Machine Learning HW1

資工系 b05502087 王竑睿

1.

先將所有的 x 座標透過ϕ_1和ϕ_2來轉換到 Z 空間

$Z_1 = \left(\phi_1(1,0), \phi_2(1,0)\right) = (-2,-2)$

$Z_2 = \left(\phi_1(0,1), \phi_2(0,1)\right) = (4,-5)$

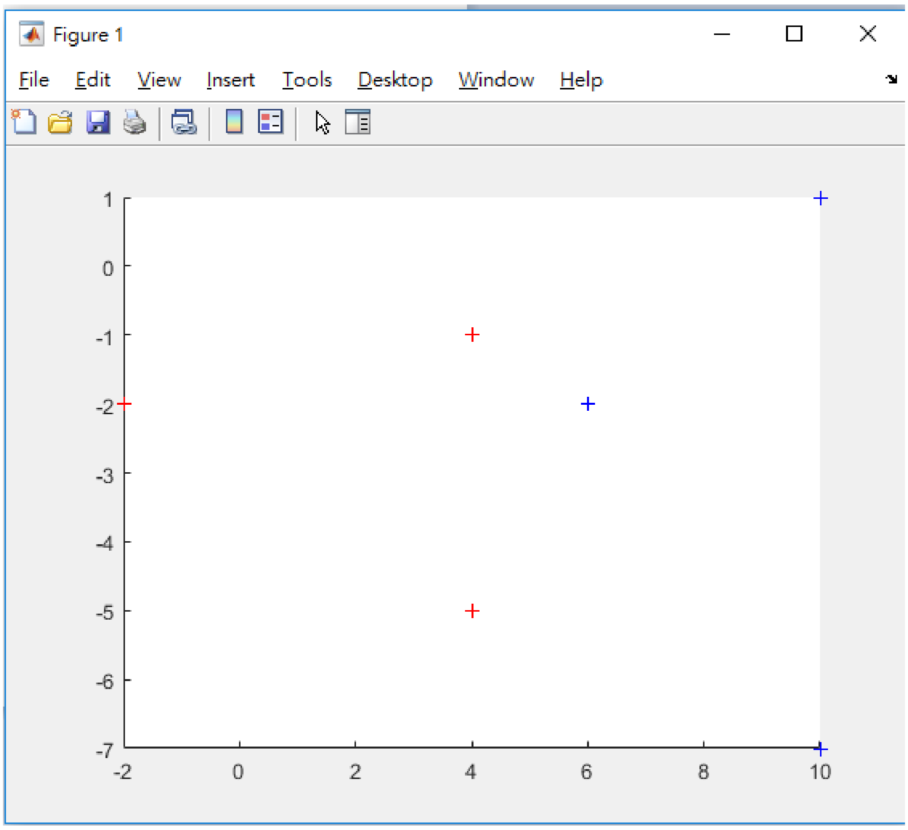
$Z_3 = \left(\phi_1(0,-1), \phi_2(0,-1)\right) = (4,-1)$

