Machine Learning Techniques - Homework 1

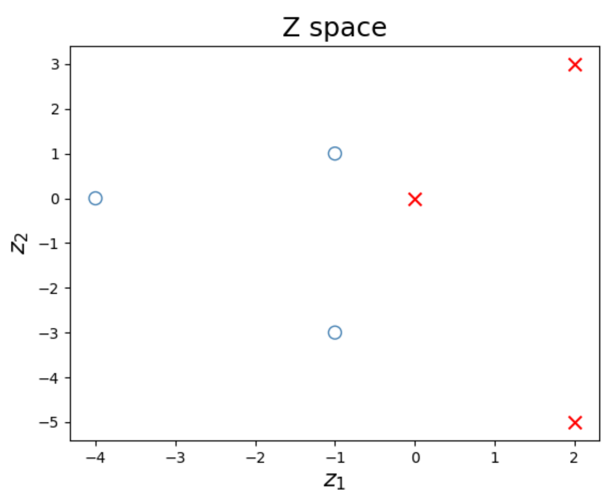
資工四 B05902023 李澤諺

Transforms: Explicit versus Implicit

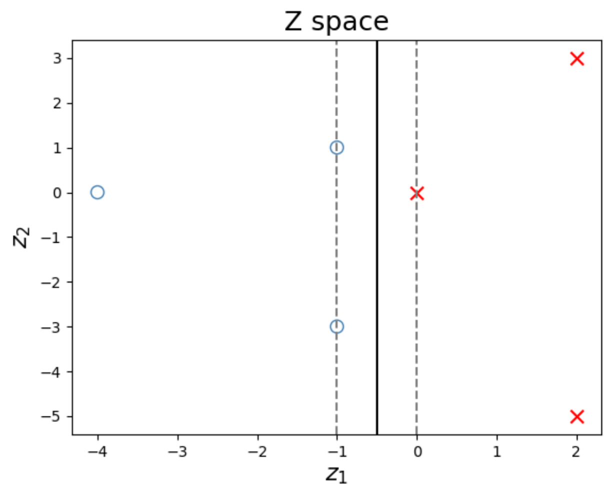
1. 將 input vector 經過題幹中的 transformation 之後會變爲

$$
\begin{aligned}
\mathbf{z}_1 &= (\phi_1(\mathbf{x}_1), \phi_2(\mathbf{x}_1)) = (-4, 0) \\
\mathbf{z}_2 &= (\phi_1(\mathbf{x}_2), \phi_2(\mathbf{x}_2)) = (-1, -3) \\
\mathbf{z}_3 &= (\phi_1(\mathbf{x}_3), \phi_2(\mathbf{x}_3)) = (-1, 1) \\
\mathbf{z}_4 &= (\phi_1(\mathbf{x}_4), \phi_2(\mathbf{x}_4)) = (0, 0) \\
\mathbf{z}_5 &= (\phi_1(\mathbf{x}_5), \phi_2(\mathbf{x}_5)) = (2, -5) \\
\mathbf{z}_6 &= (\phi_1(\mathbf{x}_6), \phi_2(\mathbf{x}_6)) = (2, 3) \\
\mathbf{z}_7 &= (\phi_1(\mathbf{x}_7), \phi_2(\mathbf{x}_7)) = (2, 3)
\end{aligned}
$$

以上的 transformed vector 在 $\mathcal{Z}$ space 中的分佈如下圖所示



由上圖可以看出，在 $\mathcal{Z}$ space 中的 optimal separating hyperplane 爲 $z_1 = -0.5$ (在 $\mathcal{X}$ space 中爲 $x_2^2 - 2x_1 - 2 = -0.5$，即 $x_2^2 - 2x_1 - 1.5 = 0$)。



2. 以下爲我實作的程式

```
import numpy as np
from sklearn.svm import SVC

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])

classifier = SVC(C = np.inf, kernel = 'poly',
                 degree = 2, gamma = 1, coef0 = 1)
classifier.fit(x, y)

print('support vector:')
print(classifier.support_vectors_)
print('y * alpha:')
print(classifier.dual_coef_)
```

由此程式可以得到

$$\alpha_1 = 0$$
$$\alpha_2 = 0.59647182$$
$$\alpha_3 = 0.81065085$$
$$\alpha_4 = 0.8887034$$
$$\alpha_5 = 0.20566488$$
$$\alpha_6 = 0.31275439$$
$$\alpha_7 = 0$$

並且可得 support vector 爲 $\mathbf{x}_2$、$\mathbf{x}_3$、$\mathbf{x}_4$、$\mathbf{x}_5$、$\mathbf{x}_6$。

3.

$$b = y_2 - \sum_{SV\ \ indices\ \ n} \alpha_n y_n K(\mathbf{x}_n, \mathbf{x}_2) \approx -1.667$$

所以在 $\mathcal{X}$ space 中的 optimal separating nonlinear curve 爲

$$\sum_{SV\ \ indices\ \ n} \alpha_n y_n K(\mathbf{x}_n, \mathbf{x}) + b \approx 0.889x_1^2 + 0.667x_2^2 - 1.778x_1 - 1.667 = 0$$

4. 因爲在這兩題中分別使用了不同的 transformation，在不同的 $\mathcal{Z}$ space 中進行求解，所以在這兩題中所得到的 optimal separating nonlinear curve 不同。

Dual Problem of Soft-Margin Support Vector Machine with Per Example Margin Goals

5. 將 (P_1') 中的限制條件改寫爲

$$\begin{aligned} \rho_n - \xi_n - y_n(\mathbf{w}^T\mathbf{x}_n + b) &\leq 0 \\ -\xi_n &\leq 0 \end{aligned}$$

利用 Lagrange multiplier，可得

$$\begin{aligned} \mathcal{L}((b, \mathbf{w}, \boldsymbol{\xi}), (\boldsymbol{\alpha}, \boldsymbol{\beta})) = {} & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\xi_n + \\ & \sum_{n=1}^{N}\alpha_n(\rho_n - \xi_n - y_n(\mathbf{w}^T\mathbf{x}_n + b)) + \sum_{n=1}^{N}\beta_n(-\xi_n) \end{aligned}$$

