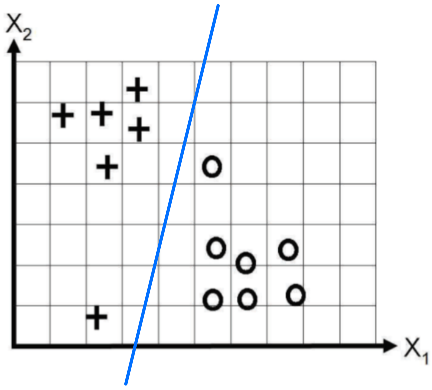
1 Written Problems

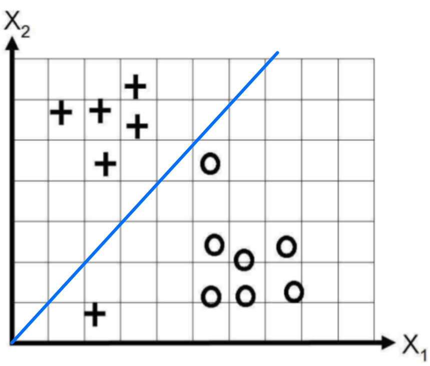
1、(1)



The answer is not unique.

There is no classification error made on the dataset.

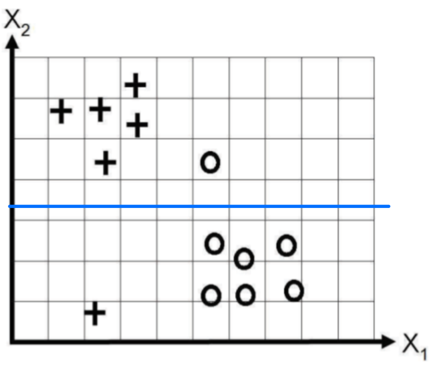
(2)



Since w_0 would be regularized to 0, so the boundary goes through the origin.

1 classification error has been made on the training set.

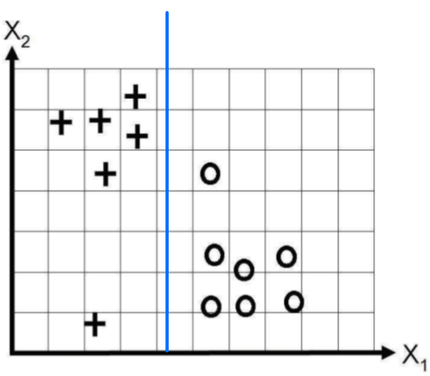
(3)



w_1 will be regularized to 0, the boundary will be horizontal.

2 classification error has been made on the training set.

(4)



w_2 will be regularized to 0, the boundary will be vertical.

0 classification error has been made on the training set.

2. (1) $\varphi(x_1) = [1, 0, 0]^T$ $\varphi(x_2) = [1, 2, 2]^T$

$\varphi(x_2) - \varphi(x_1) = [0, 2, 2]^T$

Since w is orthogonal to the decision boundary, and $\varphi(x_1)$, $\varphi(x_2)$ have decided the decision boundary, so a possible vector parallel to w can be $[0, 2, 2]^T$

(2) $d_{12} = \| \varphi(x_2) - \varphi(x_1) \|_2 = \sqrt{0+4+4} = 2\sqrt{2}$

$r = \frac{1}{2} d_{12} = \sqrt{2}$

Thus, the margin should be $\sqrt{2}$.

(3) $\frac{1}{\|w\|} = \sqrt{2} \Rightarrow \|w\| = \frac{\sqrt{2}}{2}$

From (1), we can set w to be $[0, 2a, 2a]^T$

$\|w\|_2 = \sqrt{0+4a^2+4a^2} = 2\sqrt{2}|a| = \frac{\sqrt{2}}{2}$

$\Rightarrow |a| = \frac{1}{4} \Rightarrow a = \frac{1}{4}$ or $a = -\frac{1}{4}$

Since $y_i(w^T\varphi(x_i) + w_0) \geq 1$

$$\begin{cases} -1 \cdot w_0 \geq 1 \\ 1 \cdot (4a+4a+w_0) \geq 1 \end{cases} \Rightarrow \begin{cases} -w_0 \geq 1 \\ 8a+w_0 \geq 1 \end{cases} \Rightarrow 8a \geq 2 \Rightarrow a \geq \frac{1}{4}$$

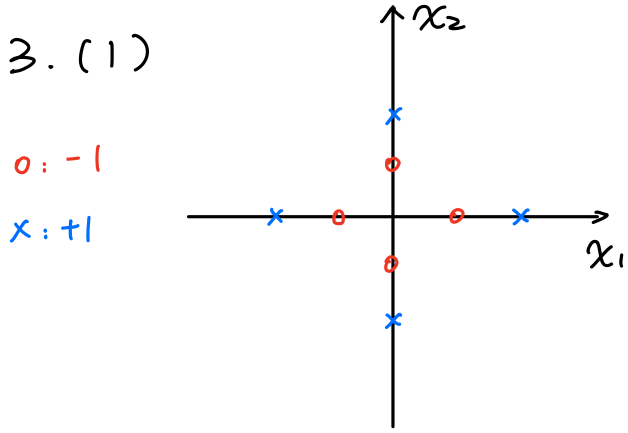
Thus, $a = \frac{1}{4}$, so $w = [0, \frac{1}{2}, \frac{1}{2}]^T$

