ตัวอย่าง: เลือก $\{x_1, x_2, x_3, x_4, x_5\} \subset \mathbb{R}^4$

ที่สามารถแบ่งด้วย perceptron ได้

X_-, X_+ ที่ $\operatorname{conv}(X_-) \cap \operatorname{conv}(X_+) = \emptyset$

ก็ต่อเมื่อ มี $(w, \theta) \in \mathbb{R}^{4+1}$ ที่ทำให้

$$\langle w, x_i \rangle - \theta \begin{cases} < 0 & , x_i \in X_- \\ \geq 0 & , x_i \in X_+ \end{cases}$$

เลือก $w \in \mathbb{R}^4$, $\theta \in \mathbb{R}$

$w = (0, 1, -1, 0)$, $\theta = 1$

เลือก x_i ที่ $\langle w, x_i \rangle - \theta \geq 0 \Rightarrow x_i \in X_+$

x_i ที่ $\langle w, x_i \rangle - \theta < 0 \Rightarrow x_i \in X_-$

Ex: $x_1 = (0, 1, 1, 0) \Rightarrow \langle w, x_1 \rangle - \theta$
$= 0 - 1 < 0 \Rightarrow x_1 \in X_-$

$x_2 = (0, 1, -1, 0) \Rightarrow \langle w, x_2 \rangle - \theta$
$= 2 - 1 = 1 > 0 \Rightarrow x_2 \in X_+$

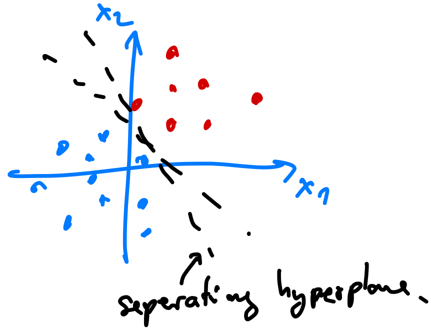
$\vdots$

$x_5 = \cdots$

$$X = X_- \cup X_+ = \left\{ \underbrace{\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}}_{x_1}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \cdots\cdots, \right\}$$

$\Rightarrow$ Perceptron Learning Algorithm:

ได้ว่า X_+ และ X_- เป็น finite set ที่แบ่งแยก affinely separable ได้



แสดงว่า $\mathrm{conv}(X_-) \cap \mathrm{conv}(X_+) = \emptyset$

(สามารถแบ่งได้ แต่ไม่ unique.)

ปัญหาในเรื่องนี้: จะสามารถสร้าง seperating hyperplane ที่แบ่ง X_+ และ X_- ได้อย่างไร

Perceptron Learning Problem:
(PL Problem)

ให้ $X_+, X_- \subset \mathbb{R}^n$ เป็น finite sets ที่ $\mathrm{conv}(X_+) \cap \mathrm{conv}(X_-) = \emptyset$ นั่นคือ สามารถแบ่งได้โดยใช้ perceptron. ขอใช้อัลกอริทึมต่อไปนี้ หา

$(\omega, \theta) \in \mathbb{R}^{n+1}$ ที่ว่า

$$\mathrm{sign}(\underbrace{\langle \omega, x \rangle - \theta}_{\langle \hat{\omega}, \hat{x} \rangle}) = \begin{cases} +1 & \text{ถ้า } x \in X_+ . \\ -1 & \text{ถ้า } x \in X_- . \end{cases}$$

ในเรื่องต่อไปนี้ เราจะปรับปัญหา โดยตัดค่า threshold θ. โดยให้ $\hat{x} = \begin{pmatrix} x \\ 1 \end{pmatrix}$, $\hat{\omega} = \begin{pmatrix} \omega \\ -\theta \end{pmatrix}$ ในโดเมน $\mathbb{R}^n \times \{1\}$

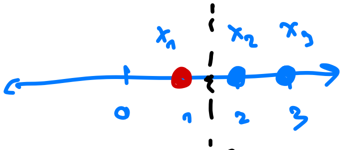
และพิจารณาปัญหา.

