Jian Xu

jxu72364@usc.edu

Problem1

1. (a)

$g(x) = w_0 + w^T x = 0$

so, we can get $w_1 x_1 + w_2 x_2 + \cdots w_n x_n + w_0 = 0$

for any 2 points x_1 and x_2 on the H

we can get that $\begin{cases} w^T x_1 = b \\ w^T x_2 = b \end{cases}$

and for $x_1 - x_2$ let $\vec{v} = x_1 - x_2$

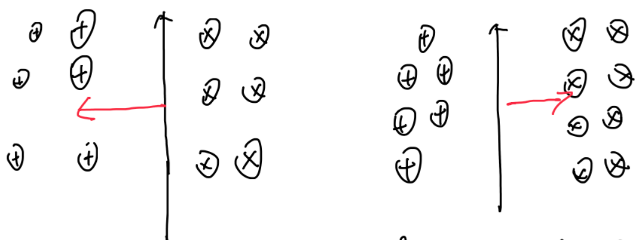
we can get $w^T \cdot \vec{v} = w^T(x_1 - x_2) = w^T x_1 - w^T x_2 = 0$

so, any vector on the H multiply w will be 0.

so, we can know that vector $\vec{w}$ is normal to H.

(b) $y = \operatorname{sign}(w^T x + b)$

we can know that $y=1$ will define the positive class and $y=-1$ will define the nagative class



we can proof by these 2 pictures. the black line is the H, and the red line is the normal vector of the H. The circles with the plus sign are posite class and with cross are negatie class.

the normal vector will hae the form $w_T = [\pm 1, 0]^T$

so, if the posite is +1, then $w^T x$ will be positive if w points to the direction of the positive areas. otherwise, if $w^T x$ is negative, for the nagative class, and our output should be -1. so, the function will always have the correct sign to produce the correct class. If we invert the class labels, the weight vector will flip its direction in order to point to the negative class.

As a result, w always points to the posite side of H.

(c) the hyperplane $g(x) = w^t x = 0$ in augmented feature space

$$r = \| proj_w(x - X_i) \| = \left\| \frac{(x - X_i) \cdot w}{w \cdot w} w \right\| = | x \cdot w - X_i \cdot w | \frac{\|w\|}{\|w\|^2}$$

