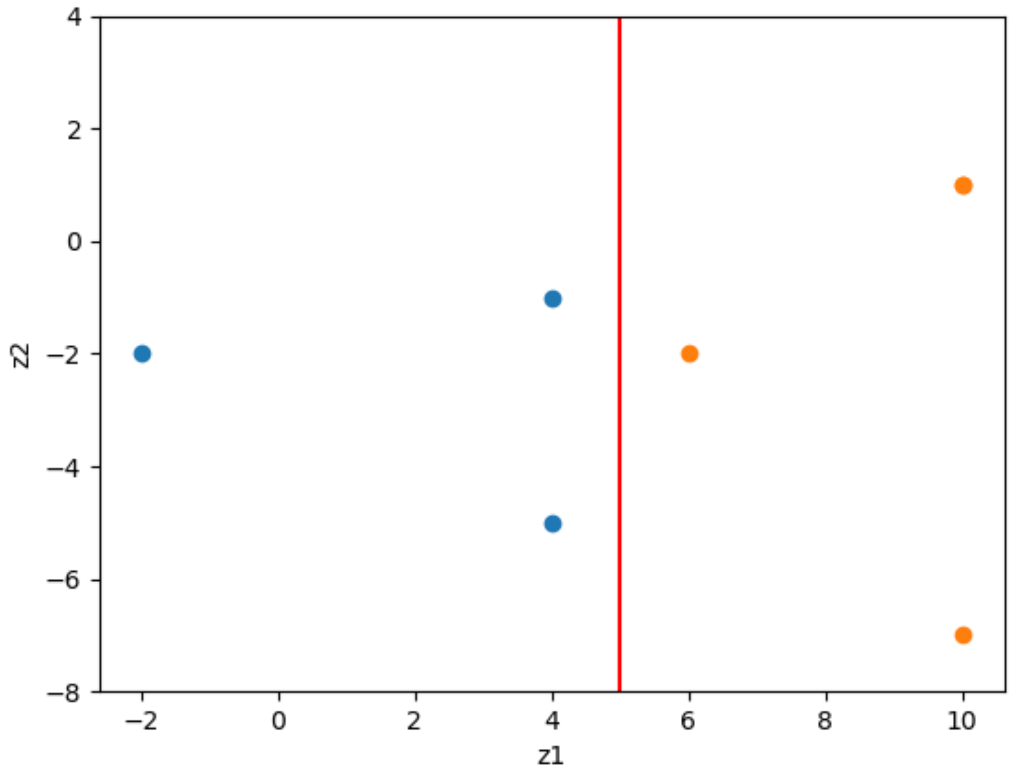
Machine Learning HW1

r07922100 資工碩一 楊力權

1.

$z_1 = (-2, -2), z_2 = (4, -5), z_3 = (4, -1), z_4 = (6, -2), z_5 = (10, -7), z_6 = (10,1), z_7 = (10,1)$



可知道在Z空間中最好的hyperplane是$z_1 = 5$

2.

使用package sklearn.svm.SVC，參數SVC(C=1e10, kernel='poly', coef0=1, degree=2, gamma=1)表示使用Hard-Margin、polynomial-2 kernel$(1 + x^T x)^2$

得到α如下

x	(1,0)	(0,1)	(0,-1)	(-1,0)	(0,2)	(0,-2)	(-2,0)
α	0.0	0.5965	0.8107	0.8887	0.2056	0.3128	0.0

support vector：(0,1), (0,-1), (-1,0), (0,2), (0,-2)

3.

$b = y_s - \sum_{SV \text{ indices n}} \alpha_n y_n K(x_n, x_s) = \sum_{SV \text{ indices n}} \alpha_n y_n (1 + x_n^T x_s)^2$其中$(x_s, y_s)$是SV

得到b = -1.6665544170710516

$\sum_{SV \text{ indices n}} \alpha_n y_n (1 + x_n^T x)^2 + b = 0$ 設$x = (v_1, v_2)$