6. 因爲

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = C - \alpha_n - \beta_n$$

因此，若令 $\frac{\partial \mathcal{L}}{\partial \xi_n} = C - \alpha_n - \beta_n = 0$，則可得到 $\beta_n = C - \alpha_n$，並且，由於 $\beta_n \geq 0$，因此可得 $\alpha_n \leq C$，所以原問題可以改寫如下

$$\begin{aligned} & \max_{0 \leq \alpha_n \leq C} \min_{(b, \boldsymbol{w}, \boldsymbol{\xi})} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N}(C - \alpha_n - \beta_n)\xi_n + \sum_{n=1}^{N}\alpha_n(\rho_n - y_n(\mathbf{w}^T\mathbf{x}_n + b)) \\ = {} & \max_{0 \leq \alpha_n \leq C} \min_{(b, \boldsymbol{w})} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N}\alpha_n(\rho_n - y_n(\mathbf{w}^T\mathbf{x}_n + b)) \end{aligned}$$

接著，因爲

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

因此，若令 $\frac{\partial \mathcal{L}}{\partial b} = -\sum_{n=1}^{N} \alpha_n y_n = 0$，即 $\sum_{n=1}^{N} \alpha_n y_n = 0$，則可將問題繼續改寫如下

$$
\begin{aligned}
&\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \min_{(b, \boldsymbol{w})} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(\rho_n - y_n \mathbf{w}^T \mathbf{x}_n) - \sum_{n=1}^{N} \alpha_n y_n \cdot b \\
= &\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \min_{\boldsymbol{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(\rho_n - y_n \mathbf{w}^T \mathbf{x}_n)
\end{aligned}
$$

接著，因爲

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n
$$

因此，若令 $\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = \mathbf{0}$，即 $\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$，則可將問題繼續改寫如下

$$
\begin{aligned}
&\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{w}^T \mathbf{x}_n + \sum_{n=1}^{N} \alpha_n \rho_n \\
= &\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \frac{1}{2}\mathbf{w}^T\mathbf{w} - \mathbf{w}^T \left(\sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n \right) + \sum_{n=1}^{N} \alpha_n \rho_n \\
= &\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \frac{1}{2}\mathbf{w}^T\mathbf{w} - \mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n \rho_n \\
= &\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} -\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n \rho_n \\
= &\max_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n \mathbf{x}_m + \sum_{n=1}^{N} \alpha_n \rho_n
\end{aligned}
$$

上式等同於

$$
\min_{0 \le \alpha_n \le C, \sum_{n=1}^{N} \alpha_n y_n = 0} \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n \mathbf{x}_m - \sum_{n=1}^{N} \alpha_n \rho_n
$$

因此可得 dual problem 爲

$$
\begin{aligned}
&\textit{minimize}\ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n \mathbf{x}_m - \sum_{n=1}^{N} \alpha_n \rho_n \\
&\textit{variables}\ \boldsymbol{\alpha} = (\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_N) \\
&\textit{subject to}\ 0 \le \alpha_n \le C,\ \sum_{n=1}^{N} \alpha_n y_n = 0 \\
&\textit{implicitly}\ \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n,\ \beta_n = C - \alpha_n
\end{aligned}
$$

7. 首先，將 (P_1) 和 (P_1') 改寫爲 unconstrained form。
因爲

$$\begin{aligned}&y_n(\mathbf{w}^T\mathbf{x}_n+b)\geq\rho_n-\xi_n\ and\ \xi_n\geq 0\\ \Leftrightarrow\ &\xi_n\geq\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b)\ and\ \xi_n\geq 0\\ \Leftrightarrow\ &\xi_n\geq max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)\end{aligned}$$

因此，(P_1') 可以改寫爲

$$\begin{aligned}&minimize\ \frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{n=1}^{N}\xi_n\\ &variables\ \mathbf{w},\ b,\ \boldsymbol{\xi}\\ &subject\ to\ \xi_n\geq max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)\end{aligned}$$

接著，設 $(\mathbf{w}',b',\boldsymbol{\xi}')$ 爲 (P_1') 的一個 optimal solution，其中 $\exists\ k\in\{1,2,\cdots,N\}$ 使得 $\xi_k'>max(\rho_k-y_k(\mathbf{w}^T\mathbf{x}_k+b),0)$，令 $\boldsymbol{\xi}''$ 爲

$$\xi_n''=\begin{cases}max(\rho_k-y_k(\mathbf{w}^T\mathbf{x}_k+b)<\xi_k' & if\ n=k\\ \xi_n' & if\ n\neq k\end{cases}$$

則 $(\mathbf{w}',b',\boldsymbol{\xi}'')$ 符合 (P_1') 的限制條件，並且

$$\begin{aligned}&\frac{1}{2}\mathbf{w}'^T\mathbf{w}'+C\sum_{n=1}^{N}\xi_n''\\ =&\frac{1}{2}\mathbf{w}'^T\mathbf{w}'+C\left(\xi_k''+\sum_{1\leq n\leq N,n\neq k}\xi_n''\right)\\ <&\frac{1}{2}\mathbf{w}'^T\mathbf{w}'+C\left(\xi_k'+\sum_{1\leq n\leq N,n\neq k}\xi_n'\right)\\ =&\frac{1}{2}\mathbf{w}'^T\mathbf{w}'+C\sum_{n=1}^{N}\xi_n'\end{aligned}$$

因此 $(\mathbf{w}',b',\boldsymbol{\xi}'')$ 比 $(\mathbf{w}',b',\boldsymbol{\xi}')$ 更爲 optimal，其與 $(\mathbf{w}',b',\boldsymbol{\xi}')$ 爲 (P_1') 的一個 optimal solution 矛盾，由此可知，若 $(\mathbf{w}',b',\boldsymbol{\xi}')$ 爲 (P_1') 的一個 optimal solution，則必定有 $\xi_n'=max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)$，$\forall\ n\in\{1,2,\cdots,N\}$，因此在求 (P_1') 的 optimal solution 的過程中，其實不需要 $\xi_n\geq max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)$ 如此寬鬆的限制條件，將限制條件限縮爲 $\xi_n=max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)$ 並不會影響到求解，因此可以將 (P_1') 繼續改寫爲

$$\begin{aligned}&minimize\ \frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{n=1}^{N}\xi_n\\ &variables\ \mathbf{w},\ b,\ \boldsymbol{\xi},\ where\ \xi_n=max(\rho_n-y_n(\mathbf{w}^T\mathbf{x}_n+b),0)\end{aligned}$$