$$= \frac{| x \cdot w - X_i \cdot w |}{\|w\|}$$

$$r = \| proj_w(x - X_i) \| = \frac{| x \cdot w |}{\|W\|}$$

so we can get that $r = \frac{g(x)}{\|w\|}$

(d) we know that

$$w = w_p + r \frac{x}{\|x\|}$$

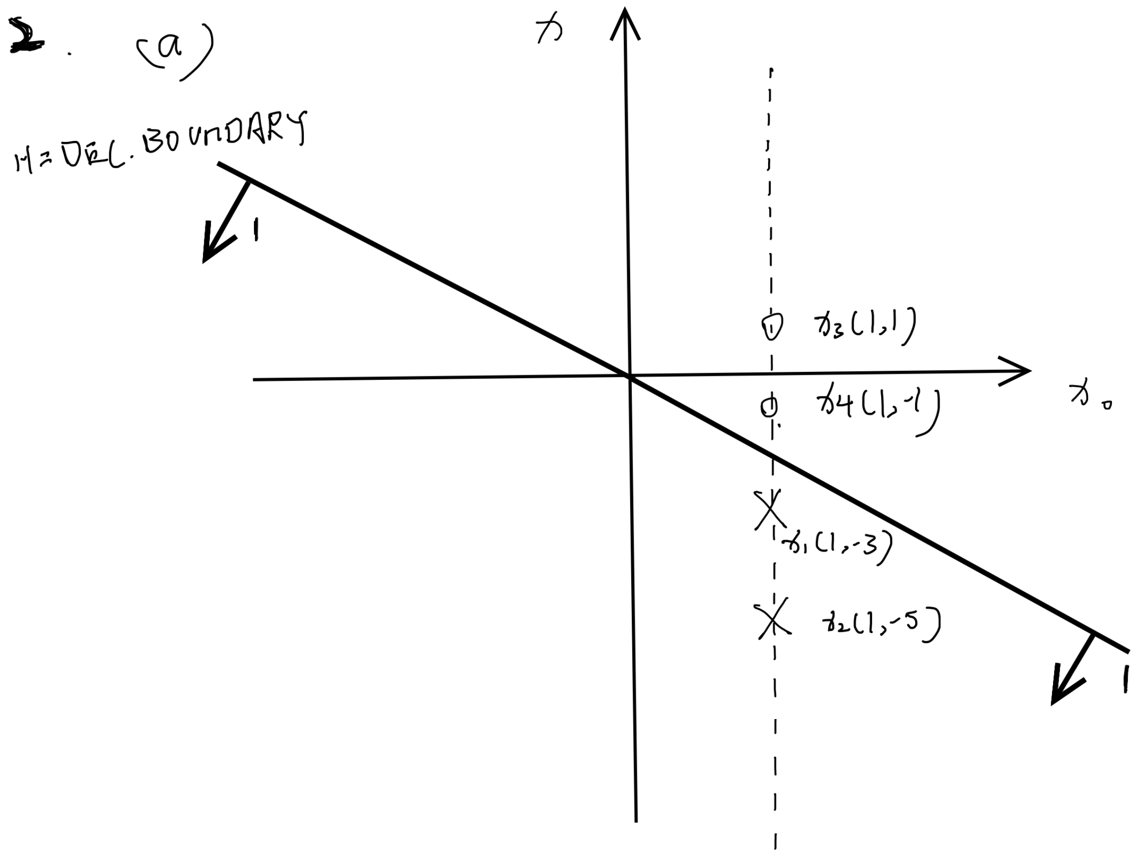