$Z_4 = \left(\phi_1(-1,0), \phi_2(-1,0)\right) = (6,-2)$

$Z_5 = \left(\phi_1(0,2), \phi_2(0,2)\right) = (10,-7)$

$Z_6 = \left(\phi_1(0,-2), \phi_2(0,-2)\right) = (10,1)$

$Z_7 = \left(\phi_1(-2,0), \phi_2(-2,0)\right) = (10,1)$



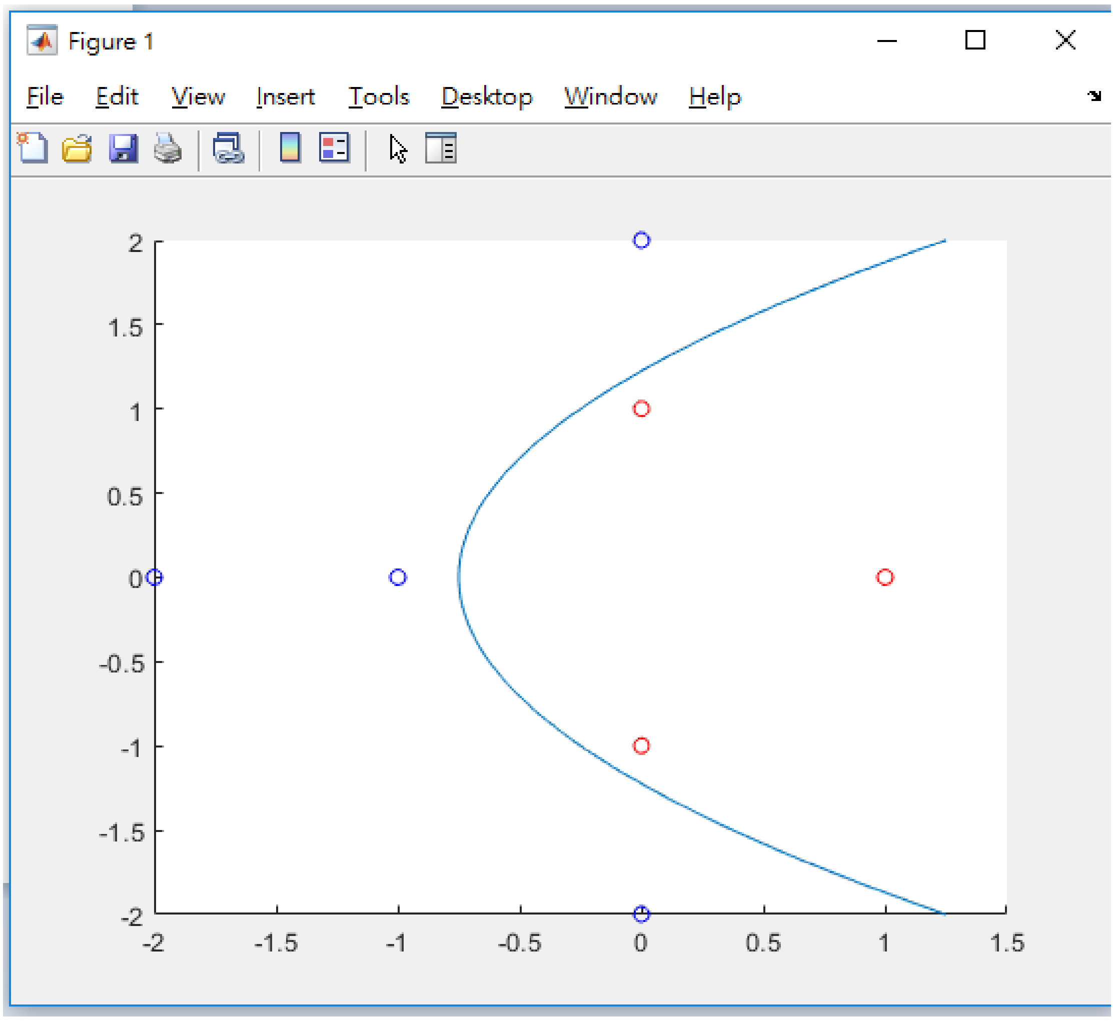
從圖中可以看出 optimal separating hyperplane 位在 $Z_1 = 5$ 的位置

$Z_1 = \phi_1 = 5$

⇨ $2x_2^2 - 4x_1 + 2 = 5$

⇨ $x_1 = \frac{2x_2^2 - 3}{4}$

得到下圖



所以 optimal separating hyperplane 為

$$x_1 = \frac{2x_2^2 - 3}{4}$$

2.

本題使用 quadratic programming 來解

$q_{n,m} = y_n y_m K(x_n, x_m)$

$$= \begin{matrix} 4 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 4 & 0 & -1 & -9 & -1 & -1 \\ 1 & 0 & 4 & -1 & -1 & -9 & -1 \\ 0 & -1 & -1 & 4 & 1 & 1 & 9 \\ -1 & -9 & -1 & 1 & 25 & 9 & 1 \\ -1 & -1 & -9 & 1 & 9 & 25 & 1 \\ -1 & -1 & -1 & 9 & 1 & 1 & 25 \end{matrix}$$

$\mathrm{p} = -1_{\mathrm{N}}$

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

目標函數一次式的係數

A=

1	1	1	-1	-1	-1	-1
-1	-1	-1	1	1	1	1
-1	0	0	0	0	0	0
0	-1	0	0	0	0	0
0	0	-1	0	0	0	0
0	0	0	-1	0	0	0
0	0	0	0	-1	0	0
0	0	0	0	0	-1	0
0	0	0	0	0	0	-1

A 的第一列表示$\sum \mathrm{y}_{\mathrm{n}} \alpha_n \geq 0$

A 的第二列表示$\sum \mathrm{y}_{\mathrm{n}} \alpha_n \leq 0$

A 的第三列到第九列表示 $\alpha_{\mathrm{n}} \geq 0 \;\; n = 1,2, \ldots ,7$

c=[0,0,0,0,0,0,0,0,0]

c 是條件式的常數，共有九個條件，故為九維向量

由 matlab `alpha=quadprog(Q,p,A,c)`

得到

`alpha=[0.0000,0.7037,0.7037,0.8889,0.2593,0.2593,0.0000]`

其中 alpha 的第一個元素和第七個元素為 0

所以X_1和X_7是 support vectors

3.

由上題所求出的 alpha

知 $\mathrm{b} = \mathrm{y}_{\mathrm{s}} - w^T z_s = y_s - (\sum \alpha_n y_n z_n) z_s = y_s - \sum \alpha_n y_n K(x_n, x_s)$