$$\mathrm{sign}(\langle \hat{\omega}, \hat{x} \rangle) = \begin{cases} +1, & \hat{x} \in \hat{X}_+ \\ -1, & \hat{x} \in \hat{X}_- \end{cases}$$

โดยที่: $\hat{X}_+ = X_+ \times \{1\}$, $\hat{X}_- = X_- \times \{1\} \subset \mathbb{R}^n \times \{1\}$

ดังนั้น ในเรื่องต่อไปนี้ เราจะต้องหา linear separating hyperplane ของ $\hat{X}_+, \hat{X}_- \subset \mathbb{R}^{n+1}$

Idea:

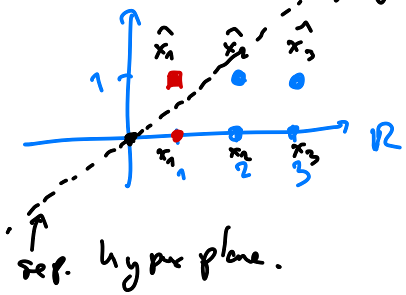


, $x_1 = 1,\ x_2 = 2,\ x_3 = 3$

Sep. hyperplane. $x - 1.5 = 0$ $\left(\begin{array}{l} w = 1 \\ \theta = 1.5 \end{array}\right)$

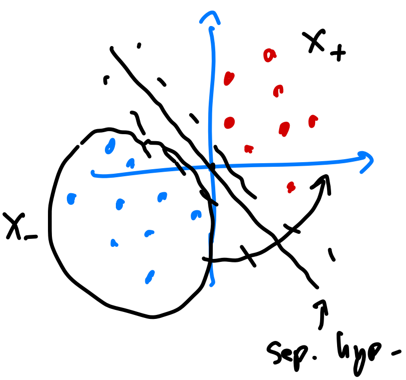
ดังนั้น.

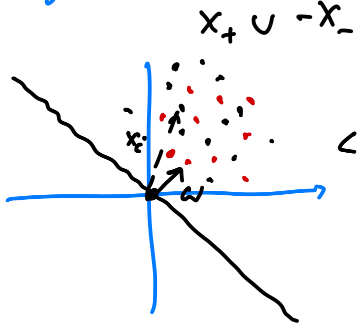
$\Rightarrow$



$$\hat{x}_1 = \begin{pmatrix} x_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$\hat{x}_2 = \begin{pmatrix} x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$
$$\hat{x}_3 = \begin{pmatrix} x_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$\Rightarrow$ ถ้าข้อมูลอยู่ในรูป linearly separable.





$\langle \hat{x}_i, w \rangle \geq 0 \quad \forall x_i$

เวกเตอร์ $\hat{x}$ ที่อยู่ในรูปของ

$$\hat{X} := \hat{X}_+ \cup \{-\hat{x} : \hat{x} \in \hat{X}_-\} \subset \mathbb{R}^{n+1}$$

จะเปลี่ยนโจทย์ หา $\hat{\omega} \in \mathbb{R}^{n+1}$ ที่ทำให้

$$\hat{\omega} \cdot \hat{x} > 0 \quad \forall \hat{x} \in \hat{X}$$

ดังนั้น PL Problem จะสามารถเขียนได้ในรูป.

Modified PL Problem: ให้หาเวกเตอร์ $\hat{\omega} \in \mathbb{R}^{n+1}$

ที่ซึ่ง

$$\langle \hat{w}, \hat{x} \rangle > 0 \quad \text{สำหรับทุก } \hat{x} \in \hat{X} \qquad (1)$$

ถ้า $\hat{X} = \{\hat{x}_1, \ldots, \hat{x}_N\}$ แล้วจะสามารถสร้างเมทริกซ์

$A = [\hat{x}_1, \ldots, \hat{x}_N]^T$ และจะเขียน (1) ได้ในรูป

$$A\hat{w} > 0$$

มากกว่า 0 ในทุกเวกเตอร์คอมโพเนนท์.