其等同於以下的 unconstrained form

$$minimize\ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N} max(\rho_n - y_n(\mathbf{w}^T\mathbf{x}_n + b), 0)$$

而 (P_1) 爲 (P_1') 的特例，只要將以上敘述中的 ρ_n 皆換爲 1，即可得到 (P_1) 的 unconstrained form

$$minimize\ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N} max(1 - y_n(\mathbf{w}^T\mathbf{x}_n + b), 0)$$

接著，說明若 $(b_*', \mathbf{w}_*')$ 爲 (P_1') 在 ρ_n 皆爲 1 時的 optimal solution，則 $(2b_*', 2\mathbf{w}_*')$ 爲 (P_1) 在將 C 換爲 $2C$ 時的 optimal solution。因爲 $\forall\ (b'', \mathbf{w}'')$，皆有

$$\begin{aligned}
&\frac{1}{2}(2\mathbf{w}_*')^T(2\mathbf{w}_*') + 2C\sum_{n=1}^{N} max(1 - y_n((2\mathbf{w}_*')^T\mathbf{x}_n + 2b_*'), 0)\\
&= 4\left(\frac{1}{2}\mathbf{w}_*'^T\mathbf{w}_*' + C\sum_{n=1}^{N} max(\rho_n - y_n(\mathbf{w}_*'^T\mathbf{x}_n + b_*'), 0)\right)\\
&\leq 4\left(\frac{1}{2}(\frac{1}{2}\mathbf{w}'')^T(\frac{1}{2}\mathbf{w}'') + C\sum_{n=1}^{N} max(\rho_n - y_n((\frac{1}{2}\mathbf{w}'')^T\mathbf{x}_n + \frac{1}{2}b''), 0)\right)\\
&= \frac{1}{2}\mathbf{w}''^T\mathbf{w}'' + 2C\sum_{n=1}^{N} max(1 - y_n(\mathbf{w}''^T\mathbf{x}_n + b''), 0)
\end{aligned}$$

因此可得，若 $(b_*', \mathbf{w}_*')$ 爲 (P_1') 在 ρ_n 皆爲 1 時的 optimal solution，則 $(2b_*', 2\mathbf{w}_*')$ 爲 (P_1) 在將 C 換爲 $2C$ 時的 optimal solution。

Hard-Margin versus Soft-Margin

8. 因爲 $\boldsymbol{\alpha}^*$ 爲 hard-margin SVM 的一個 optimal solution，所以 $\boldsymbol{\alpha}^*$ 會滿足 hard-margin SVM 的限制條件

$$\alpha_n^* \geq 0,\ \sum_{n=1}^{N} \alpha_n^* y_n = 0$$

又 $C \geq \max_{1\leq n\leq N} \alpha_n^*$，因此可得

$$0 \leq \alpha_n^* \leq C,\ \sum_{n=1}^{N} \alpha_n^* y_n = 0$$

故 $\boldsymbol{\alpha}^*$ 滿足 soft-margin SVM 的限制條件。接著，設在 soft-margin SVM 中，$\boldsymbol{\alpha}'$ 比 $\boldsymbol{\alpha}^*$ 更爲 optimal，意即，$\boldsymbol{\alpha}'$ 亦滿足 soft-margin SVM 的限制條件

$$0 \leq \alpha_n' \leq C,\ \sum_{n=1}^{N} \alpha_n' y_n = 0$$

並且 $\boldsymbol{\alpha}'$ 可以使得 soft-margin SVM 的目標函數 $\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$ 有更小的值，即

$$\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n'\alpha_m' y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n' < \sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n^*\alpha_m^* y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n^*$$

因此，$\boldsymbol{\alpha}'$ 亦滿足

$$\alpha_n' \geq 0,\ \sum_{n=1}^{N}\alpha_n' y_n = 0$$

即 $\boldsymbol{\alpha}'$ 滿足 hard-margin SVM 的限制條件，並且，由於 hard-margin SVM 和 soft-margin SVM 的目標函數相同，因此

$$\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n'\alpha_m' y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n' < \sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n^*\alpha_m^* y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n^*$$

代表 $\boldsymbol{\alpha}'$ 可以使得 hard-margin SVM 的目標函數 $\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$ 有更小的值，綜合以上所述，可得在 hard-margin SVM 中，$\boldsymbol{\alpha}'$ 比 $\boldsymbol{\alpha}^*$ 更爲 optimal，其與 $\boldsymbol{\alpha}^*$ 爲 hard-margin SVM 的一個 optimal solution 矛盾，故假設錯誤，可得在 soft-margin SVM 中，不存在 $\boldsymbol{\alpha}'$ 比 $\boldsymbol{\alpha}^*$ 更爲 optimal，$\boldsymbol{\alpha}^*$ 即爲 soft-margin SVM 的一個 optimal solution。

Operation of Kernels

9. [a] 若

$$\mathbf{Q}_1 = \begin{pmatrix} K_1(\mathbf{x}_1,\mathbf{x}_1) & K_1(\mathbf{x}_1,\mathbf{x}_2) \\ K_1(\mathbf{x}_2,\mathbf{x}_1) & K_1(\mathbf{x}_2,\mathbf{x}_2) \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$$

($\mathbf{Q}_1$ 爲 symmetric，並且，因爲 $\mathbf{Q}_1$ 所有的 principal minor 的 determinant 爲

$$\begin{vmatrix} 0.9 \end{vmatrix} = 0.9 \geq 0$$
$$\begin{vmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{vmatrix} = 0.8 \geq 0$$