則做內積可得含x=(v1, v2)的等式如下：

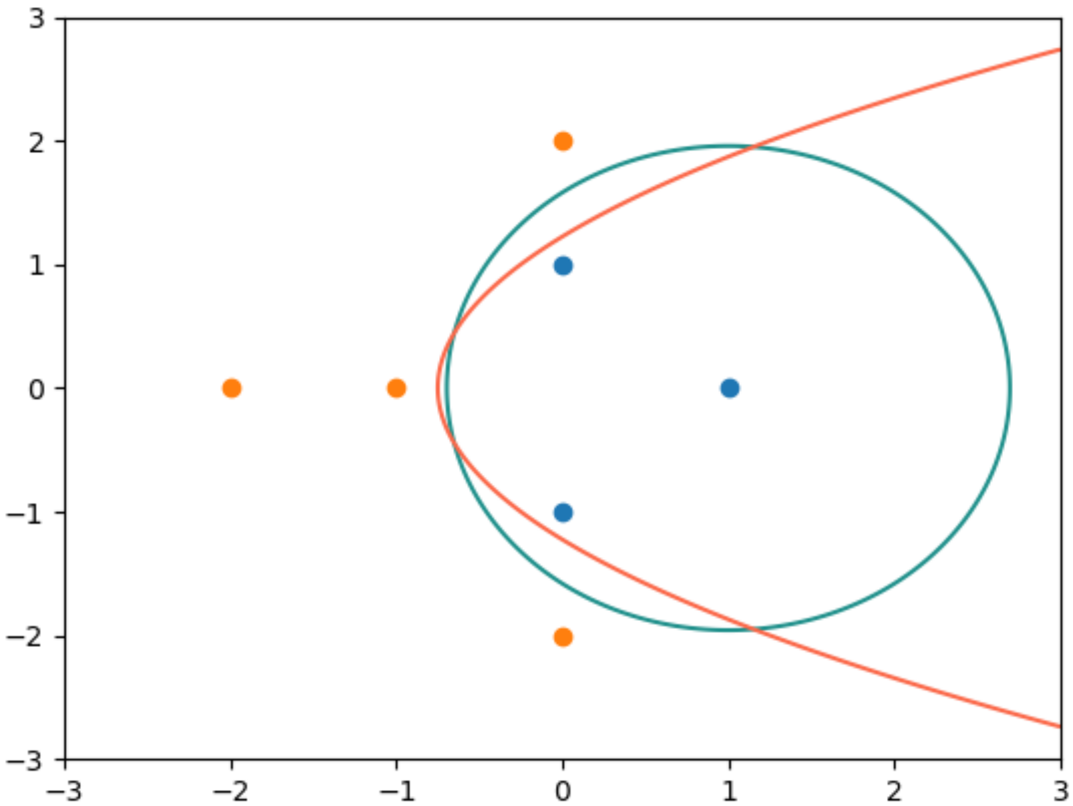
0.888703401608899*v1**2 − 1.7774068032178*v1 + 0.666554417071052*v2**2 + 2.22044604925031e-16*v2 − 1.66655441707105 = 0

4.

Q1是透過non-linear transform將2維空間轉成另一個2維空間做SVM。

Q3是透過non-linear transform將2維空間轉成7維空間做SVM。

因此結果如下圖並不相同(紅線為Q1，綠線為Q3)。



5.

$$\phi(x) = e^{-x^2}(1, \sqrt{\frac{2^1}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \cdots), \tilde{\phi}(x) = (1, \sqrt{\frac{2^1}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \cdots)$$

已知$\frac{\tilde{\phi}(x)^T\tilde{\phi}(x)}{\|\tilde{\phi}(x)\|^2} = 1$

$$\tilde{\phi}(x)^T\tilde{\phi}(x) = 1 + \frac{2^1}{1!}x^2 + \frac{2^2}{2!}x^4 + \cdots + \frac{2^n}{n!}x^{2n} + \cdots \overset{\text{Taylor}}{=} e^{2x^2}$$

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

6.

kernel function為 $\cos(x, x') = \dfrac{x^T x'}{\|x\|\|x'\|}$, transform為$\phi(x) = \dfrac{x}{\|x\|}$

且$\cos(x, x') = \dfrac{x^T x'}{\|x\|\|x'\|} = \dfrac{x'^T x}{\|x'\|\|x\|} = cos(x', x)$，K矩陣必symmetric

$$K = \begin{bmatrix} \phi(x_1)^T\phi(x_1) & \phi(x_1)^T\phi(x_2) & \cdots \\ \phi(x_2)^T\phi(x_1) & \phi(x_2)^T\phi(x_2) & \cdots \\ & \vdots & \end{bmatrix} = [\phi(x_1) \ \ \phi(x_2) \ \ \cdots \ \ \phi(x_n)]^T[\phi(x_1) \ \ \phi(x_2) \ \ \cdots \ \ \phi(x_n)]$$

必為半正定，符合Mercer’s condition。

7.