取$\mathrm{x}_{\mathrm{s}} = (0,1) \;\; \mathrm{y}_{\mathrm{s}} = -1$

可以算出 b=-1.6671

Nonlinear curve 為

$\mathrm{w}^{\mathrm{T}} \phi(x) + b = 0$

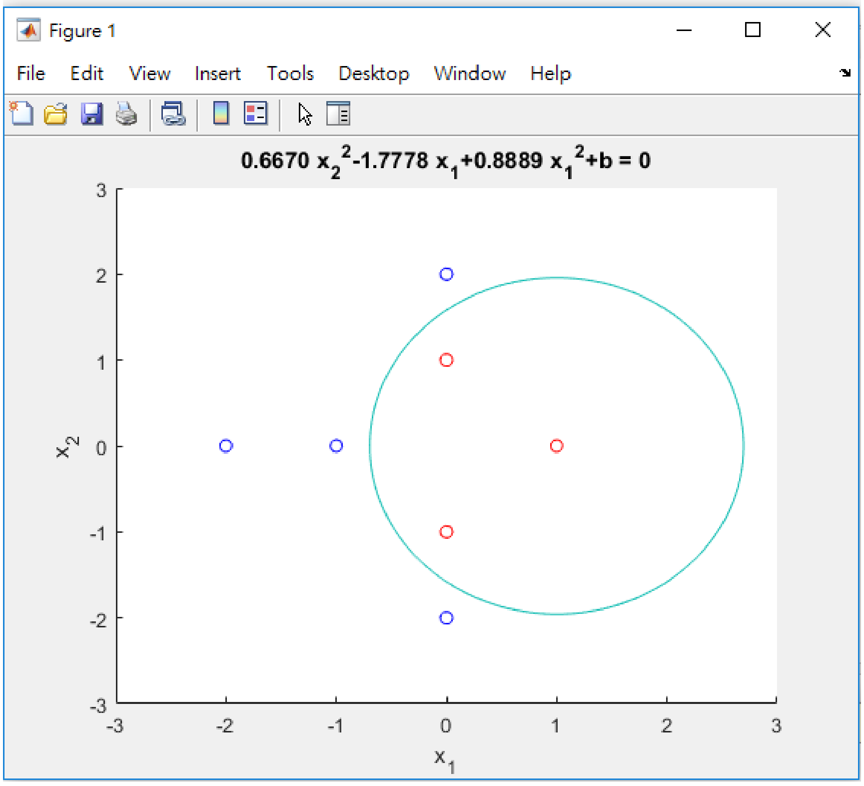
⇨ $\sum \alpha_{\mathrm{n}} y_n K(x_n, x) + b = 0$

Kernel function 和 α_{n} 和 y_{n} 和 b 都是已知

$\mathrm{x} = (\mathrm{x}_1, x_2)$代入即可以解得曲線

$$0.6670\mathrm{x}_2^2 + 0.8889 x_1^2 - 1.7778 x_1 - 1.6671 = 0$$

曲線如圖



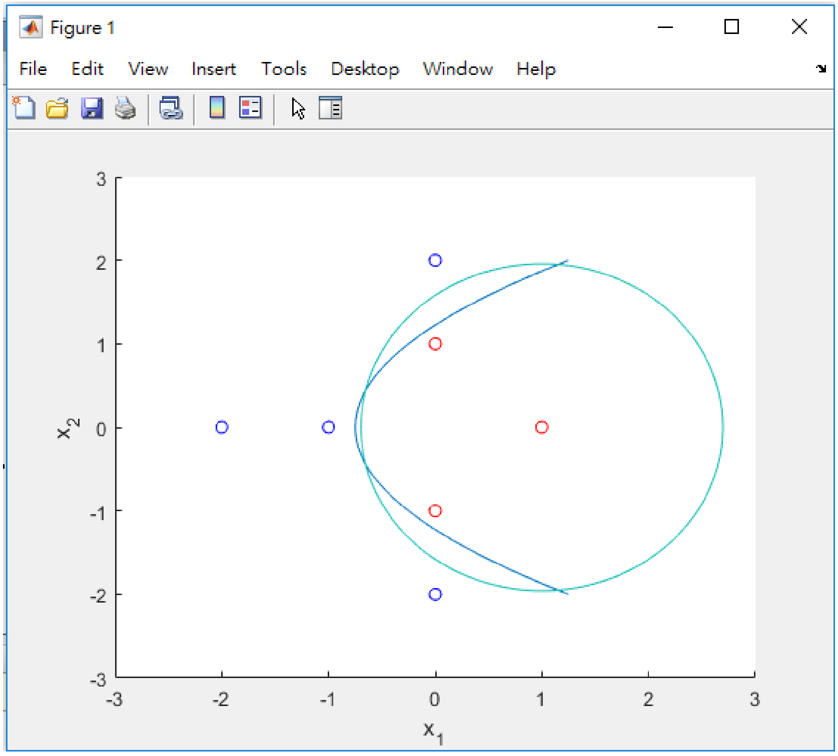
4.

不同，因為兩題的 transformation 不一樣。

Question 1 是使用$T_1(x_1,x_2) = (2x_2^2 - 4x_1 + 2, x_1^2 - 2x_2 - 3)$

Question 3 是使用$T_2(x_1,x_2) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_1x_2, x_2x_1, x_2^2)$

從圖形上也可看出不同



5.

$$\left(\frac{1}{\exp(-\mathrm{x}^2)}\right)^2 = \exp(2\mathrm{x}^2) = \frac{1}{0!} + \frac{2x^2}{1!} + \frac{(2x^2)^2}{2!} + \cdots + \frac{(2x^2)^n}{n!} + \cdots = \left\|\tilde{\phi}(x)\right\|^2$$

⇨ $\frac{1}{\exp(-\mathrm{x}^2)} = \|\tilde{\phi}(x)\|$ (兩者都是≥ 0，必定同號)

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

6.