因此由 Sylvester's criterion 可知 $\mathbf{Q}_1$ 爲 positive semi-definite)，則有

$$\begin{aligned} \mathbf{Q} &= \begin{pmatrix} K(\mathbf{x}_1,\mathbf{x}_1) & K(\mathbf{x}_1,\mathbf{x}_2) \\ K(\mathbf{x}_2,\mathbf{x}_1) & K(\mathbf{x}_2,\mathbf{x}_2) \end{pmatrix} \\ &= \begin{pmatrix} 1-K_1(\mathbf{x}_1,\mathbf{x}_1) & 1-K_1(\mathbf{x}_1,\mathbf{x}_2) \\ 1-K_1(\mathbf{x}_2,\mathbf{x}_1) & 1-K_1(\mathbf{x}_2,\mathbf{x}_2) \end{pmatrix} \\ &= \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \end{aligned}$$

因為 $\mathbf{Q}$ 所有的 principal minor 的 determinant 為

$$\begin{vmatrix} 0.1 \end{vmatrix} = 0.1 \geq 0$$
$$\begin{vmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{vmatrix} = -0.8 < 0$$

因此由 Sylvester's criterion 可知 $\mathbf{Q}$ 不為 positive semi-definite，故由 Mercer's condition 可知 $K(\mathbf{x}, \mathbf{x}')$ 不為一個 valid kernel。

9. [b] 因為

$$K(\mathbf{x}, \mathbf{x}') = (1 - K_1(\mathbf{x}, \mathbf{x}'))^0 = 1$$

所以 $\forall$ $\mathbf{x}$、$\mathbf{x}'$，皆有

$$K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})$$

故 $K(\mathbf{x}, \mathbf{x}')$ 為 symmetric，並且，因為 $\forall$ $\mathbf{x}_1$、$\mathbf{x}_2$、$\cdots$、$\mathbf{x}_N$，皆有

$$\mathbf{Q} = \begin{pmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) & \cdots & K(\mathbf{x}_1, \mathbf{x}_N) \\ K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) & \cdots & K(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & & \vdots \\ K(\mathbf{x}_N, \mathbf{x}_1) & K(\mathbf{x}_N, \mathbf{x}_2) & \cdots & K(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

所以 $\forall$ $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \in \mathbb{R}^N$ 且 $\mathbf{u} \neq \mathbf{0}$，皆有

$$\mathbf{u}^T \mathbf{Q} \mathbf{u} = \sum_{i=1}^{N} \sum_{j=1}^{N} u_i u_j = (u_1 + u_2 + \cdots + u_N)^2 \geq 0$$

因此可得 $\mathbf{Q}$ 必定為 positive semi-definite，故由 Mercer's condition 可知 $K(\mathbf{x}, \mathbf{x}')$ 為一個 valid kernel。

Lemma

(1) 若 $K_1(\mathbf{x}, \mathbf{x}')$ 和 $K_2(\mathbf{x}, \mathbf{x}')$ 皆為 valid kernel，則 $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') K_2(\mathbf{x}, \mathbf{x}')$ 亦為 valid kernel。

(2) 若 $\forall$ $i \in \mathbb{N}$，$K_i(\mathbf{x}, \mathbf{x}')$ 皆為 valid kernel，並且 $\sum_{i=1}^{\infty} K_i(\mathbf{x}, \mathbf{x}')$ 存在，則 $K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} K_i(\mathbf{x}, \mathbf{x}')$ 亦為 valid kernel。

Proof **(1)** 因為 $K_1(\mathbf{x}, \mathbf{x}')$ 和 $K_2(\mathbf{x}, \mathbf{x}')$ 皆為 valid kernel，所以 $\exists$ $\Phi_1(\mathbf{x})$、$\Phi_2(\mathbf{x})$ 使得

$$K_1(\mathbf{x}, \mathbf{x}') = \Phi_1(\mathbf{x})^T \Phi_1(\mathbf{x}')$$
$$K_2(\mathbf{x}, \mathbf{x}') = \Phi_2(\mathbf{x})^T \Phi_2(\mathbf{x}')$$

令

$$\Phi(\mathbf{x}) = (\ \cdots \ \Phi_1^i(\mathbf{x})\Phi_2^j(\mathbf{x}) \ \cdots \ )^T$$

(其中 $\Phi_n^k(\mathbf{x})$ 爲 $\Phi_n(\mathbf{x})$ 的第 k 個 element，因爲 $\{\Phi_1^i(\mathbf{x})\}$ 和 $\{\Phi_2^j(\mathbf{x})\}$ 皆爲 countable set，所以 $\{\Phi_1^i(\mathbf{x})\Phi_2^j(\mathbf{x})\} \simeq \{\Phi_1^i(\mathbf{x})\} \times \{\Phi_2^j(\mathbf{x})\}$ 亦爲 countable set，因此才能將 $\{\Phi_1^i(\mathbf{x})\Phi_2^j(\mathbf{x})\}$ 列爲 $\Phi(\mathbf{x})$ 的各個 element)，則有

$$\begin{aligned}
& K(\mathbf{x},\mathbf{x}') \\
= \ & K_1(\mathbf{x},\mathbf{x}')K_2(\mathbf{x},\mathbf{x}') \\
= \ & \left(\Phi_1(\mathbf{x})^T\Phi_1(\mathbf{x}')\right)\left(\Phi_2(\mathbf{x})^T\Phi_2(\mathbf{x}')\right) \\
= \ & \left(\sum_i \Phi_1^i(\mathbf{x})\Phi_1^i(\mathbf{x}')\right)\left(\sum_j \Phi_2^j(\mathbf{x})\Phi_2^j(\mathbf{x}')\right) \\
= \ & \sum_{i,j} \Phi_1^i(\mathbf{x})\Phi_1^i(\mathbf{x}')\Phi_2^j(\mathbf{x})\Phi_2^j(\mathbf{x}') \\
= \ & \sum_{i,j} \left(\Phi_1^i(\mathbf{x})\Phi_2^j(\mathbf{x})\right)\left(\Phi_1^i(\mathbf{x}')\Phi_2^j(\mathbf{x}')\right) \\
= \ & \Phi(\mathbf{x})^T\Phi(\mathbf{x}')
\end{aligned}$$

因此可得 $K(\mathbf{x},\mathbf{x}')$ 爲一個 valid kernel。

(2) 因爲 $\forall\ i \in \mathbb{N}$，$K_i(\mathbf{x},\mathbf{x}')$ 皆爲 valid kernel，因此由 Mercer's condition 可知，$K_i(\mathbf{x},\mathbf{x}')$ 爲 symmetric，即 $\forall\ \mathbf{x}$、$\mathbf{x}'$，皆有

$$K_i(\mathbf{x},\mathbf{x}') = K_i(\mathbf{x}',\mathbf{x})$$

所以

$$K(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{\infty} K_i(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{\infty} K_i(\mathbf{x}',\mathbf{x}) = K(\mathbf{x}',\mathbf{x})$$