since $g(w_p) = 0$

so we have $g(w) = x^t w = r \|x\|$

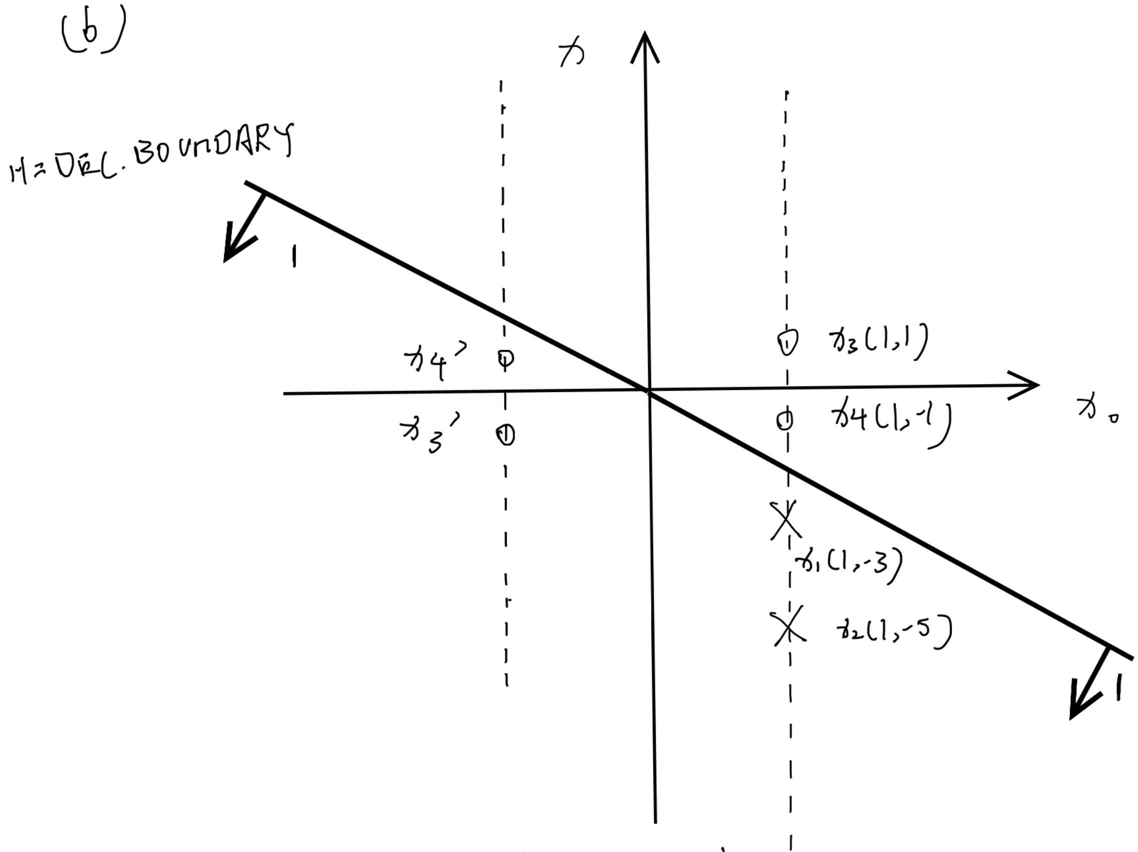
so, $r = \frac{g(w)}{\|x\|}$

Problem2

2. (a)

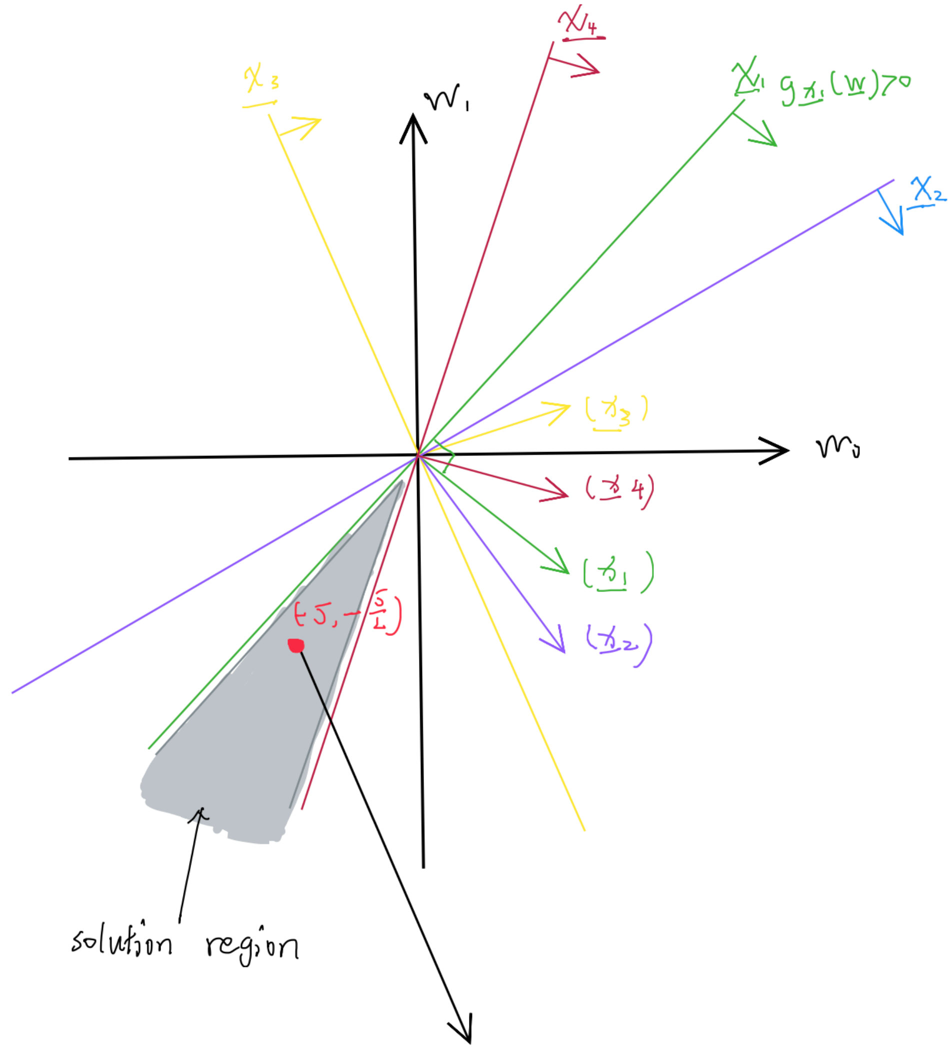


(b)



Yes, it still classify them correctly.

