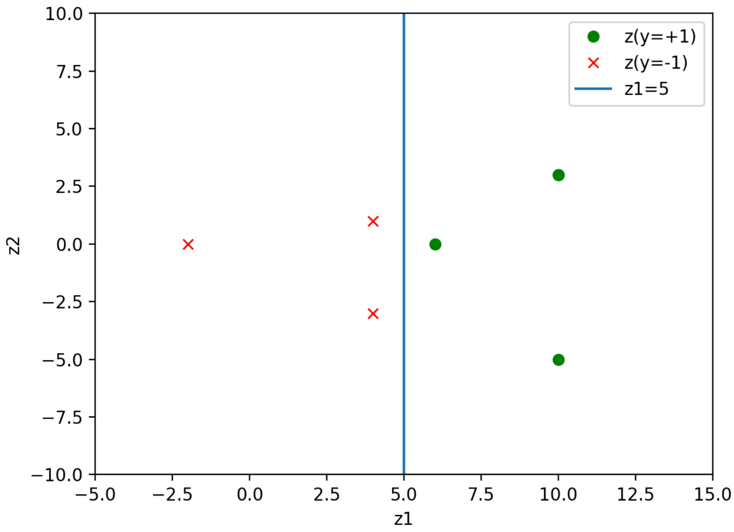
Machine Learning Techniques HW1

B04902004 王佑安

1.



2. $\alpha = [0, 0.21970.28020.33320.06810.0984, 0]$
Support vectors = x2, x3, x4, x5, x6

3. $\sum_{\text{SV indices n}} \alpha_n y_n K(x_n, x) + b = 0$
$\Rightarrow \sum_{\text{SV indices n}} \alpha_n y_n (1 + 2x_n^T x)^2 + (y_s - \sum_{\text{SV indices n}} \alpha_n y_n (1 + 2x_n^T x_s)) = 0$
$\Rightarrow 1.33x_1^2 + 0.67x_2^2 - 1.33x_1 - 1.67 = 0$

4.
Q1:
$\phi_1(x) = 5$
$\Rightarrow 2x_2^2 + 4x_1 - 3 = 0$
Q3:
$1.33x_1^2 + 0.67x_2^2 - 1.33x_1 - 1.67 = 0$
Two curves are diffrent in X space.

5. $\mathcal{L}((b, \mathbf{w}, \xi), (\alpha, \beta)) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N} u_n \xi_n + \sum_{n=1}^{N} \alpha_n (p_n - \xi_n - y_n(\mathbf{w}^T\mathbf{x}_n + b)) + \sum_{n=1}^{N} \beta_n(-\xi_n)$

6. For optimal $(\xi, \mathbf{w}, b)$

   $\frac{\delta\mathcal{L}}{\delta\xi} = \sum_{n=1}^{N} Cu_n - \alpha_n - \beta_n = 0$

   $\frac{\delta\mathcal{L}}{\delta\mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n x_n = 0$

   $\frac{\delta\mathcal{L}}{\delta b} = \sum_{n=1}^{N} -\alpha_n y_n = 0$

   $\Rightarrow \mathcal{L}((b, \mathbf{w}, \xi), (\alpha, \beta)) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(p_n - y_n(\mathbf{w}^T\mathbf{x}_n)) + \sum_{n=1}^{N} -\alpha_n y_n b + \sum_{n=1}^{N} (Cu_n - \alpha_n - \beta_n)\xi_n$

   $\Rightarrow \mathcal{L}((b, \mathbf{w}, \xi), (\alpha, \beta)) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(p_n - y_n(\mathbf{w}^T\mathbf{x}_n))$

   $\Rightarrow \mathcal{L}((b, \mathbf{w}, \xi), (\alpha, \beta)) = -\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n p_n$

   $P_1' = \max_{Cu_n \geq \alpha_n \geq 0, \mathbf{w}=\sum_{n}^{N} \alpha_n y_n x_n, \sum_{n=1}^{N} \alpha_n y_n = 0, \beta = Cu_n - \alpha_n} \min_{(b,\mathbf{w},\xi)} \mathcal{L}((b, \mathbf{w}, \xi), (\alpha)) = -\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n p_n$

