1、Written Questions

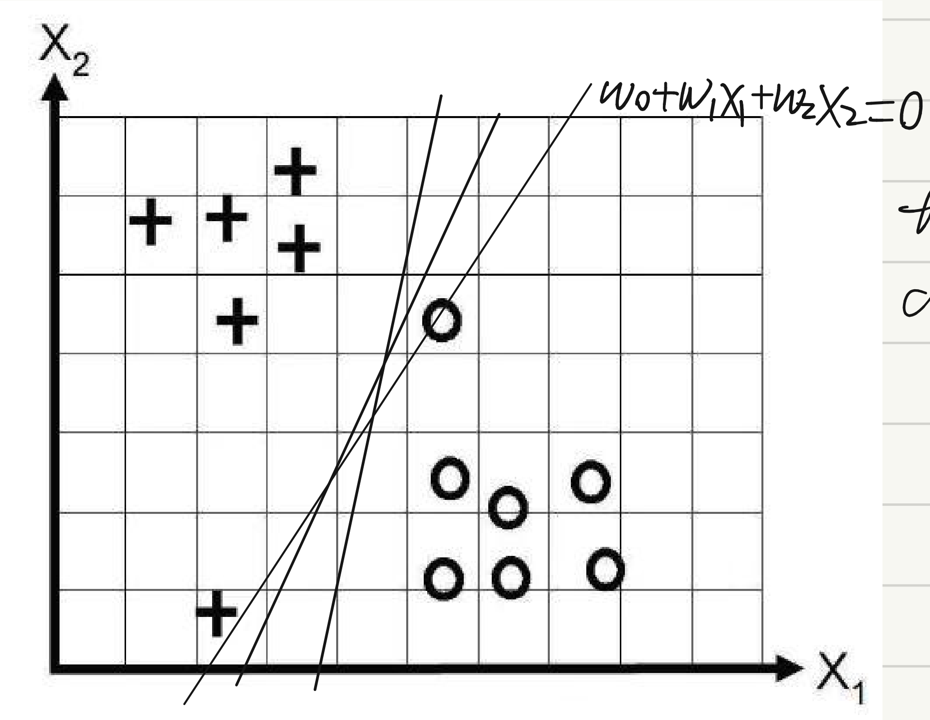
1.1 Question 1

a、 the decision boundary is given by:

$$w_0 + w_1 x_1 + w_2 x_2 = 0$$

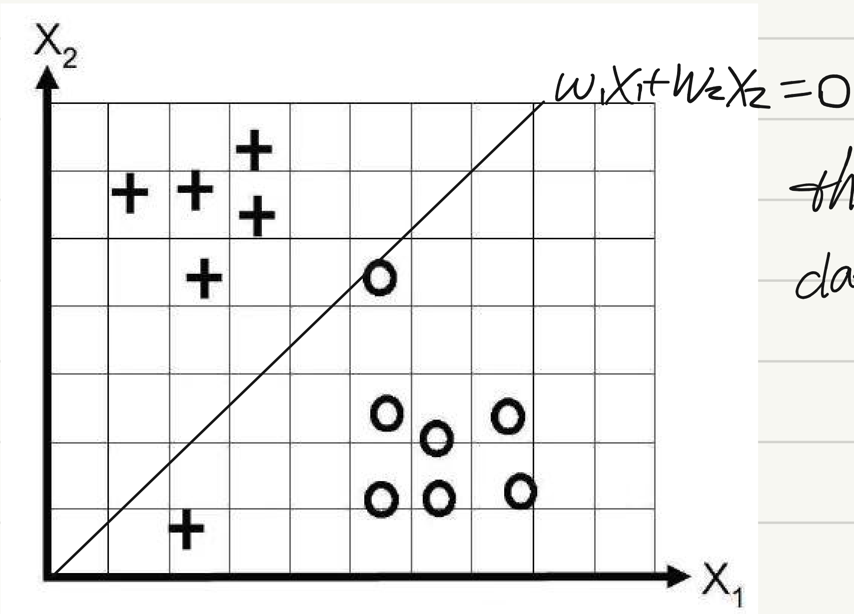
it is a line depends on $w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$