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

8.

KKT

$\|z_n - c\|^2 \leq R^2$

$\lambda_n \geq 0$

$\lambda_n(\|z_n - c\|^2 - R^2) = 0$

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

過程：

$$\frac{\partial L}{\partial c} = 0 \Rightarrow \sum_{n=1}^{N} \lambda_n(-2z_n + 2c) = 0 \Rightarrow c = \frac{\sum_{n=1}^{N} \lambda_n z_n}{\sum_{n=1}^{N} \lambda_n}$$

$\dfrac{\partial L}{\partial R} = 0 \Rightarrow 2R + \sum_{n=1}^{N} \lambda_n(-2R) = 0 \Rightarrow \sum_{n=1}^{N} \lambda_n = 1$ (R不等於0)

$$\because \sum_{n=1}^{N} \lambda_n = 1 \therefore c = \sum_{n=1}^{N} \lambda_n z_n$$

$\dfrac{\partial L}{\partial c} = 0 \Rightarrow \sum_{n=1}^{N} \lambda_n(-2z_n + 2c) = 0 \Rightarrow$ Optimal $c = \dfrac{\sum_{n=1}^{N} \lambda_n z_n}{\sum_{n=1}^{N} \lambda_n}$ if $\sum_{n=1}^{N} \lambda_n \neq 0$

9.

$c = \sum_{n=1}^{N} \lambda_n z_n$；$\sum_{n=1}^{N} \lambda_n = 1$更改原式

$$\max_{\lambda_n \geq 0} \min_{R,c} R^2 + \sum_{n=1}^{N} \lambda_n (\|z_n - c\|^2 - R^2)$$

$$\equiv \max_{\lambda_n \geq 0} R^2 + 1(-R^2) + \sum_{n=1}^{N} \lambda_n (z_n^T z_n - 2z_n^T c + c^T c)$$

$$\equiv \max_{\lambda_n \geq 0} \sum_{n=1}^{N} \lambda_n z_n^T z_n - 2\sum_{n=1}^{N} \lambda_n z_n^T c + \sum_{n=1}^{N} \lambda_n c^T c$$

$$\equiv \max_{\lambda_n \geq 0} \sum_{n=1}^{N} \lambda_n z_n^T z_n - 2\sum_{n=1}^{N} \lambda_n z_n^T \sum_{n=1}^{N} \lambda_n z_n + (\sum_{n=1}^{N} \lambda_n z_n)^T \sum_{n=1}^{N} \lambda_n z_n$$

$$\equiv \max_{\lambda_n \geq 0} \sum_{n=1}^{N} \lambda_n z_n^T z_n - (\sum_{n=1}^{N} \lambda_n z_n)^T \sum_{n=1}^{N} \lambda_n z_n$$

$$\equiv \max_{\lambda_n \geq 0} \sum_{n=1}^{N} \lambda_n z_n^T z_n - \sum_{n=1}^{N} \sum_{m=1}^{N} \lambda_n \lambda_m z_n^T z_m \text{ subject to } \sum_{n=1}^{N} \lambda_n = 1$$

10.