ในกรณีหา $\hat{w}$ จำเป็นต้องสร้างชุดของการเรียนรู้ (training set) นิยามโดย:

Def: ให้ $\hat{X} \subset \mathbb{R}^{n+1}$ เป็น finite set จะเรียกลำดับ $u : \mathbb{N} \to \hat{X}$ ว่าเป็น training sequence สำหรับ $\hat{X}$ ถ้าทุก $\hat{x} \in \hat{X}$ และ $k_0 \in \mathbb{N}$ แล้วจะมี $k \geq k_0$ ที่ซึ่ง $u(k) = \hat{x}$.

(หรืออีกนัยหนึ่ง คือ แต่ละ $\hat{x} \in \hat{X}$ จะปรากฏ $\hat{x}$ ไม่จำกัดจำนวนครั้ง ในลำดับ training set u).

Ex: $\hat{X} = \{x_1, x_2, x_3\}$

$k_0 = 3$

$$u = (\underbrace{x_1, x_3, x_2}, \underbrace{x_1, x_2, x_3}, \underbrace{x_3, x_2, x_1}, \ldots)$$

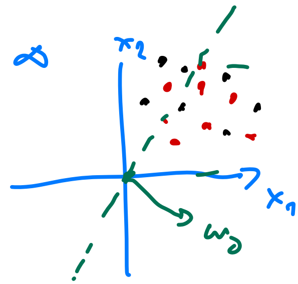
เรียงลำดับกันในทุกจำนวนสมาชิกของ $\hat{X}$

PL Algorithm: ให้ $A = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]^T \in \mathbb{R}^{N \times m}$, $(m = n+1)$

u เป็น training sequence สำหรับ $\hat{X} = \{\hat{x}_1, \dots, \hat{x}_N\}$

และให้ $\varphi : \mathbb{N} \to \mathbb{R}$ เป็นลำดับ ที่ซึ่ง

$$0 < \inf_{k \in \mathbb{N}} \varphi(k) \leq \sup_{k \in \mathbb{N}} \varphi(k) < \infty$$



แล้ว ขน. ในแต่ละรอบ จะทำซ้ำ.

ข0: เลือก $w(0) \in \mathbb{R}^m$, ให้ $k := 0$

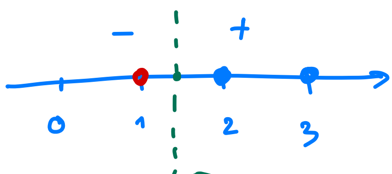
ข1: set

$$w(k+1) = \begin{cases} w(k) & , \langle u(k), w(k) \rangle > 0 \\ w(k) + \varphi(k) u(k) & , \langle u(k), w(k) \rangle \leq 0 \end{cases}$$

ข2: If $Aw(k+1) > 0$ stop.

Else $k = k+1$ GoTo ข1.

Ex:



aff. sep. $\langle w, x \rangle - \theta = 0$

$(x - 1.5 = 0 \Rightarrow w = 1, \theta = 1.5)$

PL Alg: $X = \{x_1, x_2, x_3\}$, $x_1 = 1, x_2 = 2, x_3 = 3$

$X_+ = \{x_2, x_3\} = \{2, 3\}$

$X_- = \{x_1\} = \{1\}$

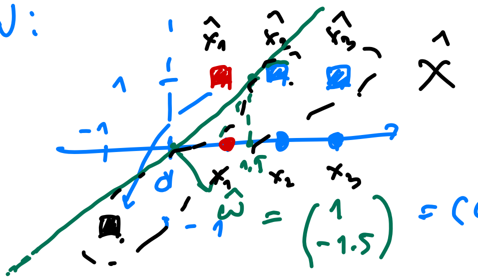
จะได้ $\hat{X}_+ = \left\{ \begin{pmatrix} x_2 \\ 1 \end{pmatrix}, \begin{pmatrix} x_3 \\ 1 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$

