Problem 2.

我用 sklearn 去實作 kernel SVM，因為是 hard margin SVM，所以我把 C 設 1e10，才不會有錯誤發生，其他參數如題目所述，常數 ζ 設 1，γ 設 1，degree 設 2。最後求出的 dual coefficient 是 support vector 所乘上的係數，也就是 y*α，所以真正的 α 值還需要乘上 y：

α = [0.0, 0.59647182, 0.81065085, 0.8887034, 0.20566488, 0.31275439, 0.0]

由上面 α 可以判斷出，support vector 為 α > 0 的那幾個點，也就是：

SV = [(0, 1), (0, -1), (-1, 0), (0, 2), (0, -2)]

Source code:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import svm

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

C = 1e10
svm_poly = svm.SVC(C=C, kernel = 'poly',coef0 = 1, gamma = 1, degree = 2).fit(x,y)

print(svm_poly.dual_coef_)
print("support vectors: ", svm_poly.support_vectors_)
```

Problem 3.

$g_{curve} = wz + b$

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

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

已知 $K(x, x') = (1+x^T x')^2 = 1 + 2x^T x' + (x^T x')(x^T x')$

$w = \sum_{n=1}^{N} \alpha_n y_n [x_n[0]^2, x_n[1]^2, 2x_n[0], 2x_n[1], 1]$

$g_{curve} = 0.8887x_0^2 + 0.6666\, x_1^2 - 1.7774\, x_0 - 1.6666$

Source code:

```
x = np.array(x)
b = 0
for i,ya in zip(svm_poly.support_, svm_poly.dual_coef_[0]):
    b += (ya*(np.dot(x[1],x[i])+1)**2)
b = (y[1] - b)
print(b)

w = [0,0,0,0,0]
np.array(w)
for i,ya in zip(svm_poly.support_, svm_poly.dual_coef_[0]):
    w += np.array([ya*1, ya*2*x[i][0], ya*2*x[i][1], ya*x[i][0]**2, ya*x[i][1]**2])
print('w=',w)
```

Problem 4.

Q1 curve: φ 1(x) = 2x 2 2 − 4x1 + 2 = 5

Q3 curve: $0.8887x_0^2 + 0.6666\, x_1^2 - 1.7774\, x_0 - 1.6666 = 0$

兩個 curve 長得不一樣，因為他們是由不同 z 空間最佳化出來的。

Problem 7.

根據 Lagrange function，可以直接得到：

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

當$\|x_n - c\| \leq R^2, \lambda_n \geq 0$:

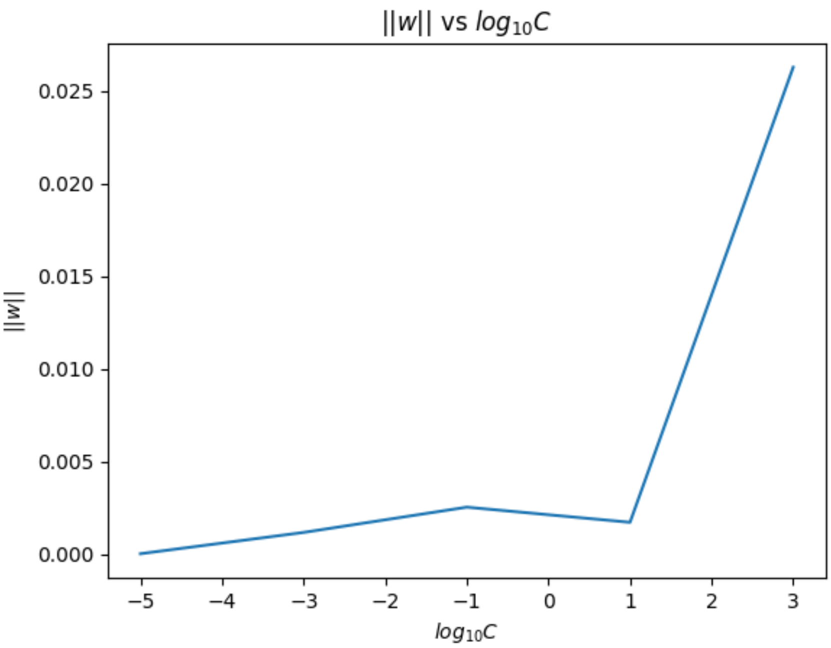
max $\lambda_n (\|z_n - c\|^2 - R^2) = 0$;for n = 1,2,⋯⋯,N

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

min max $L(R, c, \lambda) =$ min R^2

Problem 13.