   $= \max_{Cu_n \geq \alpha_n \geq 0, \mathbf{w}=\sum_{n}^{N} \alpha_n y_n x_n, \sum_{n=1}^{N} \alpha_n y_n = 0, \beta = Cu_n - \alpha_n} -\frac{1}{2}\| \sum_{n=1}^{N} \alpha_n y_n x_n \|^2 + \sum_{n=1}^{N} \alpha_n p_n$

   Dual problem:

   $\min_{\alpha} \frac{1}{2} \sum_{n_=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m x_n^T x_m - \sum_{n=1}^{N} \alpha_n p_n$

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

7. (P_1') $\min_{\mathbf{w}_*', b_*', \xi} \frac{1}{2}\mathbf{w}_*'^T\mathbf{w}_*' + C\sum_{n=1}^{N} \xi_n$

   subject to $y_n(\mathbf{w}_*'^T\mathbf{x}_n + b_*') \geq 0.25 - \xi_n, \xi_n \geq 0$

   $\Leftrightarrow \min_{\mathbf{w}_*', b_*', \xi} \frac{1}{16}\frac{1}{2}4\mathbf{w}_*'^T 4\mathbf{w}_*' + \frac{1}{4}C\sum_{n=1}^{N} 4\xi_n$

   subject to $y_n(4\mathbf{w}_*'^T\mathbf{x}_n + 4b_*') \geq 1 - 4\xi_n, 4\xi_n \geq 0$

   $\Leftrightarrow \min_{\mathbf{w}_*', b_*', \xi} \frac{1}{2}4\mathbf{w}_*'^T 4\mathbf{w}_*' + C\sum_{n=1}^{N} 4\xi_n$

   subject to $y_n(4\mathbf{w}_*'^T\mathbf{x}_n + 4b_*') \geq 1 - 4\xi_n, 4\xi_n \geq 0$

   $\Rightarrow$ For (P_1), when $(\mathbf{w}, b) = (4\mathbf{w}_*', 4b*')$ is the optimal solution.

8. Dual problem of hard margin SVM $\alpha^* =$

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

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

   Dual problem of soft margin SVM $\alpha^* =$

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

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

   when $C \geq max_{\alpha^*}$, two problems are equivalent, so they have same optimal solution α^*.

9. (a) For $K_1(x,x') = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = K_1(x,x')^T K_1(x,x')$ is positive semi-definite

   $(2 - K_1(x,x'))^2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, eigenvalues = [-1, 3], is not positive semi-definite $\Rightarrow$ invalid kernel.

   (b) $(2 - K_1(x,x'))^0 = I = I^T I$ is positive semi-definite $\Rightarrow$ valid kernel.

   (c) A is positive semi-definite $\Rightarrow \forall x \in \mathbb{R}^n, x^T A x \geq 0$

   $\forall$ positive semi-definite A, B, $x^T(A+B)x = x^T A x + x^T B x \Rightarrow$ positive semi-definite is closed under addition.

   $A^n = (X^T)^n X^n$ is positive semi-definite $\Rightarrow$ positive semi-definite is closed under power operation.

   $(2 - K_1(x,x'))^{-1} = \sum\limits_{n=0}^{\infty} \frac{K_1(x,x')^n}{2^{n+1}}$ is positive semi-definite $\Rightarrow$ valid kernel.

   (d) $(2 - K_1(x,x'))^{-2} = ((2 - K_1(x,x'))^{-1})^2$

   From (c) we know positive semi-definite is closed under power operation and $(2 - K_1(x,x'))^{-1}$ is positive semi-definite

   $\Rightarrow (2 - K_1(x,x'))^{-2}$ is positive semi-definite $\Rightarrow$ valid kernel.