$\cos(\mathrm{x},\mathrm{x}') = \cos(\theta) = \frac{\mathrm{x}\cdot\mathrm{x}'}{\|x\|\|x'\|} = (\frac{\mathrm{x}}{\|x\|}) \cdot \left(\frac{\mathrm{x}'}{\|x'\|}\right)$

所以 transform 為$\phi(\mathrm{x}) = \frac{\mathrm{x}}{\|x\|}$

所以 cos(x,x')是一 valid kernel

7.

由 lagrange multiplier 可將原本的目標函數和限制式寫成

$\min_{R,c}[R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$

令所有的$\lambda_\mathrm{n} \geq 0$，這樣如果有任何一個$(\|z_n - c\|^2 - R^2) \geq 0$，

則$\max_{\lambda_\mathrm{n}\geq 0}[R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$會跑向無限大並且被外面的 min 淘汰

也就是限制式可以用 max 來取代

可以寫成

$$\mathrm{L}(\mathrm{R},\mathrm{c},\lambda)=[R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$$

8.

原問題 P 為：$\mathrm{L}(\mathrm{R},\mathrm{c},\lambda)= [R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$

對偶問題 D 為：$\mathrm{L}(\mathrm{R},\mathrm{c},\lambda)= [R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$

Primal feasible：$\|z_n - c\|^2 \leq R^2$

Dual feasible：$\lambda_\mathrm{n} \geq 0$

Primal-inner optimal：$\lambda_n(\|z_n - c\|^2 - R^2) = 0$

Dual-inner optimal：

$$\frac{\sum_{n=1}^{N} \lambda_n z_n}{\sum_{n=1}^{N} \lambda_n} = 0 \quad \sum_{n=1}^{N} \lambda_n=1$$

解 dual problem：

$\frac{\partial}{\partial \mathrm{c}}[R^2 + \sum_{n=1}^{N} \lambda_n(\|z_n - c\|^2 - R^2)]$

⇨ $\sum_{n=1}^{N} \frac{\partial}{\partial c} \lambda_n \big((z_n - c) \cdot (z_n - c) - R^2 \big)$

⇨ $\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial c} (z_n \cdot z_n - 2z_n \cdot c + c \cdot c)$

⇨ $\sum_{n=1}^{N} \lambda_n (2c - 2z_n)$

⇨ $2c \sum_{n=1}^{N} \lambda_n - 2 \sum_{n=1}^{N} \lambda_n z_n$

令$\frac{\partial}{\partial \mathrm{c}} L(R, c, \lambda) = 0 \;\; (\sum_{n=1}^{N} \lambda_n \neq 0)$

⇨ $\frac{\sum_{n=1}^{N} \lambda_n z_n}{\sum_{n=1}^{N} \lambda_n}$=c=$\sum_{n=1}^{N} \lambda_n z_n$

$$\frac{\partial}{\partial \mathrm{R}} \left[R^2 + \sum_{n=1}^{N} \lambda_n (\|z_n - c\|^2 - R^2) \right]$$

⇨ 2R-2$\sum_{n=1}^{N} \lambda_n R$

令$\frac{\partial}{\partial \mathrm{R}} L(R, c, \lambda) = 0$

⇨ $\sum_{n=1}^{N} \lambda_n$=1

9.

L(R,c,λ)=$[R^2 + \sum_{n=1}^{N} \lambda_n (\|z_n - c\|^2 - R^2)]$

代入$\sum_{n=1}^{N} \lambda_n$=1

⇨ L(R,c,λ)=$\big[R^2 + \sum_{n=1}^{N} \lambda_n \big((z_n - c) \cdot (z_n - c) - R^2 \big) \big]$

⇨ L(R,c,λ)=$[R^2 + \sum_{n=1}^{N} \lambda_n (z_n - c) \cdot (z_n - c) - (\sum_{n=1}^{N} \lambda_n) R^2]$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n - c) \cdot (z_n - c)]$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n - 2z_n \cdot c + c \cdot c)]$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n) - 2c \sum_{n=1}^{N} \lambda_n \cdot z_n + c \cdot c \sum_{n=1}^{N} \lambda_n)]$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n) - 2c \sum_{n=1}^{N} \lambda_n \cdot z_n + c \cdot c)]$

代入 c=$\sum_{n=1}^{N} \lambda_n z_n$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n) - 2(\sum_{n=1}^{N} \lambda_n z_n)(\sum_{n=1}^{N} \lambda_n z_n) +$ $(\sum_{n=1}^{N} \lambda_n z_n)(\sum_{n=1}^{N} \lambda_n z_n)]$

⇨ L(R,c,λ)=$[\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n) - (\sum_{n=1}^{N} \lambda_n z_n)(\sum_{n=1}^{N} \lambda_n z_n)]$