因此 $K(\mathbf{x},\mathbf{x}')$ 亦爲 symmetric，並且，因爲 $\forall\ \mathbf{x}_1$、$\mathbf{x}_2$、$\cdots$、$\mathbf{x}_N$

$$\mathbf{Q}_i = \begin{pmatrix} K_i(\mathbf{x}_1,\mathbf{x}_1) & K_i(\mathbf{x}_1,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_1,\mathbf{x}_N) \\ K_i(\mathbf{x}_2,\mathbf{x}_1) & K_i(\mathbf{x}_2,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_2,\mathbf{x}_N) \\ \vdots & \vdots & & \vdots \\ K_i(\mathbf{x}_N,\mathbf{x}_1) & K_i(\mathbf{x}_N,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix}$$

皆爲 positive semi-definite，即 $\forall\ \mathbf{u} \in \mathbb{R}^N$ 且 $\mathbf{u} \neq \mathbf{0}$，皆有 $\mathbf{u}^T\mathbf{Q}_i\mathbf{u} \geq 0$，因此，若令

$$\mathbf{Q} = \begin{pmatrix} K(\mathbf{x}_1,\mathbf{x}_1) & K(\mathbf{x}_1,\mathbf{x}_2) & \cdots & K(\mathbf{x}_1,\mathbf{x}_N) \\ K(\mathbf{x}_2,\mathbf{x}_1) & K(\mathbf{x}_2,\mathbf{x}_2) & \cdots & K(\mathbf{x}_2,\mathbf{x}_N) \\ \vdots & \vdots & & \vdots \\ K(\mathbf{x}_N,\mathbf{x}_1) & K(\mathbf{x}_N,\mathbf{x}_2) & \cdots & K(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix}$$

則有

$$\begin{aligned}
\mathbf{Q} &= \begin{pmatrix} \sum_{i=1}^{\infty} K_i(\mathbf{x}_1,\mathbf{x}_1) & \sum_{i=1}^{\infty} K_i(\mathbf{x}_1,\mathbf{x}_2) & \cdots & \sum_{i=1}^{\infty} K_i(\mathbf{x}_1,\mathbf{x}_N) \\ \sum_{i=1}^{\infty} K_i(\mathbf{x}_2,\mathbf{x}_1) & \sum_{i=1}^{\infty} K_i(\mathbf{x}_2,\mathbf{x}_2) & \cdots & \sum_{i=1}^{\infty} K_i(\mathbf{x}_2,\mathbf{x}_N) \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{\infty} K_i(\mathbf{x}_N,\mathbf{x}_1) & \sum_{i=1}^{\infty} K_i(\mathbf{x}_N,\mathbf{x}_2) & \cdots & \sum_{i=1}^{\infty} K_i(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix} \\
&= \sum_{i=1}^{\infty} \begin{pmatrix} K_i(\mathbf{x}_1,\mathbf{x}_1) & K_i(\mathbf{x}_1,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_1,\mathbf{x}_N) \\ K_i(\mathbf{x}_2,\mathbf{x}_1) & K_i(\mathbf{x}_2,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_2,\mathbf{x}_N) \\ \vdots & \vdots & & \vdots \\ K_i(\mathbf{x}_N,\mathbf{x}_1) & K_i(\mathbf{x}_N,\mathbf{x}_2) & \cdots & K_i(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix} \\
&= \sum_{i=1}^{\infty} \mathbf{Q}_i
\end{aligned}$$

所以 $\forall\ \mathbf{u} \in \mathbb{R}^N$ 且 $\mathbf{u} \neq \mathbf{0}$，皆有

$$\mathbf{u}^T\mathbf{Q}\mathbf{u} = \mathbf{u}^T\left(\sum_{i=1}^{\infty}\mathbf{Q}_i\right)\mathbf{u} = \sum_{i=1}^{\infty}\mathbf{u}^T\mathbf{Q}_i\mathbf{u} \geq 0$$

故 $\mathbf{Q}$ 爲 positive semi-definite，因此由 Mercer's condition 可知，$K(\mathbf{x},\mathbf{x}')$ 爲一個 valid kernel。

9. [**c**] 因爲 $0 < K_1(\mathbf{x},\mathbf{x}') < 1$，所以

$$K(\mathbf{x},\mathbf{x}') = (1 - K_1(\mathbf{x},\mathbf{x}'))^{-1} = \sum_{i=0}^{\infty} K_1(\mathbf{x},\mathbf{x}')^i$$

其中，由 **9.** [**b**] 和 **Lemma (1)** 可知，$\forall\ i \in \mathbb{N} \cup \{0\}$，$K(\mathbf{x},\mathbf{x}')^i$ 皆爲 valid kernel，因此由 **Lemma (2)** 可知，$K(\mathbf{x},\mathbf{x}') = \sum_{i=0}^{\infty} K_1(\mathbf{x},\mathbf{x}')^i$ 爲一個 valid kernel。

9. [**d**] 因爲

$$K(\mathbf{x},\mathbf{x}') = (1 - K_1(\mathbf{x},\mathbf{x}'))^{-2} = (1 - K_1(\mathbf{x},\mathbf{x}'))^{-1} \cdot (1 - K_1(\mathbf{x},\mathbf{x}'))^{-1}$$

其中，由 **9.** [**c**] 可知，$(1 - K_1(\mathbf{x},\mathbf{x}'))^{-1}$ 爲一個 valid kernel，因此由 **Lemma (1)** 可知，$K(\mathbf{x},\mathbf{x}') = (1 - K_1(\mathbf{x},\mathbf{x}'))^{-1} \cdot (1 - K_1(\mathbf{x},\mathbf{x}'))^{-1}$ 爲一個 valid kernel。