10. Optimal solution of $\tilde{K}(x,x') =$

    $\min\limits_{\tilde{\alpha}} \frac{1}{2} \sum\limits_{n=1}^{N} \sum\limits_{m=1}^{N} \tilde{\alpha}_n \tilde{\alpha}_m y_n y_m p K(x_n^T, x_m) - \sum\limits_{n=1}^{N} \tilde{\alpha}_n$

    subject to $\sum\limits_{n=1}^{N} y_n \tilde{\alpha}_n = 0, C \geq \tilde{\alpha}_n \geq 0$

    $\min\limits_{\tilde{\alpha}} \frac{1}{2} \sum\limits_{n=1}^{N} \sum\limits_{m=1}^{N} \tilde{\alpha}_n \tilde{\alpha}_m y_n y_m p K(x_n^T, x_m) - \sum\limits_{n=1}^{N} \tilde{\alpha}_n$

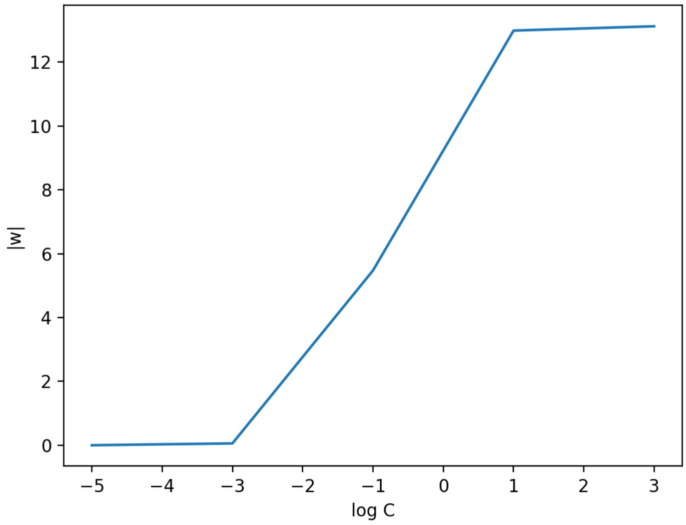
    $= p \min\limits_{\tilde{\alpha}} \frac{1}{2} \sum\limits_{n=1}^{N} \sum\limits_{m=1}^{N} \tilde{\alpha}_n \tilde{\alpha}_m y_n y_m p K(x_n^T, x_m) - p \sum\limits_{n=1}^{N} \tilde{\alpha}_n$

    $= \min\limits_{\tilde{\alpha}} \frac{1}{2} \sum\limits_{n=1}^{N} \sum\limits_{m=1}^{N} (p\tilde{\alpha}_n)(p\tilde{\alpha}_m) y_n y_m K(x_n^T, x_m) - \sum\limits_{n=1}^{N} (p\tilde{\alpha}_n)$

    $\Rightarrow$ Optimal solution of $K(x,x') : \alpha = (p\tilde{\alpha}) \Rightarrow \tilde{\alpha} = \frac{\alpha}{p}$

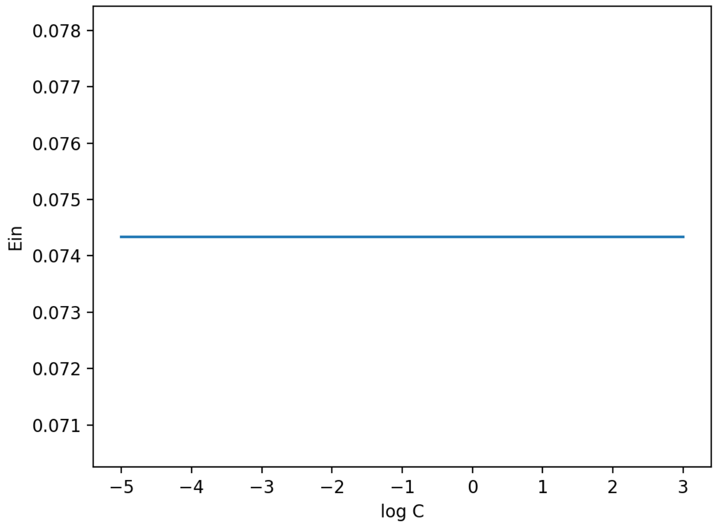
    $\Rightarrow$ Upper bound of $\tilde{\alpha} : \tilde{C} = \frac{C}{p}$

11.