$\hat{X}_- = \left\{ \begin{pmatrix} x_1 \\ 1 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$

$$\hat{X} = \hat{X}_+ \cup \{ -\hat{x} : \hat{x} \in \hat{X}_- \}$$

$$= \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} = \left\{ \underset{x_3}{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}, \underset{x_2}{\begin{pmatrix} 3 \\ 1 \end{pmatrix}}, \underset{x_1}{\begin{pmatrix} -1 \\ -1 \end{pmatrix}} \right\}$$

รูป :



$\hat{\omega} = \begin{pmatrix} 1 \\ -1.5 \end{pmatrix} = (\omega, -\theta)$ ซึ่ง $\langle \hat{\omega}, \hat{x}_i \rangle = \langle \omega, x \rangle - \theta > 0$

$\forall\, i = 1, 2, 3.$

I0: เลือก $\omega(0) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\hat{X} = \left\{ \overset{x_1}{\begin{pmatrix} -1 \\ -1 \end{pmatrix}}, \overset{x_2}{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}, \overset{x_3}{\begin{pmatrix} 3 \\ 1 \end{pmatrix}} \right\}$

สร้าง $A = [\hat{x}_1, \hat{x}_2, \hat{x}_3] = \left[\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right]^T = \begin{bmatrix} -1 & -1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$

ตรวจสอบ $A\omega(0) = \begin{bmatrix} -1 & -1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \not> 0$ ← ต้องแก้ไข x_1

I1: ให้ $\varphi(0) = 1$, $u(0) = x_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

ทำให้ $\langle \omega(0), u(0) \rangle = 0$ จริง

$$\omega(1) = \omega(0) + \varphi(0) u(0)$$

$$= \begin{pmatrix} 1 \\ -1 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

ตรวจสอบ. $A\omega(1)$

$\Rightarrow \quad Aw(1) = \begin{bmatrix} -1 & -1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ -2 \end{pmatrix}$ ← x_2, ← x_3

I2: ให้ $\varphi(1) = 1$ ($\frac{1}{2}$), $u(1) = x_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

ทำได้ $\langle w(1), u(1) \rangle = \left\langle \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\rangle = -2 < 0$

จะได้ $w(2) = w(1) + \varphi(1)\, u(1)$ ($\frac{1}{2}$)

$= \begin{pmatrix} 0 \\ -2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$

$\varphi(1) = \frac{1}{2}$

$\Rightarrow w(2) = \begin{pmatrix} 0 \\ -2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$= \begin{pmatrix} 1 \\ -1.5 \end{pmatrix}$ ☒

ตรวจสอบ. $Aw(2) = \begin{bmatrix} -1 & -1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 5 \end{pmatrix}$ ← x_1

$\vdots$ (ให้ $\varphi(k) = 1$) $\Rightarrow w(7) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$

I8: ให้ $\varphi(7) = 1$, $u(7) = x_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

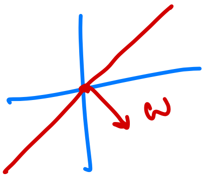
$w(8) = \begin{pmatrix} 3 \\ -2 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$

ตรวจสอบ $Aw(8) = \begin{bmatrix} -1 & -1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} > 0$ ✓

สรุป $\hat{w} = w(8) = \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} w \\ -\theta \end{pmatrix} \Rightarrow w = 2, \ \theta = 3$

ดังนั้น $h(x) = \langle w, x \rangle - \theta = 2x - 3 = 0$ ☒

$(x - 1.5 = 0)$



สำหรับ PL Algorithm จะมีจำนวนในการทำซ้ำเป็นจำนวนจำกัด.
ด้วยทฤษฎีบทต่อไปนี้.

<u>ทบ!</u> (Perceptron convergence thm.)