所以

Objective(λ)= $\sum_{n=1}^{N} \lambda_n (z_n \cdot z_n) - (\sum_{n=1}^{N} \lambda_n z_n)(\sum_{n=1}^{N} \lambda_n z_n)$

Subject to $\sum_{n=1}^{N} \lambda_n$=1

10.

$$\sum_{n=1}^{N}\lambda_n(z_n\cdot z_n)-(\sum_{n=1}^{N}\lambda_n z_n)(\sum_{n=1}^{N}\lambda_n z_n)$$

$$=\ \sum_{n=1}^{N}\lambda_n(z_n\cdot z_n)-(\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m\, z_n\cdot z_m)$$

以 K(x,x')=$z^{T}z$代入

$$=\ \sum_{n=1}^{N}\lambda_n K(x_n\cdot x_n)-(\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m\, K(\, x_n\cdot x_m))$$

$$=-\sum_{n=1}^{N}\lambda_n(-K(x_n\cdot x_n))+\frac{1}{2}(\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m(-2)\cdot K(\, x_n\cdot x_m))$$

此形式可以使用 quadratic programming 來解

$\mathrm{Q}=[q_{n,m}]=-2K(x_n,x_m),\mathrm{p}=1_{\mathrm{N}}$

$$\mathrm{A}=\begin{bmatrix}1 & 1 & \dots & 1\\ -1 & -1 & \dots & -1\\ -1 & 0 & \dots & 0\\ 0 & -1 & \dots & 0\\ \dots & & & \\ 0 & 0 & \dots & -1\end{bmatrix}$$

C=[1,1,0,0,…,0]

λ=`quadprog(Q,p,A,c)`

解出的λ是一個 N 維度的向量，取其中大於 0 的第 i 個 element λ_{i}

代入 Primal-inner optimal：$\lambda_n(\|z_n-c\|^2-R^2)=0$

得到$\lambda_i(\|z_i-c\|^2-R^2)=0$

⇨ $\|z_i-c\|^2-R^2=0$

⇨ $\mathrm{R}=\sqrt{(\mathrm{z_i}-\sum_{n=1}^{N}\lambda_n z_n)\cdot(\mathrm{z_i}-\sum_{n=1}^{N}\lambda_n z_n)}$

⇨ $\mathrm{z_i}$和λ_{n}以及$\mathrm{z_n}$都是已知

⇨ 可以求出 R

11.

使用反證法

假設α^{*}不是最佳解，則存在一個$\alpha^{*\prime}$滿足

$\sum_{\mathrm{n}=1}^{\mathrm{N}}y_n\alpha_n=0$

$0\le\alpha_{\mathrm{n}}\le C,n=1,2,\dots,N$

並且能使

$$\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m z_n^T z_m-\sum_{n=1}^{N}\alpha_n$$

最小

則此$\alpha^{*\prime}$顯然也滿足 hard-margin 的條件

$\sum_{\mathrm{n}=1}^{\mathrm{N}}y_n\alpha_n=0$

$0\le\alpha_{\mathrm{n}}\ \ ,n=1,2,\dots,N$

並且比α^*能使目標函數

$$\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m z_n^T z_m - \sum_{n=1}^{N}\alpha_n$$

更小

則α^*不是 hard-margin SVM 的 optimal solution

⇒矛盾，因此α^*也應該是 soft-margin SVM 的 optimal solution

12.

Soft margin SVM dual problem 的目標函數和限制式為

$$\min_{\alpha}\frac{1}{2}\alpha^T Q\alpha + r^T\alpha$$

Subject to $A\alpha \leq u$

$Q=[q_{n,m} = y_n y_m K(x_n, x_m)]$

$A=\begin{bmatrix} y^T \\ -y^T \\ -I_{NxN} \\ I_{NxN} \end{bmatrix}$ $u=\begin{bmatrix} 0 \\ 0 \\ 0_{NxN} \\ CxI_{NxN} \end{bmatrix}$ $r=-1_N$

若 $\widetilde{K}(x,x') = pK(x,x')$且 $\tilde{C} = \frac{C}{p}$

則 $\tilde{Q}=[q_{n,m} = y_n y_m \widetilde{K}(x_n, x_m)]=[y_n y_m pK(x_n, x_m)]=pQ$

A 和 r 不變

$\tilde{u}=\begin{bmatrix} 0 \\ 0 \\ 0_{NxN} \\ \frac{C}{p}xI_{NxN} \end{bmatrix}=\frac{u}{p}$

因此，原本的目標函數可以改成

$$\frac{1}{2}\alpha^T\tilde{Q}\alpha + r^T\alpha = \frac{1}{2}\alpha^T(pQ)\alpha + r^T\alpha = \frac{1}{p}[\frac{1}{2}(p\alpha)^T Q(p\alpha) + r^T(p\alpha)]$$