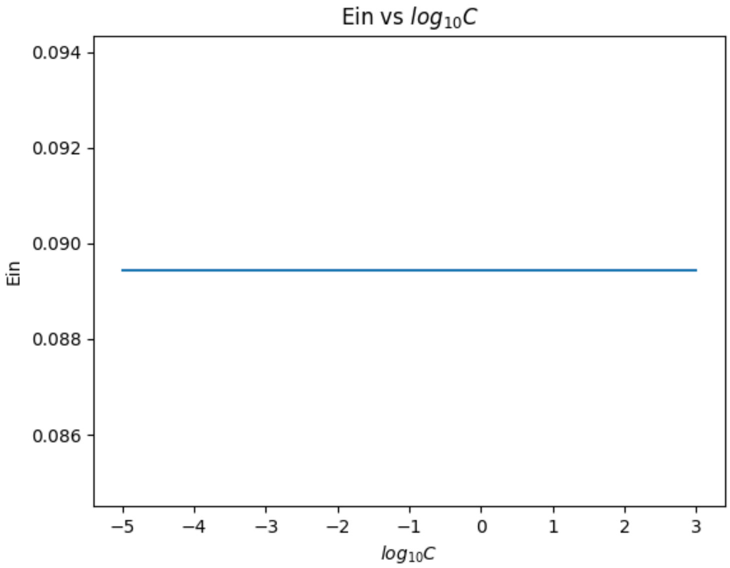
我用 sklearn 來作 SVM，最後結果如下：



可以看到 C 越大||w||也越大，因為 C 越大代表這個 SVM 越不能忍受有 X 犯錯，因此會越來越接近 hard-margin SVM，而代價就是為了追求正確率，margin = 1/||w||會越來越小。由此可知 C 跟||w||兩者成正相關。

Problem 14.

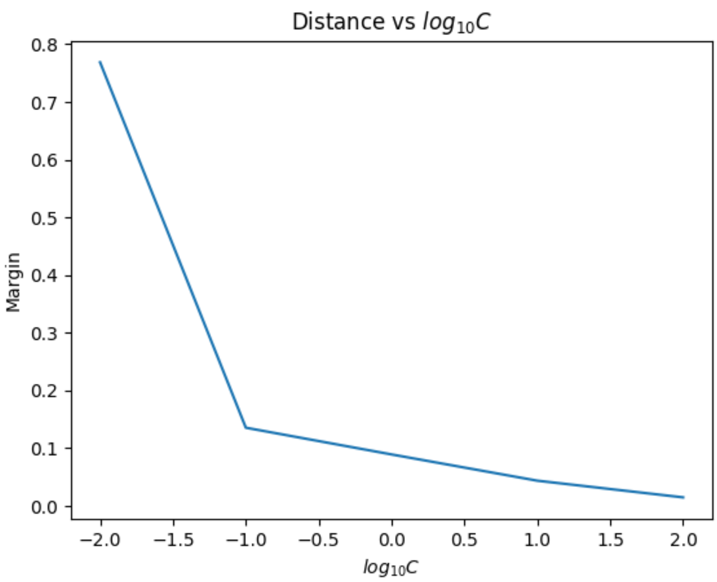
我用 sklearn 來作 SVM，最後結果如下：



理論上來說 C 越大 Ein 應該要越小，但沒想到 Ein 做出來都一樣。於是我把 predict X_train 的結果印出來，發現全部都是-1，而且 Ein 的值剛好跟 (Y == True)/(num of Y)一樣，因此推測得出所有 C 的 g_{svm} 都為-1，因此 Ein 才會都一樣。

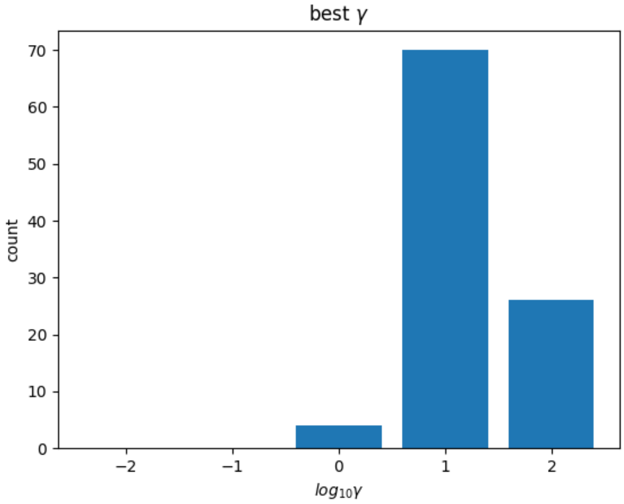
Problem 15.

我用 sklearn 來作 SVM，最後結果如下：



這裡的 distance 就是 margin = 1/llwll，而 C 越大代表能容忍的錯誤越少，但也因此會犧牲 margin 來換取成功率，所以 C 越大 margin 也會越小。