the decision boundary is not unique



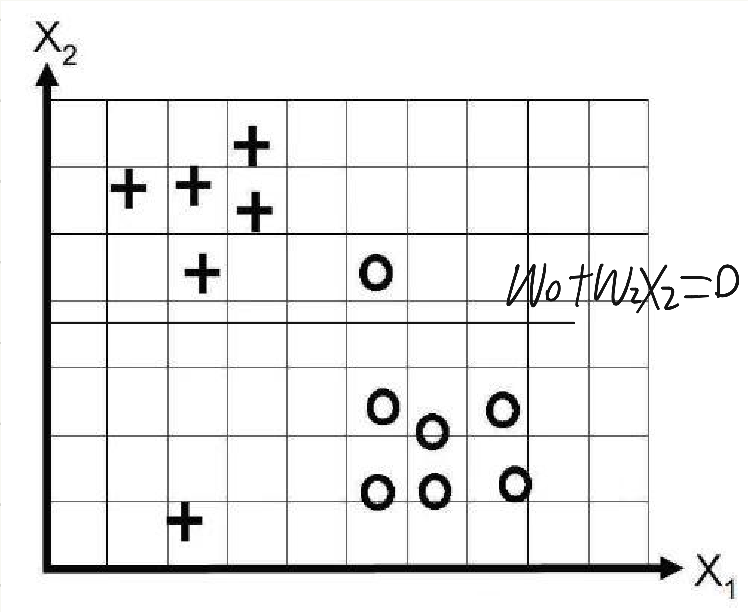
there is no classification error

b、 since w_0 all the way to 0, the decision is a line that passes the origin. $w_1x_1+w_2x_2=0$



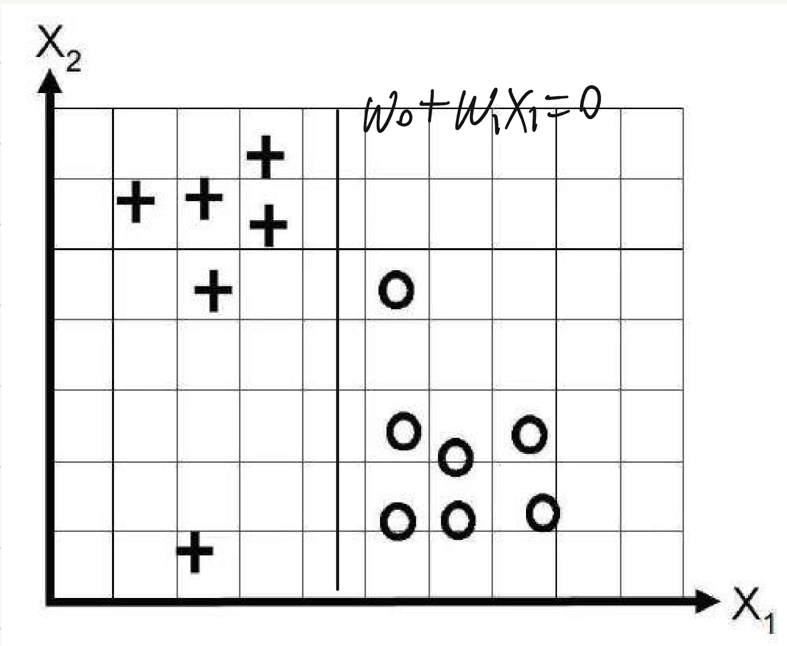
there is 1 classification error

c. w_1 all the way to 0, so the decision boundary is a line parallel to X_1 axis.



there are 2 classification errors.

d. w_2 all the way to 0, so the decision boundary is a line parallel to X_2 axis.



there is no classification error.

1.2 Question 2

a. $\phi(x_1) = [1, 0, 0]^T$

$\phi(x_2) = [1, 2, 2]^T$

decision boundary has normal $[0, 1, 1]^T$ and passes the point $[1, 1, 1]^T$.

$\because$ a vector that is parallel to w is also perpendicular to the decision boundary, so the vector can be $[0, 1, 1]^T$

b. the margin is the distance between $[1, 1, 1]^T$ and $[1, 2, 2]^T$

$$\sqrt{(2-1)^2 + (2-1)^2 + (1-1)^2} = \sqrt{2}$$

c. $\frac{1}{\|w\|} = \sqrt{2} \Rightarrow \frac{1}{\|w\|} = \frac{\sqrt{2}}{2}$

$w = k \cdot [0, 1, 1]^T$

$\therefore \sqrt{k^2 + k^2} = \sqrt{2}k = \frac{\sqrt{2}}{2} \Rightarrow k = \frac{1}{2}$

$\therefore w = [0, \frac{1}{2}, \frac{1}{2}]^T$

d. $w^T \phi(x_1) = 0$

$\therefore -w_0 = 1 \Rightarrow w_0 = -1$

e. $f(x) = \frac{\sqrt{2}}{2}x + \frac{x^2}{2} - 1$

1.3 Question 3

No. the resulting decision boundary can't guaranteed to separate the classes, since the margin can be increased by considering the slack variable, which allows some points locate inside the margin and may appear on the wrong side.