因為 p>0，因此取 min 時可以忽略

原目標函數可寫成

$$\min_{\alpha}(\frac{1}{2}(p\alpha)^T Q(p\alpha) + r^T(p\alpha))$$

原本的限制式則可以改成

$$A\alpha \leq \tilde{u} = \frac{u}{p} \Rightarrow A(p\alpha) \leq u$$

因此，得到一個新的 QP 問題，且此新問題的形式和原問題相同

若原問題的解為α^*，則新問題的 pα 的解即為α^*

意即新問題的解為 $\frac{1}{p}\alpha^*$ 記為$\widetilde{\alpha}^*= \frac{1}{p}\alpha^*$

$$\tilde{b}^* = y_s - \sum_{n=1}^{N} \tilde{\alpha}_n^* y_n \tilde{K}(x_n, x_s)$$

$$= y_s - \sum_{n=1}^{N} \frac{1}{p} \alpha_n^* y_n p K(x_n, x_s)$$

$$= y_s - \sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_s) = b^*$$

$$\tilde{g}_{SVM}(x) = \mathrm{sign}\left(\sum_{n=1}^{N} \tilde{\alpha}_n^* y_n \tilde{K}(x_n, x) + \tilde{b}^*\right)$$

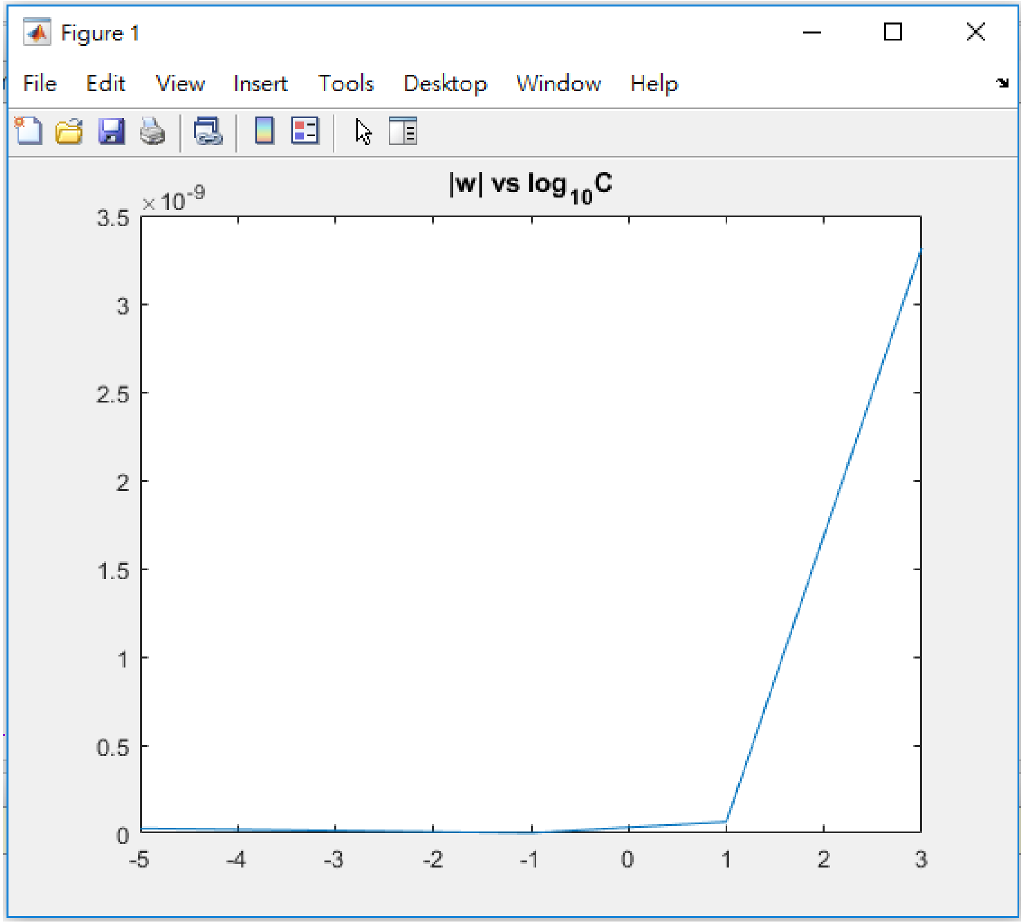
$$= \mathrm{sign}\left(\sum_{n=1}^{N} \left(\frac{1}{p}\alpha_n^*\right) y_n \left(pK(x_n, x)\right) + b^*\right)$$

$$= \mathrm{sign}\left(\sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x) + b^*\right)$$