(4) $$\begin{cases} -1 \cdot w_0 \geq 1 \\ 1 \cdot (2+w_0) \geq 1 \end{cases} \Rightarrow \begin{cases} w_0 \leq -1 \\ w_0 \geq -1 \end{cases} \Rightarrow w_0 = -1$$

(5) $f(x) = w_0 + w^T\varphi(x) = -1 + [0 \ \frac{1}{2} \ \frac{1}{2}] \begin{bmatrix} 1 \\ \sqrt{2}x \\ x^2 \end{bmatrix}$

$= -1 + \frac{\sqrt{2}}{2}x + \frac{1}{2}x^2$

3. (1)

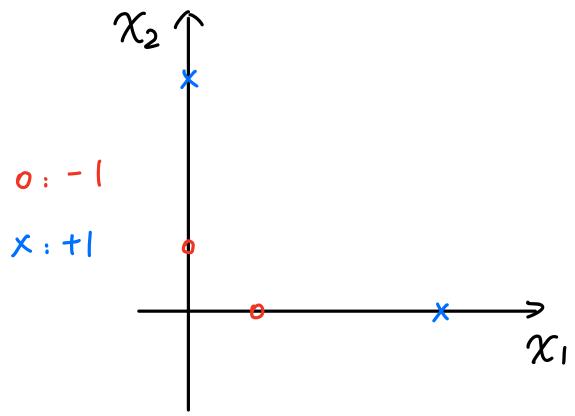


We cannot find a svm classifier without slack variable for this dataset, since the data is not linearly separable in the original dimension space.

(2)

Class -1 : $\begin{bmatrix} (1 & 0) \\ (0 & 1) \end{bmatrix}$ Class +1: $\begin{bmatrix} (4 & 0) \\ (0 & 4) \end{bmatrix}$

Then we draw the plot below:



After the transformation by the kernel function, the data becomes linearly separable.

Then, we fit the SVM classifier, let $w = [w_1 \ w_2]^T$

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad -1 \cdot (w^T \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + b) \geq 1 \qquad \alpha_1$$

$$-1 \cdot (w^T \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b) \geq 1 \qquad \alpha_2$$

$$1 \cdot (w^T \cdot \begin{bmatrix} 0 \\ 4 \end{bmatrix} + b) \geq 1 \qquad \alpha_3$$

$$1 \cdot (w^T \cdot \begin{bmatrix} 4 \\ 0 \end{bmatrix} + b) \geq 1 \qquad \alpha_4$$

$$L = \frac{1}{2}(w_1^2 + w_2^2) + \alpha_1(1 + w_2 + b) + \alpha_2(1 + w_1 + b) + \alpha_3(1 - 4w_2 - b) + \alpha_4(1 - 4w_1 - b)$$

Stationarity:

$$\frac{\partial L}{\partial w_1} = w_1 + \alpha_2 - 4\alpha_4 = 0 \quad , \quad \frac{\partial L}{\partial w_2} = w_2 + \alpha_1 - 4\alpha_3 = 0$$

$$\frac{\partial L}{\partial b} = \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$$

Feasibility: $\alpha_i \geq 0$, $1 - y_i(w^T x_i + b) \leq 0$, $i = 1, 2, 3, 4$.

Complementarity slackness:

$\alpha_1(1 + w_2 + b) = 0$, $\alpha_2(1 + w_1 + b) = 0$, $\alpha_3(1 - 4w_2 - b) = 0$, $\alpha_4(1 - 4w_1 - b) = 0$

After the calculation, $w = [\frac{2}{3} \ \frac{2}{3}]^T$, $b = -\frac{5}{3}$

The decision boundary should be $\frac{2}{3}x_1 + \frac{2}{3}x_2 - \frac{5}{3} = 0 \Rightarrow 2x_1 + 2x_2 - 5 = 0$

Let $x_1 = [1 \ 2]^T$, $\varphi(x_1) = [1 \ 4]^T$

$w^T \varphi(x_1) + b = \frac{2}{3} + \frac{8}{3} - \frac{5}{3} = \frac{5}{3} > 0$, so the label of $[1 \ 2]^T$ should be class $+1$.

4. The dual problem of the optimization problem in the question is.

$$\min_{w,b} \ \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \ 1 - y_i(w^T x_i + b) \leq 0 \quad \forall i$$

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 + \sum_i^m \alpha_i(1 - y_i(w^T x_i + b))$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_i^m \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_i^m \alpha_i y_i = 0$$

By strong duality theorem,

$$\sum_{n=1}^N \alpha_n - \frac{1}{2}\sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m x_n^T x_m = \frac{1}{2}\|w\|^2$$

Since $\sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m x_n^T x_m = (\sum_{i=1}^N \alpha_i y_i x_i)^T (\sum_{j=1}^N \alpha_j y_j x_j) = w^T w = \|w\|^2$,

so the above equation can be rewritten as:

$$\sum_{n=1}^{N} a_n - \frac{1}{2}\|w\|^2 = \frac{1}{2}\|w\|^2$$

$$\|w\|^2 = \sum_{n=1}^{N} a_n$$

Since γ is the margin, then $\gamma = \frac{1}{\|w\|}$

$$\Rightarrow \frac{1}{\gamma^2} = \|w\|^2 = \sum_{n=1}^{N} a_n \qquad \text{Q.E.D.}$$