10. 令 $\boldsymbol{\alpha}^*$ 爲以下 (S_1) 的 optimal solution

$$\begin{aligned}
&\textit{minimize}\ \sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m) - \sum_{n=1}^{N}\alpha_n \\
&\textit{subject to}\ 0 \leq \alpha_n \leq C,\ \sum_{n=1}^{N}\alpha_n y_n = 0
\end{aligned}$$

首先，説明 $\frac{\boldsymbol{\alpha}^*}{p}$ 爲以下 (S_2) 的 optimal solution

$$
\begin{aligned}
&minimize\ \sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha_n \\
&subject\ to\ 0\le\alpha_n\le\tilde{C},\ \sum_{n=1}^{N}\alpha_n y_n=0
\end{aligned}
$$

因爲 $\boldsymbol{\alpha}^*$ 爲 (S_1) 的 optimal solution，所以 $\boldsymbol{\alpha}^*$ 會滿足 (S_1) 的限制條件

$$
0\le\alpha_n^*\le C,\ \sum_{n=1}^{N}\alpha_n^* y_n=0
$$

因此可得

$$
0\le\frac{\alpha_n^*}{p}\le\frac{C}{p}=\tilde{C},\ \sum_{n=1}^{N}\frac{\alpha_n^*}{p}y_n=\frac{1}{p}\sum_{n=1}^{N}\alpha_n^* y_n=0
$$

故 $\frac{\boldsymbol{\alpha}^*}{p}$ 會滿足 (S_2) 的限制條件。接著，設在 (S_2) 中，$\boldsymbol{\alpha}'$ 比 $\frac{\boldsymbol{\alpha}^*}{p}$ 更爲 optimal，亦即，$\boldsymbol{\alpha}'$ 亦符合 (S_2) 的限制條件

$$
0\le\alpha_n'\le\tilde{C},\ \sum_{n=1}^{N}\alpha_n' y_n=0
$$

並且 $\boldsymbol{\alpha}'$ 可以使得 (S_2) 的目標函數 $\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m\tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha_n$ 有更小的值，即

$$
\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n'\alpha_m' y_n y_m\tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha_n' < \sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\alpha_n^*}{p}\frac{\alpha_m^*}{p}y_n y_m\tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\frac{\alpha_n^*}{p}
$$

因此，$p\boldsymbol{\alpha}'$ 會滿足

$$
0\le p\alpha_n'\le p\tilde{C}=C,\ \sum_{n=1}^{N}(p\alpha_n')y_n=p\sum_{n=1}^{N}\alpha_n' y_n=0
$$

即 $p\boldsymbol{\alpha}'$ 會滿足 (S_1) 的限制條件，並且

$$
\begin{aligned}
&\sum_{n=1}^{N}\sum_{m=1}^{N}(p\alpha'_n)(p\alpha'_m)y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}p\alpha'_n\\
&=p\left(\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha'_n\alpha'_m y_n y_m pK(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha'_n\right)\\
&=p\left(\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha'_n\alpha'_m y_n y_m \tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha'_n\right)\\
&<p\left(\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\alpha^*_n}{p}\frac{\alpha^*_m}{p}y_n y_m \tilde{K}(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\frac{\alpha^*_n}{p}\right)\\
&=p\left(\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\alpha^*_n}{p}\frac{\alpha^*_m}{p}y_n y_m pK(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\frac{\alpha^*_n}{p}\right)\\
&=\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha^*_n\alpha^*_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha^*_n
\end{aligned}
$$

代表 $p\boldsymbol{\alpha}'$ 可以讓 (S_1) 的目標函數 $\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)-\sum_{n=1}^{N}\alpha_n$ 有更小的值，綜合以上所述，可得在 (S_1) 中，$p\boldsymbol{\alpha}'$ 比 $\boldsymbol{\alpha}^*$ 更爲 optimal，其與 $\boldsymbol{\alpha}^*$ 爲 (S_1) 的 optimal solution 矛盾，故假設錯誤，可得在 (S_2) 中，不存在 $\boldsymbol{\alpha}'$ 比 $\frac{\boldsymbol{\alpha}^*}{p}$ 更爲 optimal，$\frac{\boldsymbol{\alpha}^*}{p}$ 即爲 (S_2) 的一個 optimal solution。接著，由

$$
0\le\alpha^*_n\le C\Leftrightarrow 0\le\frac{\alpha^*_n}{p}\le\frac{C}{p}=\tilde{C}
$$

可知 (S_1) 和 (S_2) 有相同的 support vector (以及 free support vector)，因此，若 (S_1) 所得到的 classifier 爲

$$
g_{SVM}(\mathbf{x})=sign\left(\sum_{SV\ \ indices\ \ n}\alpha^*_n y_n K(\mathbf{x}_n,\mathbf{x})+b\right)
$$

(S_2) 所得到的 classifier 爲

$$
\tilde{g}_{SVM}(\mathbf{x})=sign\left(\sum_{SV\ \ indices\ \ n}\frac{\alpha^*_n}{p}y_n\tilde{K}(\mathbf{x}_n,\mathbf{x})+\tilde{b}\right)
$$

則有

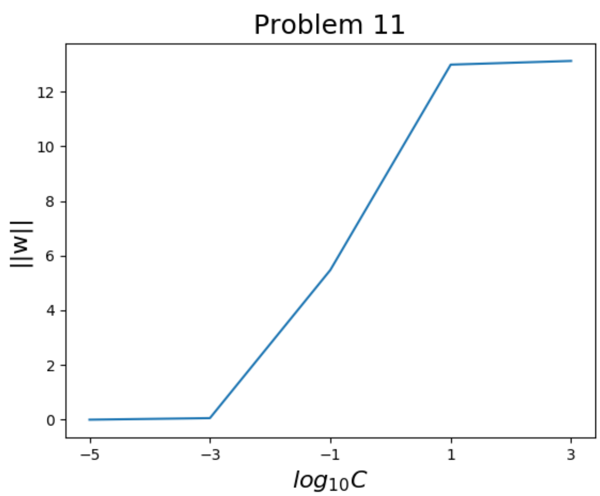
$$
\begin{aligned}
b&=y_s-\sum_{SV\ \ indices\ \ n}\alpha^*_n y_n K(\mathbf{x}_n,\mathbf{x}_s)\\
&=y_s-\sum_{SV\ \ indices\ \ n}\frac{\alpha^*_n}{p}y_n pK(\mathbf{x}_n,\mathbf{x}_s)\\
&=y_s-\sum_{SV\ \ indices\ \ n}\frac{\alpha^*_n}{p}y_n\tilde{K}(\mathbf{x}_n,\mathbf{x}_s)=\tilde{b}
\end{aligned}
$$