Problem 16.



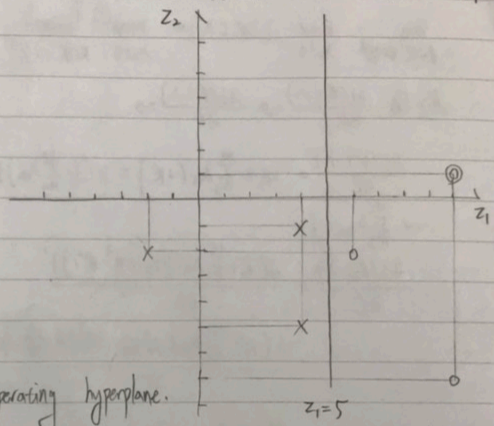
可以發現幾乎都是 gamma = 1 居多，而完全沒有-1, -2。

b05902050
黃子源

No.
Date: 2019 3 11

ML HW1

Transforms: Explicit v.s. Implicit.



1. $x_1=(1,0) \Rightarrow z_1=(-2,-2)$ $y_1=-1$

$x_2=(0,1) \Rightarrow z_2=(4,-5)$ $y_2=-1$

$x_3=(0,-1) \Rightarrow z_3=(4,-1)$ $y_3=-1$

$x_4=(-1,0) \Rightarrow z_4=(6,-2)$ $y_4=+1$

$x_5=(0,2) \Rightarrow z_5=(10,-7)$ $y_5=+1$

$x_6=(0,-2) \Rightarrow z_6=(10,1)$ $y_6=+1$

$x_7=(-2,0) \Rightarrow z_7=(10,1)$ $y_7=+1$

由圖片可看出 $z_1=5$ 是 equation of the optimal seperating hyperplane.
無法畫出对 z_2, z_3, z_4 margin都大於1的線.