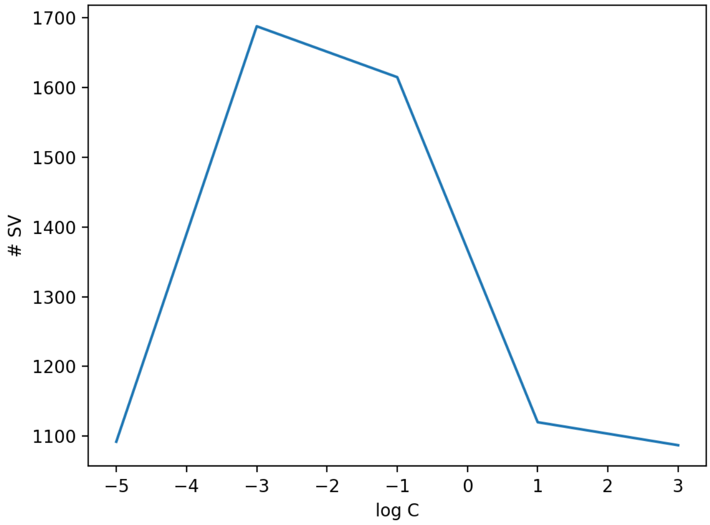
C 越大 |w| 就越大

12.



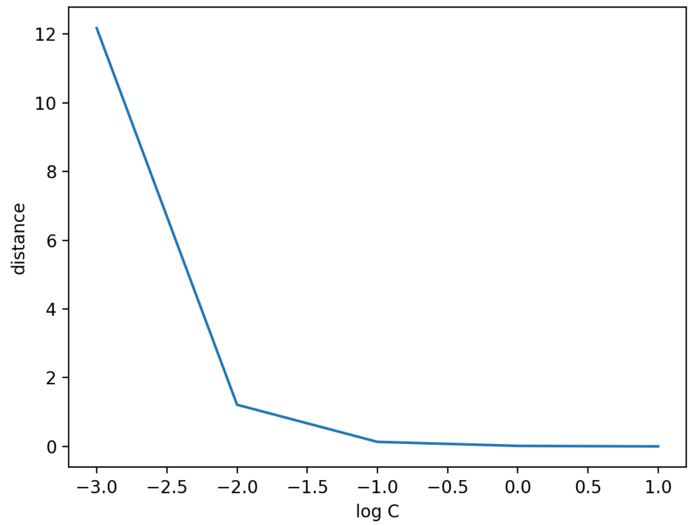
不管 C 多大在這個 task 得到的結果都一樣

13.



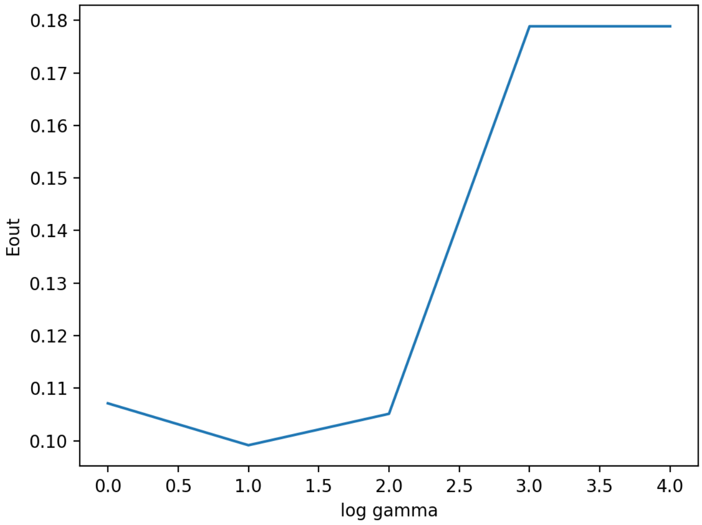
SV 的數量在 $log_{10}C = -1$~-3 時較大，其他差不多

14.



C 越小，SV 到 hyperplane 的距離越大

15.



$log_{10}C = 1$ 時 error 最小，≥ 3 時明顯變大

16.