(其中 s 爲任意一個 free support vector 的 index)

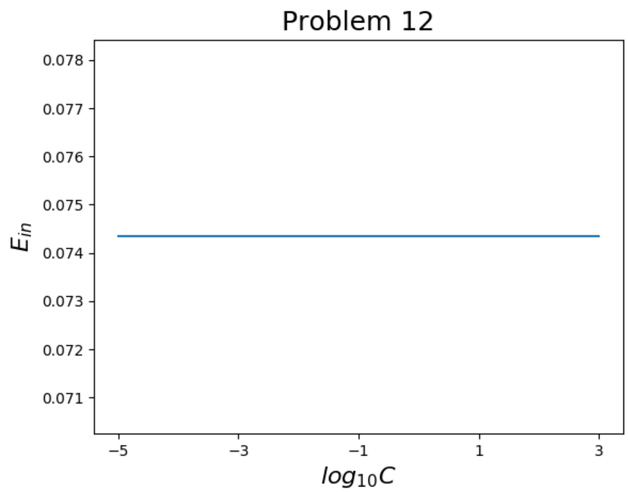
$$
\begin{aligned}
g_{SVM}(\mathbf{x}) &= sign\left(\sum_{SV\ \ indices\ \ n} \alpha_n^* y_n K(\mathbf{x}_n, \mathbf{x}) + b\right) \\
&= sign\left(\sum_{SV\ \ indices\ \ n} \frac{\alpha_n^*}{p} y_n p K(\mathbf{x}_n, \mathbf{x}) + \tilde{b}\right) \\
&= sign\left(\sum_{SV\ \ indices\ \ n} \frac{\alpha_n^*}{p} y_n \tilde{K}(\mathbf{x}_n, \mathbf{x}) + \tilde{b}\right) = \tilde{g}_{SVM}(\mathbf{x})
\end{aligned}
$$

Experiments with Soft-Margin Support Vector Machine

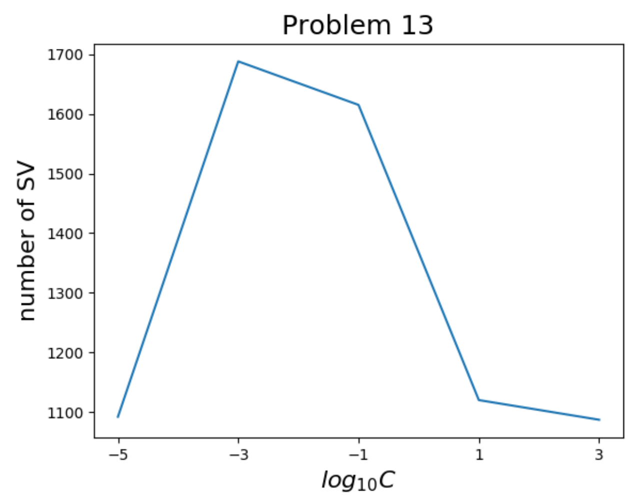
11. 由下圖可知，當 $log_{10}C$ 越大時，$\|w\|$ 也越大。



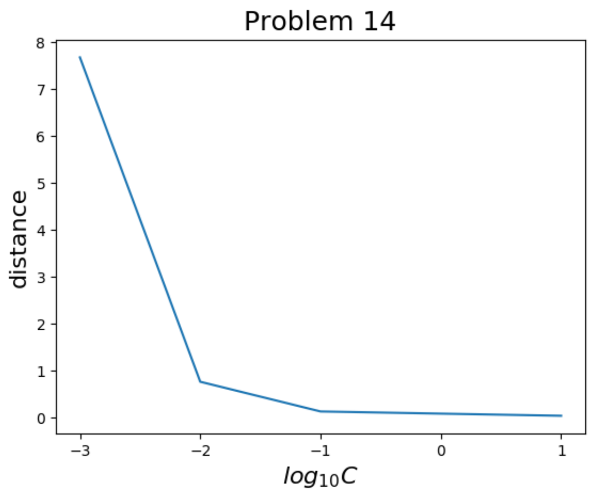
12. 由下圖可知，不論 $log_{10}C$ 爲何，E_{in} 皆相同。



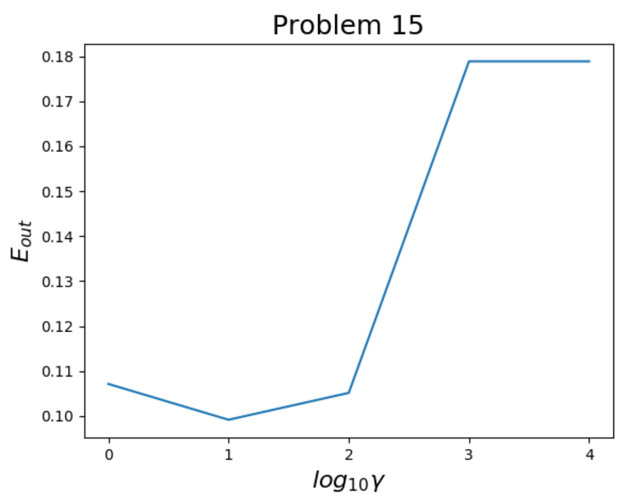
13. 由下圖可知，當 $log_{10}C$ 從 -5 增加到 -3 時，support vector 的數量上升到最大值，而當 $log_{10}C$ 從 -3 遞增到 3 時，support vector 的數量則遞減。當 $log_{10}C$ 在 -3 到 -1 之間時，support vector 的數量較多。