ให้เมทริก $A = [\hat{x}_1, \dots, \hat{x}_N]^T \in \mathbb{R}^{N \times m}$, $(m = n+1)$
ที่สร้างจาก $X_+, X_- \subset \mathbb{R}^n$ ที่ $conv(X_+) \cap conv(X_-) = \emptyset$
นั่นคือ จะมี $w^* \in \mathbb{R}^m$ ที่ทำให้

$$Aw^* > 0$$

แล้ว จาก PL Algorithm จะได้ว่าสำหรับ $w(0) \in \mathbb{R}^m$ ใดๆ
แล้วจะมี $k_L \in \mathbb{N}$ ที่ทำให้ ทุก $k > k_L$
แล้ว $w(k) = w(k_L) =: \hat{w}$ และ $A\hat{w} > 0$

<u>proof:</u> จากสมมติฐาน $Aw^* > 0$ จะมี

$$\alpha := \inf_{k \in \mathbb{N}} \varphi(k) \cdot \min_{j \in \{1, \dots, N\}} (\langle \hat{x}_j, w^* \rangle) > 0$$

กำหนด.

$$\beta := \sup_{k \in \mathbb{N}} \varphi(k) \cdot \max_{j \in \{1, \dots, N\}} \|\hat{x}_j\|_2 > 0.$$

และ $\bar{w} = \frac{\beta^2}{\alpha} w^*$

ถ้ามีการเปลี่ยนแปลง w นั่นคือ $w(k+1) \neq w(k)$
จะได้ว่า $\langle u(k), w(k) \rangle \leq 0$ นั่นคือ

$$\|w(k+1)-\bar{w}\|_2^2 = \|\underbrace{w(k)+\varphi(k)u(k)}_{w(k+1)}-\bar{w}\|_2^2$$

$$= \|\overbrace{(w(k)-\bar{w})}^{u} + \overbrace{\varphi(k)u(k)}^{v}\|_2^2$$

$$= \|w(k)-\bar{w}\|_2^2 + 2\langle \varphi(k)u(k), w(k)-\bar{w}\rangle$$

$$+ \|\varphi(k)u(k)\|_2^2 \leftarrow \le \beta^2 \qquad \langle u(k), w(k)\rangle \le 0$$

$$\left(\|u+v\|_2^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle \right)$$

$$\le \|w(k)-\bar{w}\|_2^2 - 2\underbrace{\varphi(k)\langle u(k),\bar{w}\rangle}_{\ge \beta^2} + \beta^2 \quad -(*)$$

$(\langle u(k), w(k)\rangle \le 0,)$

$\beta = \sup \varphi(k)\cdot \max \|x_j\|_2$

จะเห็นได้ว่า จาก

$$0 < \alpha = \inf_{k\in\mathbb{N}} \varphi(k)\cdot \min_{j\in\{1,\dots,N\}} \langle x_j, w^*\rangle \le \varphi(k)\langle u(k), w^*\rangle$$

ดังนั้น $\varphi(k)\langle u(k), \bar{w}\rangle = \beta^2 \underbrace{\frac{\varphi(k)\langle u(k), w^*\rangle}{\alpha}}_{\ge 1} \ge \beta^2$

$\bar{w} = \frac{\beta^2}{\alpha} w^*$

จาก $(*)$ จะได้

$$\|w(k+1)-\bar{w}\|_2^2 \le \|w(k)-\bar{w}\|_2^2 - \beta^2 \quad -(**)$$

ให้ $k_1 := 0$ และ $0 = k_1 < k_2 < k_3 < \dots$

เป็นลำดับที่ $w(k_{i-1}) \ne w(k_i)$

จะได้ว่าลำดับ. $(\omega(k_1), \omega(k_2), \dots)$ ที่ทำให้

$$0 \leq \|\omega(k_{j+1}) - \bar{\omega}\|_2^2 \overset{(**)}{\leq} \|\omega(k_j) - \bar{\omega}\|_2^2 - \beta^2$$

$$\leq \dots \leq \|\omega(0) - \bar{\omega}\|_2^2 - j\beta^2$$

จะเห็นได้ว่า $\omega(k_j) \to \bar{\omega} = \frac{\beta^2}{\alpha}\omega^*$

และ $j \leq \frac{\|\omega(0) - \bar{\omega}\|_2^2}{\beta^2}$ เป็นจำนวนที่ต้องปรับค่า ω.