$$= g_{SVM}(x)$$

因此新問題的 classifier 和原問題的 classifier 相同

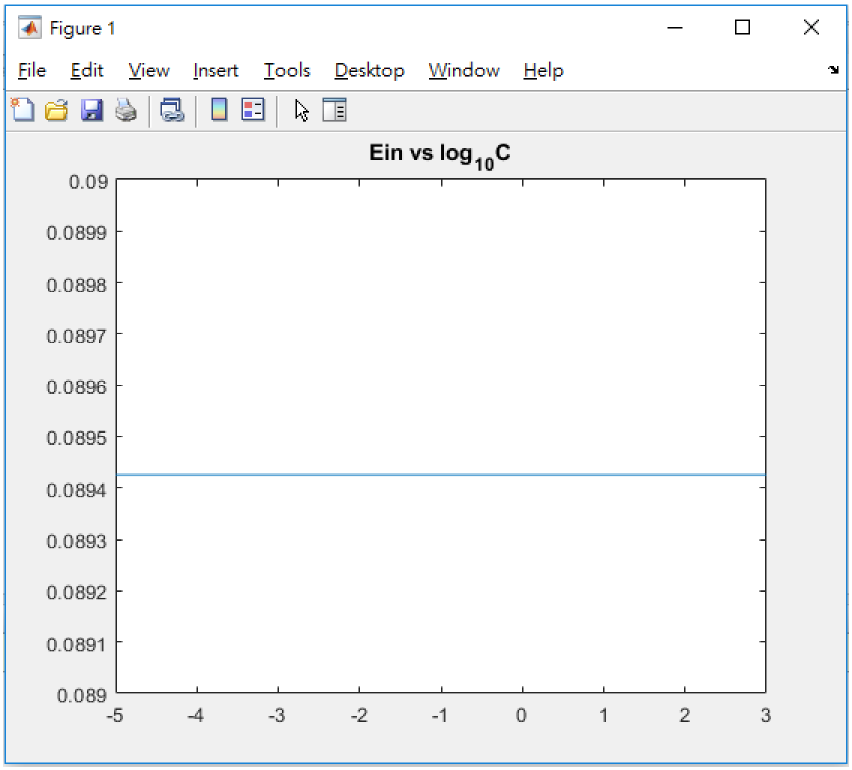
13.



C 代表發生 margin violation 時的懲罰係數，|w|則代表 margin 的大小。隨著 C 提高時，model 會盡量避免發生 margin violation，因此會縮減 margin 的大小，也就是增加|w|。

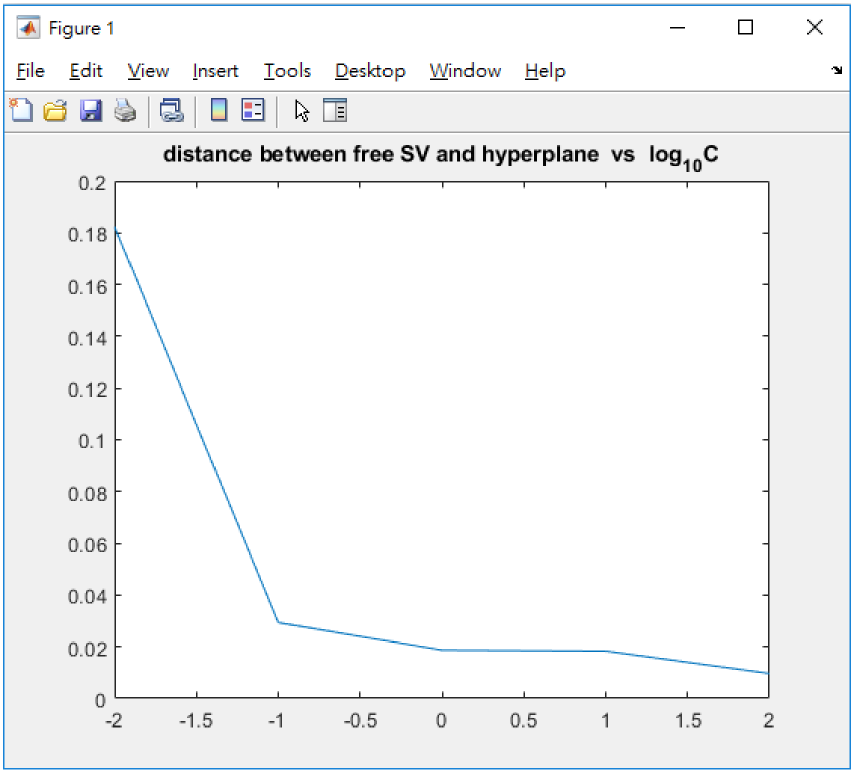
因此，|w|會隨著 C 增加而遞增。

14.