1.4 Question 4

(1) prime problem : $\min_{w,b} \frac{1}{2}\|w\|^2$

s.t. $1-y_i(w^T x_i+b) \le 0, \ \forall i$

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

the corresponding dual problem :

$$\max_\alpha \sum_i^m \alpha_i - \frac{1}{2}\sum_{i,j}^m \alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } \sum_i^m \alpha_i y_i = 0, \ \alpha_i \ge 0, \forall i$$

$\therefore \max_\alpha \sum_i^m \alpha_i - \frac{1}{2}\sum_{i,j}^m \alpha_i\alpha_j y_i y_j x_i^T x_j$

$= \max_\alpha \alpha_1+\alpha_2+\alpha_3+\alpha_4-\frac{1}{2}\alpha_1^2-\frac{1}{2}\alpha_2^2-\frac{1}{2}\alpha_3^2-\frac{1}{2}\alpha_4^2-\alpha_1\alpha_3-\alpha_2\alpha_4$

$= \max_\alpha g(\alpha)$

we have $-\alpha_1-\alpha_2+\alpha_3+\alpha_4=0$

$\Rightarrow \max_\alpha g(\alpha) = 2\alpha_1+2\alpha_2-\alpha_1^2-2\alpha_2^2-\alpha_3^2-2\alpha_1\alpha_2+2\alpha_2\alpha_3$

$\frac{\partial g}{\partial \alpha_1} = 2-2\alpha_1-2\alpha_2=0$

$\frac{\partial g}{\partial \alpha_2} = 2-4\alpha_2-2\alpha_1+2\alpha_3=0 \quad \Rightarrow \quad \alpha_1+\alpha_2=\alpha_3+\alpha_4=1, \ \alpha_2=\alpha_3$

$\frac{\partial g}{\partial \alpha_3} = -2\alpha_3+2\alpha_2=0$

$$w=-\alpha_1\begin{bmatrix}1\\0\end{bmatrix}-\alpha_2\begin{bmatrix}0\\1\end{bmatrix}+\alpha_3\begin{bmatrix}-1\\0\end{bmatrix}+\alpha_4\begin{bmatrix}0\\-1\end{bmatrix}=\begin{bmatrix}-\alpha_1-\alpha_3\\-\alpha_2-\alpha_4\end{bmatrix}=\begin{bmatrix}-\alpha_1-\alpha_2\\-\alpha_3-\alpha_4\end{bmatrix}=\begin{bmatrix}-1\\-1\end{bmatrix}$$

$$b=-1-(-\alpha_1-\alpha_3)=-1+\alpha_1+\alpha_3=-1+\alpha_1+\alpha_2=-1+1=0$$

$\therefore$ the svm : $w=\begin{bmatrix}-1\\-1\end{bmatrix}, \ b=0$

(2) since $\alpha_1,\alpha_2,\alpha_3,\alpha_4>0$, so the four given data points are all support vectors.

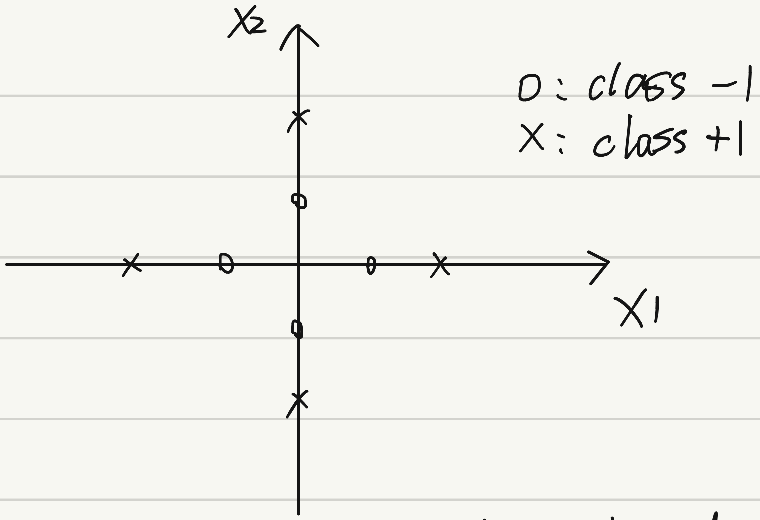
(3)

$$w^T x+b = \begin{bmatrix}-1 & -1\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = -3<0$$

$\therefore$ the predicted label of [1; 2] is -1

1.5 Question 5

(1)



Yes, we can use kernel to make the data points become separable.