5.
$$\exp(-(x-x')^2) = \exp(-(x)^2)\exp(-(x')^2)\exp(2xx')$$
$$\overset{\text{Taylor}}{=} \exp(-x^2)\exp(-x'^2)\left(\sum_{i=0}^{\infty}\frac{(2xx')^i}{i!}\right)$$

$$\|\Phi(x)\| = 1+\frac{2}{1!}x^2+\frac{4}{2!}x^4+\cdots\frac{2^k}{k!}x^{2k}\cdots$$
$$=\sum_{i=0}^{\infty}\frac{(2x^2)^i}{i!} \ \#$$

when $x=x'$:
$$\exp(-(x-x)^2) = 1 = \exp(-x^2)^2\cdot\left(\sum_{i=0}^{\infty}\frac{(2x^2)^i}{i!}\right)$$

$$\Rightarrow \exp(-x^2) = \frac{1}{\sqrt{\sum_{i=0}^{\infty}\frac{(2x^2)^i}{i!}}} = \frac{1}{\|\Phi(x)\|} \ \#$$

6.
$$x\cdot x' = \|x\|\,\|x'\|\cos\theta$$
$$\cos(x\cdot x') = x\cdot x'\frac{1}{\|x\|}\frac{1}{\|x'\|} = \phi'(x)\cdot\phi'(x')$$
$$\phi'(x) = \frac{1}{\|x\|}\cdot x = \frac{1}{\sqrt{\sum_{i=0}^{d}x_i^2}}\cdot(x_1, x_2, \cdots x_d) \ \#$$

8.

$$\min_{R\in\mathbb{R}\, c\in\mathbb{R}^d} \max_{\lambda_n\geq 0} L(R.c.\lambda) = \max_{\lambda_n\geq 0} \min_{R\in\mathbb{R}.\, c\in\mathbb{R}^d} L(R.c.\lambda)$$

極值: $\frac{\partial L(R.c.\lambda)}{\partial R}=0 \quad \frac{\partial L(R.c.\lambda)}{\partial c}=0$

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

$$\Rightarrow \sum_{n=1}^{N}\lambda_n = 1$$

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

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

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

$$\Rightarrow c=\frac{\sum_{n=1}^{N}\lambda_n z_n}{\sum_{n=1}^{N}\lambda_n} \quad \because \sum_{n=1}^{N}\lambda_n=1 \Rightarrow c=\sum_{n=1}^{N}\lambda_n z_n$$

prove that if $\sum_{n=1}^{N}\lambda_n\neq 0$, $c=\sum_{n=1}^{N}\lambda_n z_n/\sum_{n=1}^{N}\lambda_n$ #

KKT conditions:

① $\|z_n-c\|^2\leq R^2$ for $n=1,2,\dots N$ // primal feasible

② $\lambda_n\geq 0$ // dual feasible

③ $\sum_{n=1}^{N}\lambda_n=1$; $c=\sum_{n=1}^{N}\lambda_n z_n$ // dual - inner optimal

④ $\lambda_n(\|z_n-c\|^2-R^2)=0$ for $n=1,2,\dots N$ // primal - inner optimal.

9

已知 $\sum_{n=1}^{N}\lambda_n=1$，$C=\sum_{n=1}^{N}\lambda_n z_n$ 當極值發生時.

$$\max_{\lambda_n\geq 0}\ \min_{R\in\mathbb{R},C\in\mathbb{R}^d} L(R,C,\lambda)=\max_{\lambda_n\geq 0}\ \min_{R\in\mathbb{R},C\in\mathbb{R}^d} R^2+\sum_{n=1}^{N}\lambda_n(\|z_n-C\|^2-R^2)$$

$$=\max_{\lambda_n\geq 0} \cancel{R^2}+\sum_{n=1}^{N}\lambda_n\|z_n-C\|^2-\cancel{R^2\sum_{n=1}^{N}\lambda_n}$$

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

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

$$=\max_{\lambda_n\geq 0} \sum_{n=1}^{N}\lambda_n z_n^T z_n-\left(\sum_{n=1}^{N}\lambda_n z_n^T\right)\left(\sum_{n=1}^{N}\lambda_n z_n\right) \quad \text{subject to } \sum_{n=1}^{N}\lambda_n=1 \ \#$$

10.