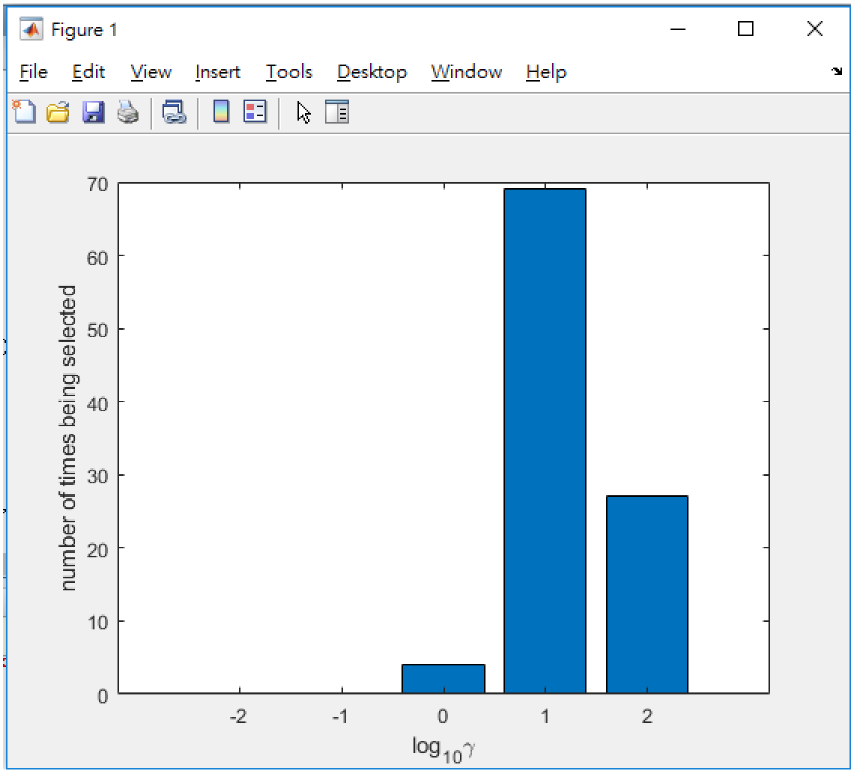
E_{in}隨著 C 提高，皆為一個介於 0.0894 和 0.0895 之間的數

15.



Free SV 和 hyperplane 的距離即為 margin 大小，隨著懲罰係數提高，margin 大小會被犧牲以換取較少的 margin violation，所以 Free SV 和 hyperplane 的距離會隨 C 遞減。

16.



大部分的最佳解都落在$\log_{10}\gamma = 1$時

17.

w_i代表 w 向量的第 i 個分量

因為 constant feature component $z_{n,i}$不論在 n 為多少時都一樣，因此可以從$\sum$中提出來

$w_i = \sum_{n=1}^{N} \alpha_n y_n z_{n,i} = z_{n,i} \sum_{n=1}^{N} \alpha_n y_n = 0 \quad (\because \sum_{n=1}^{N} \alpha_n y_n = 0)$

所以 constant feature component $z_{n,i}$ 對應到的w_i會是 0

18.

Soft margin SVM dual problem 的目標函數和限制式為

$$\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha + r^T\alpha$$

$$\text{Subject to } A\alpha \le u$$

$Q=[q_{n,m} = y_n y_m K(x_n, x_m)]$

$$A=\begin{bmatrix} y^T \\ -y^T \\ -I_{NxN} \\ I_{NxN} \end{bmatrix} \quad u=\begin{bmatrix} 0 \\ 0 \\ 0_{NxN} \\ CxI_{NxN} \end{bmatrix} \quad r=-1_N$$

若 $\tilde{K}(x, x') = K(x, x') + q$

則 $\tilde{Q}=[q_{n,m} = y_n y_m \tilde{K}(x_n, x_m)]=[y_n y_m (K(x_n, x_m) + q)]=Q + q * y^T y$

其中 y=$[y_1, y_2, \dots, y_N]$

A 和 r 和 u 不變

因此，新的目標函數為

$$\frac{1}{2}\alpha^T\tilde{Q}\alpha + r^T\alpha = \frac{1}{2}\alpha^T(\mathrm{Q} + \mathrm{q} * \mathrm{y}^\mathrm{T} y)\alpha + r^T\alpha = \frac{1}{2}\alpha^T Q\alpha + r^T\alpha$$

(因為$\sum_{\mathrm{n}=1}^{\mathrm{N}} \alpha_n y_n = 0 \Rightarrow$ yα=0)

和原本的目標函數相同

若原問題的解為α^*，則新問題的的解即為α^*，記為$\tilde{\alpha}^* = \alpha^*$

$$\begin{aligned}\tilde{b}^* &= y_s - \textstyle\sum_{n=1}^{N} \tilde{\alpha}_n^* y_n \tilde{K}(x_n, x_s) \\ &= y_s - \textstyle\sum_{n=1}^{N} \alpha_n^* y_n (K(x_n, x_s) + q) \\ &= y_s - \textstyle\sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_s) - q\textstyle\sum_{n=1}^{N} \alpha_n^* y_n = b^*\end{aligned}$$

$$\begin{aligned}\tilde{g}_{\mathrm{SVM}}(\mathrm{x}) &= \mathrm{sign}(\textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \tilde{\alpha}_n^* y_n \tilde{K}(x_n, x) + \tilde{b}^*) \\ &= \mathrm{sign}(\textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} (\alpha_n^*)\, y_n (K(x_n, x) + q) + \mathrm{b}^*) \\ &= \mathrm{sign}(\textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \alpha_n^* y_n \mathrm{K}(\mathrm{x_n}, \mathrm{x}) + \mathrm{q}\textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \alpha_n^* y_n + \mathrm{b}^*) \\ &= \mathrm{sign}(\textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \alpha_n^* y_n \mathrm{K}(\mathrm{x_n}, \mathrm{x}) + \mathrm{b}^*) \\ &= \mathrm{g}_{\mathrm{SVM}}(\mathrm{x})\end{aligned}$$

因此新問題的 classifier 和原問題的 classifier 相同