$z_n^T z_m = \phi(x_n)^T \phi(x_m) = K(x_n, x_m)$

$$\max_{\lambda_n \geq 0} \sum_{n=1}^{N} \lambda_n K(x_n, x_n) - \sum_{n=1}^{N} \sum_{m=1}^{N} \lambda_n \lambda_m K(x_n, x_m) \text{ subject to } \sum_{n=1}^{N} \lambda_n = 1$$

若一$\lambda_n > 0$則由$\lambda_n(\|z_n - c\|^2 - R^2) = 0$可知$\|z_n - c\|^2 - R^2 = 0$

$$R^2 = K(x_n, x_n) - 2z_n^T c + c^T c = K(x_n, x_n) - 2\sum_{m=1}^{N} \lambda_m K(x_n, x_m) + \sum_{m=1}^{N} \sum_{p=1}^{N} \lambda_m \lambda_p K(x_m, x_p)$$

$$R = \sqrt{K(x_n, x_n) - 2\sum_{m=1}^{N} \lambda_m K(x_n, x_m) + \sum_{m=1}^{N} \sum_{p=1}^{N} \lambda_m \lambda_p K(x_m, x_p)}$$

11.

SVM : $\min_{\alpha} \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum_{n=1}^{N} \alpha_n$

hard-margin： subject to $\sum_{n} y_n \alpha_n = 0,\ \alpha_n \geq 0$

soft-margin： subject to $\sum_{n} y_n \alpha_n = 0,\ C \geq \alpha_n \geq 0$

若hard-margin SVM的opt α^*非soft-margin SVM($C \geq \max \alpha^*$)的optimal解，則必能找到一組讓obj function更小的$C \geq \tilde{\alpha}^* \geq 0$，而$\tilde{\alpha}^* \geq 0$符合hard-margin SVM條件，說明$\alpha^*$非optimal(矛盾)⇒$\alpha^*$必也是soft-margin SVM的optimal解。

12.

使用$\tilde{K}(x,x')=pK(x,x')$；$\tilde{C}=\frac{C}{p}$

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

subject to $\sum_{n=1}^{N}y_n\alpha_n\geq 0,\ \sum_{n=1}^{N}(-y_n)\alpha_n\geq 0,\ \alpha_n\geq 0,\ -\alpha_n\geq -C$

原QP：$\min_{\alpha}\frac{1}{2}\alpha^T Q_D\alpha+(-1_N)^T\alpha$ subject to $A^T\alpha\geq c$

$$Q_D=\begin{bmatrix} q_{1,1} & q_{1,2} & \cdots \\ q_{2,1} & q_{2,2} & \cdots \\ & \vdots & \end{bmatrix}$$ 其中$q_{n,m}=y_n y_m K(x_n,x_m)$條件$A=\begin{bmatrix} y^T \\ -y^T \\ I_{NxN} \\ -I_{NxN}\end{bmatrix},c=\begin{bmatrix}0\\0\\0_N\\-C_N\end{bmatrix}$

使用$\tilde{K}(x,x')$後

$\tilde{q}_{n,m}=y_n y_m pK(x_n,x_m)$，因此$\tilde{Q}_D=pQ_D$

$$\min_{\alpha}\frac{1}{2}\alpha^T\tilde{Q}_D\alpha+(-1_N)^T\alpha\equiv\min_{\alpha}\frac{1}{p}(\frac{1}{2}(p\alpha)^T Q_D(p\alpha)+(-1_N)^T(p\alpha))$$

$\tilde{A}=A$，$\tilde{c}=\begin{bmatrix}0\\0\\0_N\\-(\frac{C}{p})_N\end{bmatrix}=\frac{c}{p}$，$\tilde{A}^T\alpha\geq\tilde{c}\equiv A^T\alpha\geq\frac{c}{p}\equiv A^T(p\alpha)\geq c$

新QP：$\min_{\alpha}\frac{1}{2}(p\alpha)^T Q_D(p\alpha)+(-1_N)^T(p\alpha)$ subject to $A^T(p\alpha)\geq c$

可以發現解$\tilde{K}(x,x')$QP的$\tilde{\alpha}*$與解原QP的$\alpha*$關係為$\tilde{\alpha}*=\frac{\alpha*}{p}$。