$$\max_{\lambda_n\geq 0}\ \min_{R\in\mathbb{R},C\in\mathbb{R}^d} L(R,C,\lambda)=\max_{\lambda_n\geq 0} \sum_{n=1}^{N}\lambda_n z_n^T z_n-\left(\sum_{n=1}^{N}\lambda_n z_n^T\right)\left(\sum_{n=1}^{N}\lambda_n z_n\right) \quad \text{subject to } \sum_{n=1}^{N}\lambda_n=1$$

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

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

可用QP solver 解：$q_{n,m}=-K(x_n,x_m)$，$p_n=K(x_n,x_n)$，$a_\geq=1_N$，$a_\leq=-1_N$，$a_n^T=$ n-th unit direction
$c_\geq=1$，$c_\leq=-1$，$c_n=0$

根據KKT condition，$\lambda_n(\|z_n-C\|^2-R^2)=0$，若有 i 使 $\lambda_i>0$，$\|z_i-C\|^2-R^2=0$

$$R^2=\|z_i-C\|^2=z_i^Tz_i-2z_i^TC+C^TC$$
$$=K(x_i,x_i)-2z_i^T\sum_{n=1}^{N}\lambda_n z_n+\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m z_n z_m$$
$$=K(x_i,x_i)-2\sum_{n=1}^{N}\lambda_n K(x_i,x_n)+\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m K(x_n,x_m)$$
$$R=\left(K(x_i,x_i)-2\sum_{n=1}^{N}\lambda_n K(x_i,x_n)+\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m K(x_n,x_m)\right)^{1/2} \ \#$$

11.

根據上課內容，上述問題跟以下 Lagrange dual problem 等架：

$$\max_{0\le\alpha_n\le C,\ \beta_n=C-\alpha_n}\left(\min_{b,w}\ \frac{1}{2}w^Tw+\sum_{n=1}^{N}\alpha_n(1-y_n(w^Tz_n+b))\right)$$

設此式的 optimal solution 為 $\alpha' \neq \alpha^*$

对於 hard-margin SVM 的 Lagrange dual problem：

$$\max_{0\le\alpha_n}\left(\min_{b,w}\ \frac{1}{2}w^Tw+\sum_{n=1}^{N}\alpha_n(1-y_n(w^Tz_n+b))\right)$$

$\because C \geq \max_{1\le n\le N}\alpha_n^*$ $\therefore \alpha'$ 的结果一定比 α^* 好

但如此一来 hard-margin SVM opimal solution 不會是最佳解 $\Rightarrow$ contradiction!

$\Rightarrow \alpha^*$ is also an optimal solution to the soft-margin SVM. #

17.

let z_i be a constant feature component, corresponds to optimal weight value w.

Assume $z_i = C$

$w_i=\sum_{n=1}^{N}y_n\alpha_n z_{ni} = C\sum_{n=1}^{N}y_n\alpha_n$，已知 $\Sigma y_n\alpha_n=0$ (KKT condition)

$\Rightarrow w_i = C\times 0 = 0$ #

12.