(2) dual problem :

$$\max_{\alpha} \sum_{i}^{m} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \phi^T(x_i)\phi(x_j)$$

$$\text{s.t.} \sum_{i}^{m} \alpha_i y_i = 0 \ , \alpha_i \geq 0, \ \forall i$$

$$\phi(x_1) = [1;0] = \phi(x_3)$$

$$\phi(x_2) = [0;1] = \phi(x_4)$$

$$\phi(x_5) = [4;0] = \phi(x_7)$$

$$\phi(x_6) = [0;4] = \phi(x_8)$$

$$\max_{\alpha} g(\alpha)$$

$$= \max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\left(\alpha_1^2 + \alpha_2^2 + 16\alpha_3^2 + 16\alpha_4^2 - 8\alpha_1\alpha_3 - 8\alpha_2\alpha_4\right)$$

$$= \max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\alpha_1^2 - \frac{1}{2}\alpha_2^2 - 8\alpha_3^2 - 8\alpha_4^2 + 4\alpha_1\alpha_3 + 4\alpha_2\alpha_4$$

$$\text{s.t.} \ -\alpha_1 - \alpha_2 + \alpha_3 + \alpha_4 = 0 \ , \text{if } \alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0$$

$$\therefore \max_{\alpha} 2\alpha_1 + 2\alpha_2 - \frac{17}{2}\alpha_1^2 - \frac{9}{2}\alpha_2^2 - 16\alpha_3^2 - 12\alpha_1\alpha_2 + 20\alpha_1\alpha_3 + 12\alpha_2\alpha_3$$

$$\triangleq \max_{\alpha} g(\alpha)$$

$$\frac{\partial g}{\partial \alpha_1} = 2 - 17\alpha_1 - 12\alpha_2 + 20\alpha_3 = 0 \ , \ \frac{\partial g}{\partial \alpha_2} = 2 - 9\alpha_2 - 12\alpha_1 + 12\alpha_3 = 0$$

$$\frac{\partial g}{\partial \alpha_3} = -32\alpha_3 + 20\alpha_1 + 12\alpha_2 = 0$$

$$w = \begin{bmatrix} -\alpha_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -\alpha_2 \end{bmatrix} + \begin{bmatrix} 4\alpha_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 4\alpha_4 \end{bmatrix} = \begin{bmatrix} 4\alpha_3 - \alpha_1 \\ 4\alpha_4 - \alpha_2 \end{bmatrix}$$

$\therefore\ w = \left[\frac{2}{3}\ \ \frac{2}{3}\right]^T,\quad b = -1-(-a_1+4a_3) = -\frac{5}{3}$

for $[1;2]$, $\phi(x) = [1;4]$

$$w^T\phi(x)+b = \frac{2}{3}+\frac{2}{3}\times 4-\frac{5}{3} = \frac{5}{3} > 0$$

$\therefore$ the predicted label of $[1;2]$ is $+1$

1.6 Question 6

let x_i be a support vector and y_i is its label

$$r = \frac{y_i(w^T x_i+b)}{\|w\|}$$

since x_i is a support vector, we have:

$$y_i(w^T x_i+b) = 1$$

$$\therefore\ r = \frac{1}{\|w\|} \Rightarrow \frac{1}{r^2} = \|w\|^2$$

$$w = \sum_{n=1}^{N} a_n t_n x_n$$

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

$$= \sum_{n=1}^{N} a_n t_n w^T x_n$$

we have $\sum_{n=1}^{N} a_n t_n = 0$

$\therefore$ multiply a constant b: $\sum_{n=1}^{N} a_n t_n b = 0$

$$\therefore\ \|w\|^2 = \sum_{n=1}^{N} a_n t_n w^T x_n + \sum_{n=1}^{N} a_n t_n b$$

$$= \sum_{n=1}^{N} a_n t_n (w^T x_n + b)$$

for support vector, $a_n t_n (w^T x_n+b) = a_n$

otherwise: $a_n = 0 \Rightarrow a_n t_n (w^T x_n+b) = 0$

$\therefore\ \|w\|^2 = \sum_{n=1}^{N} a_n \quad \therefore\ \frac{1}{r^2} = \sum_{n=1}^{N} a_n$. Q.E.D.