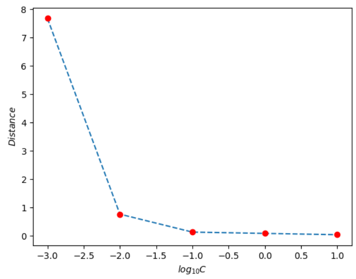
理論上,C 越大表示:在邊界周圍,能容忍的錯誤越少。但這筆 Ein 沒什麼變,而且原本 8 所佔的比例剛好就是 0.074,所以 C 的影響很小。

13.



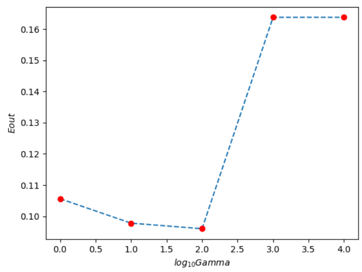
C 較小的時候,因為圖形會偏簡單,所以 Support Vector 會變少;而 C 太大的時候,他會盡量避免分類錯誤,讓 unbounded Support Vector 不要太多,所以 Support Vector 也會變少。

14.



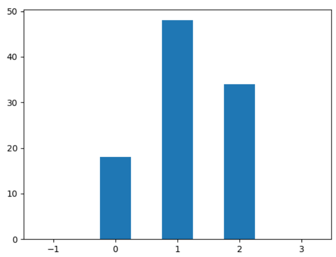
當 log10C 越大時,Z space 中 free support vector 到 optimal separating hyperplane 的距離越小。

15.



當 log10 γ 從 0 增加到 1 時,Eout 會下降到最小值,而當 log10 γ 從 1 遞增到 4 時,Eout 則會遞增。我想是因為 γ 太大會導致 overfitting,所以最後 Eout 會變大。

16.



當 log10 γ 為 1 時,被選中的次數最多，而當 log10 γ 為-1 或 3 時，則皆沒被選中。

17.

17.

Let z_c be the constant feature component, then the optimal weight value w_c is $w_c = \sum_{n=1}^{N} \alpha_n y_n z_{n,c}$

Because $z_{n,c}$ is constand, $w_c = z_{n,c} \sum_{n=1}^{N} \alpha_n y_n$

recall that when we reach optimal solution of soft-margin SVM

$\sum_{n=1}^{N} \alpha_n y_n = 0$, thus $w_c = 0$ #

18.

let. $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $Q = \begin{bmatrix} y_1 y_1 z_1^T z_1 & \cdots & y_1 y_n z_1^T z_n \\ \vdots & \ddots & \\ y_n y_1 z_n^T z_1 & & y_n y_n z_n^T z_n \end{bmatrix}$

then hard-margin dual SVM.

$= \min \frac{1}{2}\alpha^T Q \alpha - 1_N^T \alpha$

subject to $\alpha_n \geq 0$, $y^T \alpha = 0$

also $1_N^T \alpha = \sum_{n=1}^{N} \alpha_n$

$\min \max L(\alpha, \lambda, \mu)$

$= \min \max \frac{1}{2}\alpha^T Q \alpha - 1_N^T \alpha + \sum_{n=1}^{N} \lambda_n (-\alpha_n) + \mu y^T \alpha$

$= \min \max \frac{1}{2}\alpha^T Q \alpha - 1_N^T \alpha - \lambda^T \alpha + \mu y^T \alpha$

$= \min \max \frac{1}{2}\alpha^T Q \alpha - (\lambda + 1_N + \mu y)^T \alpha$

Suppose the problem is feasible

Because the problem is convex, and the constraints are all linear.

we can use strong duality

$\min \max \frac{1}{2}\alpha^T Q \alpha - (\lambda + 1_N - \mu y)^T \alpha$

$= \max \min \frac{1}{2}\alpha^T Q \alpha - (\lambda + 1_N - \mu y)^T \alpha$

By KKT condition

$Q\alpha - (\lambda + 1_N - \mu y) = 0$

$\Rightarrow \alpha = Q^{-1}(\lambda + 1_N - \mu y)$

We can rewrite the problem to.

$\max \frac{1}{2}(\lambda + 1_N - \mu y)^T (Q^{-1})^T (\lambda + 1_N - \mu y) - (\lambda + 1_N - \mu y)^T (Q^{-1})(\lambda + 1_N - \mu y)$

Because Q is symmetric.

$\Rightarrow \max \frac{-1}{2}(\lambda + 1_N - \mu y)^T Q^{-1} (\lambda + 1_N - \mu y)$

Also equal to $\min \frac{1}{2}(\lambda + 1_N - \mu y)^T Q^{-1} (\lambda + 1_N - \mu y)$