(c)



(d)

the function of decision boundry will be $g(x) = -2x$

so, the function of weight vector will be $\frac{1}{2}x$, in the weight space will be $w_2 = \frac{1}{2} w_1$

so, we can choose the point $(-5, -\frac{5}{2})$. we can find that this point is in the solution region.

Problem3

(3)

(a) we can know that

$\nabla_x f(x) = \frac{\partial f(x)}{\partial x}$ with the chain rule

so, $\nabla_{\underline{x}} f[p(x)] = \frac{d}{dp} f(p) \cdot \frac{d\, p(x)}{dx}$

$\frac{dp(x)}{dx} = \nabla_{\underline{x}}\, p(\underline{x})$

so, $\nabla_{\underline{x}} f[p(x)] = \frac{d}{dp} f(p) \cdot \frac{d\, p(x)}{dx} = \left[\frac{d}{dp} f(p)\right] \nabla_{\underline{x}}\, p(\underline{x})$

(b) relation (18) tell us $\frac{\partial}{\partial x}[x^T M x] = [M + M^T] x$

$\nabla x(x^T x) = \frac{\partial (x^T x)}{\partial x}$ let M be a identity matrix

$\nabla x(x^T x) = 2Mx$

$\nabla x(x^T x) = 2Ix$

$\nabla x(x^T x) = 2\underline{x}$

(c) $\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial (x^T x)}{\partial x} = \begin{pmatrix} \frac{\partial (x_1^2 + x_2^2 + \cdots x_n^2)}{\partial x_1} \\ \frac{\partial (x_1^2 + x_2^2 + \cdots x_n^2)}{\partial x_2} \\ \vdots \\ \frac{\partial (x_1^2 + x_2^2 + \cdots x_n^2)}{\partial x_n} \end{pmatrix}$

$\nabla x(x^T x) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix}$

$\nabla x(x^T x) = 2\underline{x}$

so, it proves the result in part (b)

(d) $\nabla_{\underline{x}} \left[(x^T x)^3 \right] = \begin{pmatrix} \dfrac{\partial (x_1^2 + x_2^2 + \cdots x_n^2)^3}{\partial x_1} \\ \vdots \\ \dfrac{\partial (x_1^2 + x_2^2 + \cdots x_n^2)^3}{\partial x_n} \end{pmatrix}$

$= \begin{pmatrix} 3(x_1^2 + x_2^2 + \cdots x_n^2)^2 \cdot 2x_1 \\ 3(x_1^2 + x_2^2 + \cdots x_n^2) \cdot 2x_2 \\ \vdots \\ 3(x_1^2 + x_2^2 + \cdots x_n^2) \cdot 2x_n \end{pmatrix}$

$= \begin{pmatrix} 6x_1 (x^T x)^2 \\ 6x_2 (x^T x)^2 \\ \vdots \\ 6x_n (x^T x)^2 \end{pmatrix}$

$= 6 \cdot \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ \vdots & x_2 & \cdots & 0 \\ 0 & \vdots & \ddots & x_n \end{pmatrix} \cdot (x^T x)^2$

$= 6 \underline{x} (x^T \underline{x})^2$

Problem4

(4)