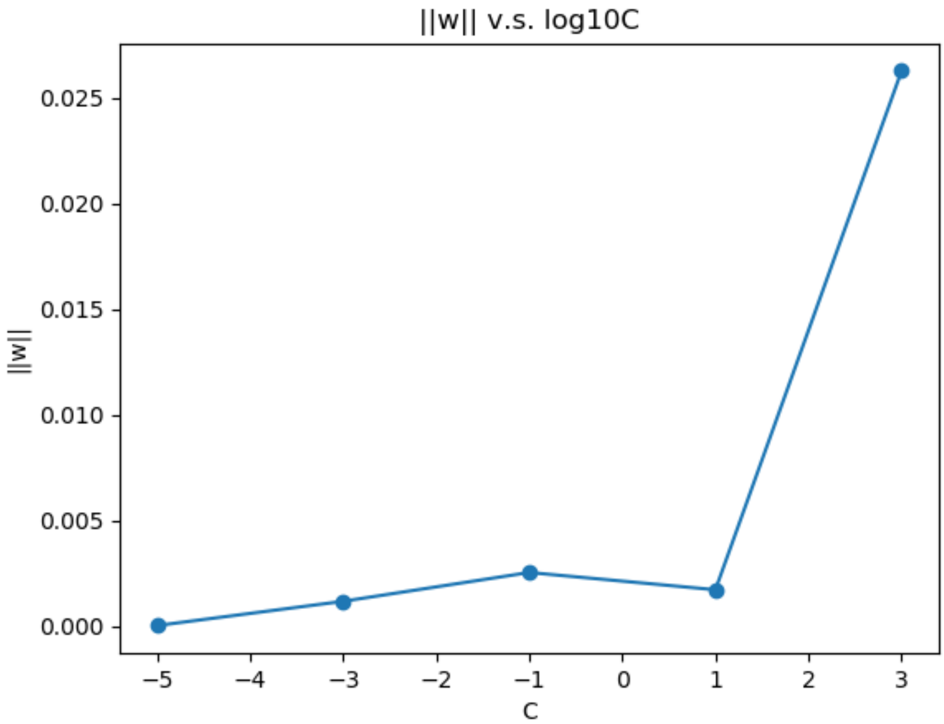
$$\tilde{w}=\sum_{n=1}^{N}\tilde{\alpha}_n y_n z_n=\frac{w}{p}$$

由complementary slackness，找一$0<\alpha_{sv}<\frac{C}{p}\Rightarrow\xi_{sv}=0$

$$\Rightarrow\tilde{b}=y_{sv}-\tilde{w}^T z_{sv}=y_{sv}-\sum_{n=1}^{N}\tilde{\alpha}_n y_n\tilde{K}(x_n,x_{sv})=y_{sv}-\sum_{n=1}^{N}\alpha_n y_n K(x_n,x_{sv})=b$$

$$\tilde{g}_{svm}(x)=sign(\sum_{n=1}^{N}\tilde{\alpha}_n y_n\tilde{K}(x,x_n)+\tilde{b})=sign(\sum_{n=1}^{N}\frac{\alpha_n}{p}y_n pK(x,x_n)+b)=g_{svm}(x)$$

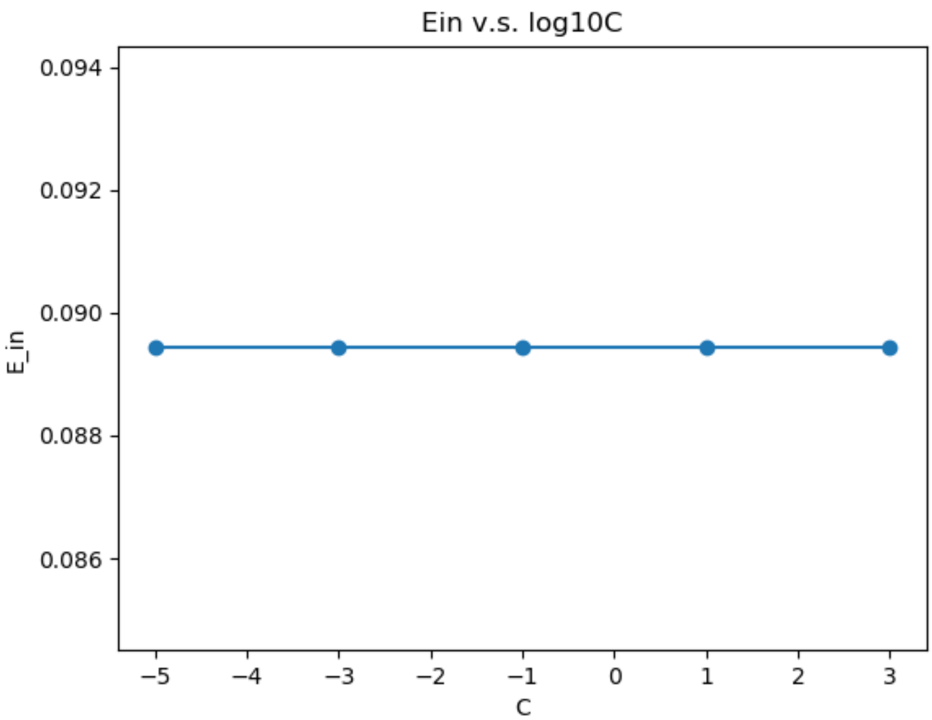
13.