原本的 SVM 問題 QP solver:

$q_{n,m} = y_n y_m K(x_n, x_m) \quad p = -1_N$

$$A = \begin{bmatrix} y^T \\ -y^T \\ I_N \\ -I_N \end{bmatrix} \quad V = \begin{bmatrix} 0 \\ 0 \\ 0_N \\ C \times I_N \end{bmatrix}$$

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

$\tilde{q}_{n,m} = y_n y_m \tilde{K}(x_n, x_m) = p y_n y_m K(x_n, x_m) \quad \tilde{p} = -1_N$

$$\tilde{A} = \begin{bmatrix} y^T \\ -y^T \\ I_N \\ -I_N \end{bmatrix} \quad \tilde{V} = \begin{bmatrix} 0 \\ 0 \\ 0_N \\ \tilde{C} \times I_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0_N \\ \frac{C}{p} I_N \end{bmatrix} \quad \Rightarrow \tilde{Q} = pQ, \ \tilde{V} = \frac{1}{p} V$$

原函式為:

$\min \frac{1}{2}\alpha^T \tilde{Q} \alpha + \tilde{p}^T \alpha = \frac{1}{2} p \alpha^T Q \alpha + 1_N \alpha$ —①

subject to $\tilde{A}\alpha \geq \tilde{V} \Rightarrow A\alpha \geq \frac{1}{p} V$ —②

$\because p > 0$, ①同乘 p 最佳解不變

$\min \frac{1}{2} p^2 \alpha^T Q \alpha + p 1_N \alpha = \min \frac{1}{2}(p\alpha^T) Q (p\alpha) + 1_N (p\alpha)$

subject to $A(p\alpha) \geq V$

把 $p\alpha$ 替換成 α', 則此 QP 問題跟原問題一樣

設原問題最佳解為 α^*, 而此 QP 問題最佳解為 $\tilde{\alpha}^*$

$p\tilde{\alpha}^* = \alpha^*$

原問題的 $g_{SVM} = \text{sign}\left(\sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x) + b\right)$

$b = y_i - w^T z_i = y_i - \sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_i)$

新問題的 $\tilde{g}_{SVM} = \text{sign}\left(\sum_{n=1}^{N} \frac{1}{p} \alpha_n^* y_n \tilde{K}(x_n, x) + \tilde{b}\right)$

$\tilde{b} = y_i - \tilde{w}^T z_i = y_i - \sum_{n=1}^{N} \frac{1}{p} \alpha_n^* y_n \tilde{K}(x_n, x_i) = y_i - \sum_{n=1}^{N} \frac{1}{p} \alpha_n^* y_n p K(x_n, x_i) = b$

$\tilde{g}_{SVM} = \text{sign}\left(\sum_{n=1}^{N} \frac{1}{p} \alpha_n^* y_n p K(x_n, x) + b\right) = g_{SVM}$ #

18.

原本的QP solver:

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

當 $\tilde{K}(x, x') = K(x, x') + q$:

$\tilde{q}_{nm} = y_n y_m \tilde{K}(x_n, x_m) = y_n y_m (K(x_n, x_m) + q)$

$\Rightarrow \tilde{Q} = Q + q y^T y$

原函式為:

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

$\because$ 限制 $\sum_{n=1}^{N} y_n \alpha_n = 0 \Rightarrow y\alpha = 0$

$\frac{1}{2}\alpha^T \tilde{Q} \alpha + p\alpha = \frac{1}{2}\alpha^T Q \alpha + \frac{1}{2}\alpha^T q y^T \overset{=0}{y\alpha} + p\alpha = \frac{1}{2}\alpha^T Q \alpha + p\alpha$

因此該QP問題的 $\underbrace{\tilde{\alpha}^*}_{\text{最佳解}}$ 跟原問題的 $\underbrace{\alpha^*}_{\text{最佳解}}$ 相同 ($\because$ Q, p, A, C 都不變)

原問題的 $g_{SVM} = \text{sign}\left(\sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x) + b\right)$

$b = y_i - w^T z_i = y_i - \sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_i)$

新問題的 $\tilde{g}_{SVM} = \text{sign}\left(\sum_{n=1}^{N} \alpha_n^* y_n \tilde{K}(x_n, x) + \tilde{b}\right)$

$\tilde{b} = y_i - \tilde{w}^T z_i = y_i - \sum_{n=1}^{N} \alpha_n^* y_n \tilde{K}(x_n, x_i)$

$= y_i - \sum_{n=1}^{N} \alpha_n^* y_n (K(x_n, x_i) + q) = y_i - \sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_i) - q\sum_{n=1}^{N} \alpha_n^* y_n$

已知 $\sum_{n=1}^{N} \alpha_n^* y_n = 0$:

$\tilde{b} = y_i - \sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x_i) - 0 = b$

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

$= \text{sign}\left(\sum_{n=1}^{N} \alpha_n^* y_n K(x_n, x) + q\overset{=0}{\sum_{n=1}^{N} \alpha_n^* y_n} + b\right)$

$= g_{SVM}$ #