ดังนั้นจะได้ว่า $\omega(k) = \omega(k_j)$ สำหรับทุก $k \geq k_j$

นั่นคือ $A\omega(k_j) > 0$ ∎

Remark:

1.) ในขั้น $\omega(k+1) = \underset{\text{(ทิศทาง)}}{\omega(k)} + \underbrace{\varphi(k)u(k)}_{\text{ควบคุมได้และควรมีขนาด 1.}}$

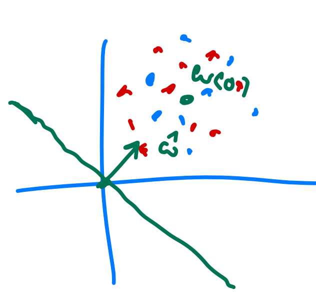
การเลือก $\varphi(k) \approx \frac{1}{\|u(k)\|_2} \Rightarrow \varphi(k)u(k) \approx 1$

Ex: $\varphi(k) = \lambda \frac{|\langle u(k), \omega(k)\rangle|}{\|u(k)\|_2^2}, \quad 0 < \lambda < 2$

2.) แทนที่จะเลือก $\omega(0) \in \mathbb{R}^{n+1}$ ใดๆ

สามารถเลือก $\omega(0)$ เป็น จุดศูนย์กลางของ $\hat{x}$

$$\left[\omega(0) = \frac{\sum_i \hat{x}_i}{N}\right]$$



การบ้าน:

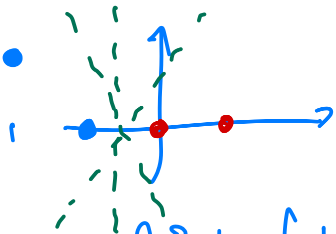
1.) ให้ยกตัวอย่าง $X_+ , X_- \subset \{0, \pm 1, \pm 2\}^2$

ที่ $\mathrm{conv}(X_+) \cap \mathrm{conv}(X_-) = \emptyset$

แล้วใช้ PL Algorithm ในการหา separating hyperplane

ของ X_+ และ X_-

Ex: $X_+ = \{(0,0),(0,1)\}$, $X_- = \{(-2,1),(-1,0)\}$

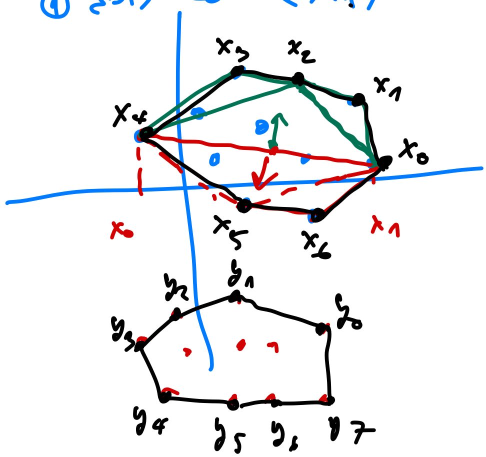


2.) Python: ให้เขียนโปรแกรมสำหรับ PL Alg.

รับ $X_+ , X_- \subset \mathbb{R}^n$ ที่ $\mathrm{conv}(X_+) \cap \mathrm{conv}(X_-) = \emptyset$.

---

① สร้าง $\mathrm{conv}(X_+)$ และ $\mathrm{conv}(X_-)$



- Quick Hull.
- check $\mathrm{conv}(X_+) \cap \mathrm{conv}(X_-) = \emptyset$.

(x_1, x_0)

(x_0, x_1)