soft-margin SVM的目標函數為：$\min_{b,w,\xi} = \frac{1}{2}w^T w + C\sum_{n=1}^{N}\xi_n$，因此C愈大，

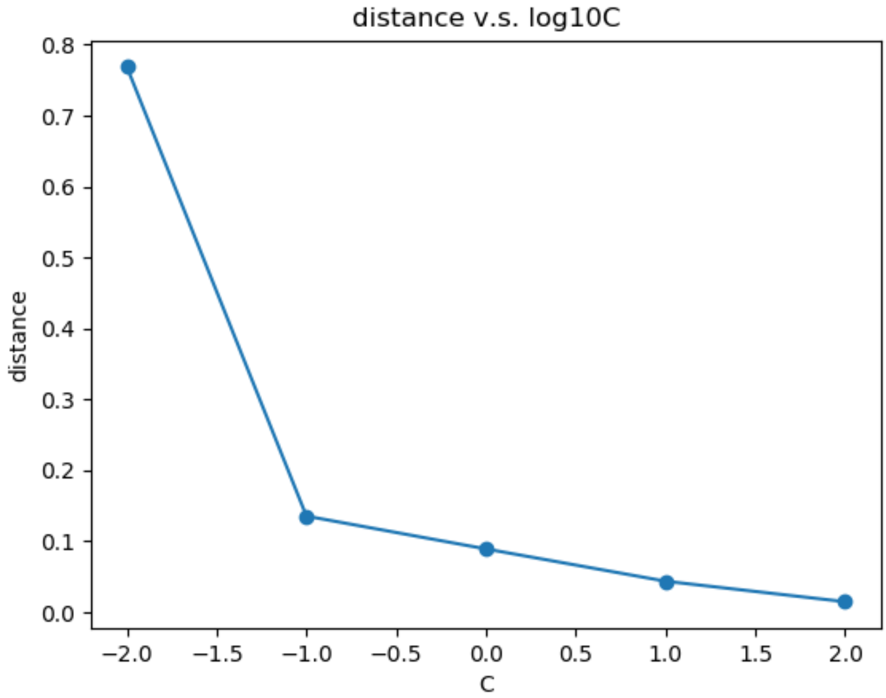
minimize $\|w\|$的成效比較差。

14.



不管哪種C對E_in都沒有差別。

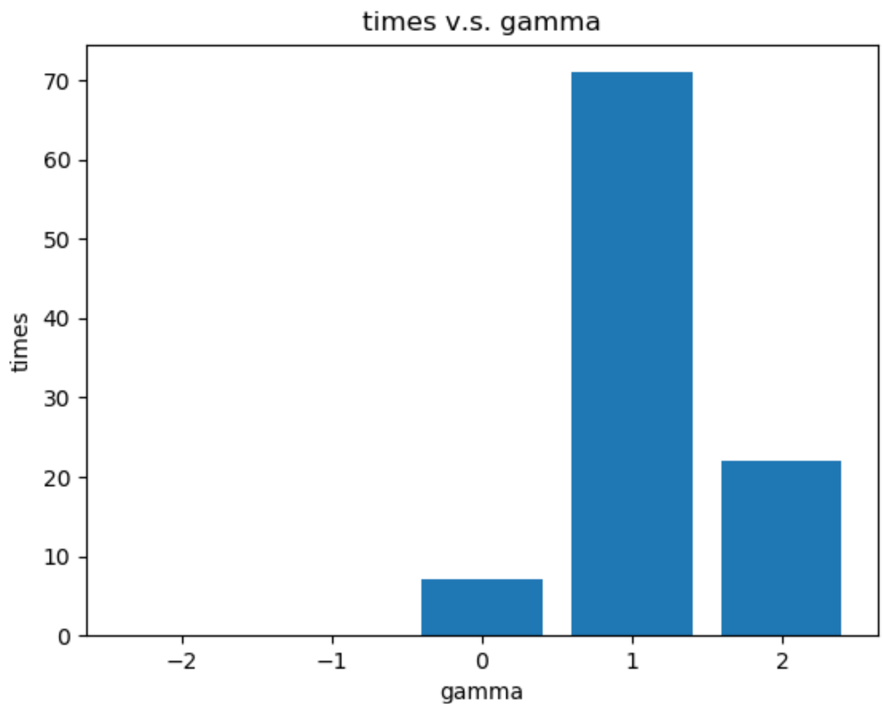
15.