(a) $\|w\|_2 = \sqrt{\sum_{i=1}^{N} w_i^2}$

let $p = w^T w$

so $\|w\|_2 = \sqrt{p}$

$$\nabla_w \|w\|_2 = \nabla_w p^{\frac{1}{2}} = \begin{pmatrix} \frac{\partial p^{\frac{1}{2}}}{w_1} \\ \vdots \\ \frac{\partial p^{\frac{1}{2}}}{w_n} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial (w_1^2 + w_2^2 + \cdots w_n^2)^{\frac{1}{2}}}{\partial w_1} \\ \vdots \\ \frac{\partial (w_1^2 + \cdots w_n^2)^{\frac{1}{2}}}{\partial w_n} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}(w_1^2 + \cdots w_n^2)^{-\frac{1}{2}} \cdot 2w_1 \\ \vdots \\ \frac{1}{2}(w_1^2 + \cdots w_n^2)^{-\frac{1}{2}} \cdot 2w_n \end{pmatrix}$$

$$= \begin{pmatrix} w_1 \cdot \frac{1}{\|w\|_2} \\ w_2 \cdot \frac{1}{\|w\|_2} \\ \vdots \\ w_n \cdot \frac{1}{\|w\|_2} \end{pmatrix}$$

$$= \underline{w} \cdot \frac{1}{\|w\|_2}$$

(b) $\nabla_w \| \underset{\sim}{M} \underset{\sim}{w} - \underset{\sim}{b} \|_2$ let $p = (Mw-b)^T (Mw-b)$

$$\nabla_w \| Mw - b \|_2 = \nabla_w p^{\frac{1}{2}}$$

$$= \begin{pmatrix} \frac{\partial p^{\frac{1}{2}}}{\partial w_1} \\ \vdots \\ \frac{\partial p^{\frac{1}{2}}}{\partial w_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial \left[(M_1 w_1 - b)^2 + \cdots (M_n w_n - b)^2 \right]^{\frac{1}{2}}}{\partial w_1} \\ \vdots \\ \frac{\partial \left[(M_n w_n - b)^2 + \cdots (M_n w_n - b)^2 \right]^{\frac{1}{2}}}{\partial w_n} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2} \left[(M_1 w_1 - b)^2 + \cdots (M_n w_n - b)^2 \right]^{-\frac{1}{2}} \cdot 2 (M_1 w_1 - b) \cdot M_1 \\ \vdots \\ \frac{1}{2} \left[(M_1 w_1 - b)^2 + \cdots (M_n w_n - b)^2 \right]^{-\frac{1}{2}} \cdot 2 (M_n w_n - b) \cdot M_n \end{pmatrix}$$

$$= \begin{pmatrix} M_1 (M_1 w_1 - b) \cdot \frac{1}{\| Mw - b \|_2} \\ M_2 (M_2 w_2 - b) \cdot \frac{1}{\| Mw - b \|_2} \\ \vdots \\ M_n (M_n w_n - b) \cdot \frac{1}{\| Mw - b \|_2} \end{pmatrix}$$

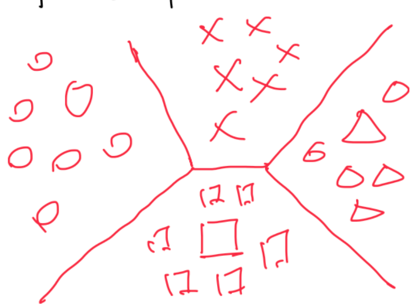
$$= \underset{\sim}{M} (\underset{\sim}{M} \underset{\sim}{w} - \underset{\sim}{b}) \cdot \frac{1}{\| Mw - b \|_2}$$

Problem5

5. we want to show that linear separability doesn't necessary imply total linear separability.

the counterexample will be like this:

for $c=4$



we can use KVM to separate them, that means they are linearly separability, but we can not find a linear function to separate "x" with the rest. So, that means they are not total linear separability.

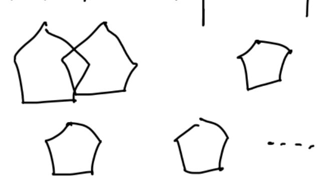
But if they are total linear separability, we want to proof that it will be linear separability.

total linear $\Rightarrow$ linear

Contraposition

not linear separability $\Rightarrow$ not total linear separability

if they are not linear separability, it means that at least 2 convex will intersect.



from the draft, we can easily find that for all classes, they are not total linear separability. So, the contraposition is proved, then we can know that total linear separability implies linear separability.