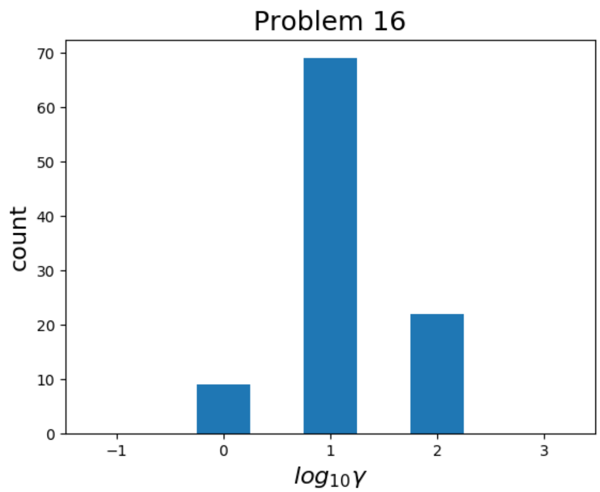
14. 由下圖可知，當 $log_{10}C$ 越大時，$\mathcal{Z}$ space 中 free support vector 到 optimal separating hyperplane 的距離越小。



15. 由下圖可知，當 $log_{10}\gamma$ 從 0 增加到 1 時，E_{out} 會下降到最小值，而當 $log_{10}\gamma$ 從 1 遞增到 4 時，E_{out} 則會遞增。



16. 當 $log_{10}\gamma$ 爲 1 時，被選中的次數最多，而當 $log_{10}\gamma$ 爲 -1 或 3 時，則皆沒被選中。



Bonus: Constant Feature for Support Vector Machine

17. 不論是 hard-margin SVM 還是 soft-margin SVM，optimal solution 的形式皆為

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n$$

因此，考慮上式的第 i 個 component，可得

$$w_i = \sum_{n=1}^{N} \alpha_n y_n z_i = \left(\sum_{n=1}^{N} \alpha_n y_n \right) z_i$$

而不論是 hard-margin SVM 還是 soft-margin SVM，optimal solution 皆必須滿足限制條件

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

因此可得

$$w_i = 0$$

Bonus: Dual of Dual

18. 令

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix},\ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$$\mathbf{Q} = \begin{pmatrix} y_1 y_1 \mathbf{z}_1^T \mathbf{z}_1 & y_1 y_2 \mathbf{z}_1^T \mathbf{z}_2 & \cdots & y_1 y_N \mathbf{z}_1^T \mathbf{z}_N \\ y_2 y_1 \mathbf{z}_2^T \mathbf{z}_1 & y_2 y_2 \mathbf{z}_2^T \mathbf{z}_2 & \cdots & y_2 y_N \mathbf{z}_2^T \mathbf{z}_N \\ \vdots & \vdots & & \vdots \\ y_N y_1 \mathbf{z}_N^T \mathbf{z}_1 & y_N y_2 \mathbf{z}_N^T \mathbf{z}_2 & \cdots & y_N y_N \mathbf{z}_N^T \mathbf{z}_N \end{pmatrix}$$

則 hard-margin dual SVM 可以寫爲

$$\begin{gathered} \textit{minimize}\ \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{1}_N^T \boldsymbol{\alpha} \\ \textit{subject to}\ \alpha_n \geq 0,\ \mathbf{y}^T \boldsymbol{\alpha} = 0 \end{gathered}$$

接著，利用 Lagrange multiplier 將以上問題改寫爲

$$\begin{aligned} & \min_{\alpha_n \geq 0} \max_{\lambda_n \geq 0, \mu} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mu) \\ = & \min_{\alpha_n \geq 0} \max_{\lambda_n \geq 0, \mu} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{1}_N^T \boldsymbol{\alpha} + \sum_{n=1}^{N} \lambda_n(-\alpha_n) + \mu \mathbf{y}^T \boldsymbol{\alpha} \\ = & \min_{\alpha_n \geq 0} \max_{\lambda_n \geq 0, \mu} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{1}_N^T \boldsymbol{\alpha} - \boldsymbol{\lambda}^T \boldsymbol{\alpha} + \mu \mathbf{y}^T \boldsymbol{\alpha} \\ = & \min_{\alpha_n \geq 0} \max_{\lambda_n \geq 0, \mu} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N)^T \boldsymbol{\alpha} \end{aligned}$$

設以上問題爲 feasible，由於該問題爲 convex，且限制條件皆爲 linear，所以可以利用 strong duality，將以上問題改寫爲

$$\max_{\lambda_n \geq 0, \mu} \min_{\alpha_n \geq 0} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N)^T \boldsymbol{\alpha}$$

接著，利用 KKT condition，令

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}} = \mathbf{Q}\boldsymbol{\alpha} - (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N) = \mathbf{0}$$

設 $\mathbf{Q}$ 爲 invertible，則有

$$\begin{gathered} \mathbf{Q}\boldsymbol{\alpha} = \boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N \\ \boldsymbol{\alpha} = \mathbf{Q}^{-1}(\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N) \end{gathered}$$

因此，以上問題可以繼續改寫爲

$$\begin{aligned} \max_{\lambda_n \geq 0, \mu} \frac{1}{2} & (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N)^T (\mathbf{Q}^{-1})^T (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N) - \\ & (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N)^T \mathbf{Q}^{-1} (\boldsymbol{\lambda} - \mu \mathbf{y} + \mathbf{1}_N) \end{aligned}$$

由於 $\mathbf{Q}$ 爲 symmetric，因此 $\mathbf{Q}^{-1}$ 亦爲 symmetric，故將 $(\mathbf{Q}^{-1})^T = \mathbf{Q}^{-1}$ 代入以上式子化簡後，可以得到 hard-margin dual SVM 的 dual problem 爲

$$\max_{\lambda_n \geq 0, \mu} -\frac{1}{2}(\boldsymbol{\lambda} - \mu\mathbf{y} + \mathbf{1}_N)^T \mathbf{Q}^{-1}(\boldsymbol{\lambda} - \mu\mathbf{y} + \mathbf{1}_N)$$

其等同於

$$\min_{\lambda_n \geq 0, \mu} \frac{1}{2}(\boldsymbol{\lambda} - \mu\mathbf{y} + \mathbf{1}_N)^T \mathbf{Q}^{-1}(\boldsymbol{\lambda} - \mu\mathbf{y} + \mathbf{1}_N)$$

其與 hard-margin SVM 的 primal problem 不同，但兩者皆是在限制條件下 minimize 某個 vector norm，爲兩者在形式上的相似之處。