distance $= \frac{1}{\|w\|} y_n(w_n^T x_n + b)$，取free support vector時，$\xi_n = 0, y_n(w_n^T x_n + b) = 1$

因此distance $= \frac{1}{\|w\|}$。所以C愈大的時候，distance會愈小也就是$\|w\|$愈大。

16.



17.

首要條件為 $\sum_{n=1}^{N} \alpha_n y_n = 0$

設Z的第一維度為constant feature C $\Rightarrow z = [C, \cdots]^T$。

optimal $w = \sum_{n=1}^{N} \alpha_n y_n z_n = \begin{bmatrix} C \sum_{n-1}^{N} \alpha_n y_n \\ \vdots \end{bmatrix} = [0, \cdots]^T$

因此 w 第一維為0。

18.

使用$\tilde{K}(x, x') = K(x, x') + q$

$$\min_{\alpha} \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum_{n=1}^{N} \alpha_n$$

subject to $\sum_{n=1}^{N} y_n \alpha_n \geq 0,\ \sum_{n=1}^{N} (-y_n) \alpha_n \geq 0,\ \alpha_n \geq 0,\ -\alpha_n \geq -C$

QP：$\min_{\alpha} \frac{1}{2} \alpha^T \tilde{Q}_D \alpha + (-1_N)^T \alpha$ subject to $A^T \alpha \geq c$

$$\tilde{Q}_D = \begin{bmatrix} \tilde{q_{1,1}} & \tilde{q_{1,2}} & \cdots \\ \tilde{q_{2,1}} & \tilde{q_{2,2}} & \cdots \\ & \vdots & \end{bmatrix} \text{其中} \tilde{q_{n,m}} = y_n y_m (K(x_n, x_m) + q) \text{條件} A = \begin{bmatrix} y^T \\ -y^T \\ I_{NxN} \\ -I_{NxN} \end{bmatrix}, c = \begin{bmatrix} 0 \\ 0 \\ 0_N \\ -C_N \end{bmatrix}$$

$$\tilde{Q}_D = \begin{bmatrix} q_{1,1} + q y_1 y_1 & q_{1,2} + q y_1 y_2 & \cdots \\ q_{2,1} + q y_2 y_1 & q_{2,2} + q y_2 y_2 & \cdots \\ & \vdots & \end{bmatrix} = Q_D + q y^T y$$

因為 $\sum_{n=1}^{N} \alpha_n y_n = 0$

$$\min_{\alpha} \frac{1}{2} \alpha^T \tilde{Q}_D \alpha + (-1_N)^T \alpha \equiv \min_{\alpha} \frac{1}{2} \alpha^T (Q_D + q y^T y) \alpha + (-1_N)^T \alpha$$

$$\equiv \min_{\alpha} \frac{1}{2} \alpha^T Q_D \alpha + (-1_N)^T \alpha + \frac{q}{2} \alpha^T y^T y \alpha \equiv \min_{\alpha} \frac{1}{2} \alpha^T Q_D \alpha + (-1_N)^T \alpha + 0$$

因此使用$\tilde{K}(x, x') = K(x, x') + q$與$K(x, x')$解出來的 α 相同。