Optimal Separation:

ให้ $X_+, X_- \subset \mathbb{R}^n$ เป็น non-empty finite sets ที่

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

ให้ $X = X_+ \cup X_- = \{x_1, \dots, x_N\}$ และให้

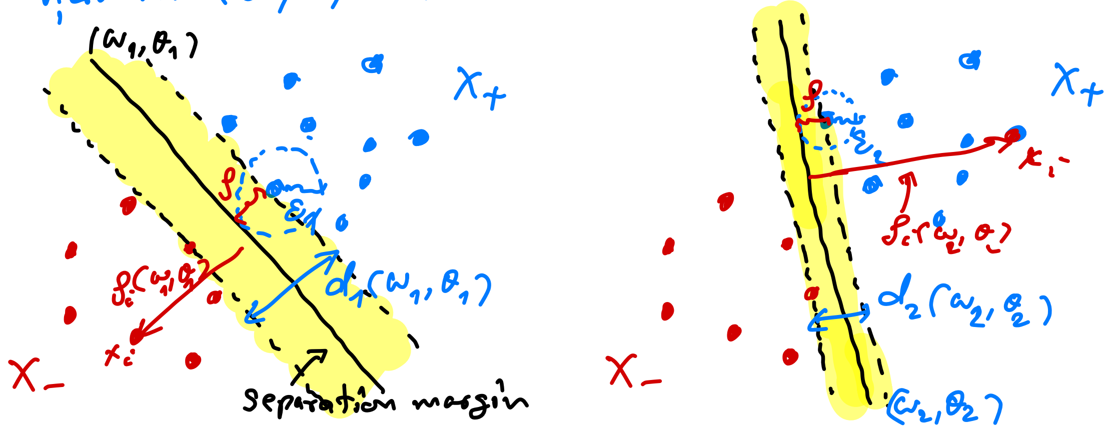
$$(x_i, y_i) = \begin{cases} (x_i, +1), & x_i \in X_+ \\ (x_i, -1), & x_i \in X_- \end{cases}$$

สมมติว่า เซตของ separating hyperplane จะอยู่ในรูปของ weight polyhedron นั่นคือ

$$\Gamma = \{ (\underset{\mathbb{R}^n}{\underset{\uparrow}{w}}, \underset{\mathbb{R}}{\underset{\uparrow}{\theta}}) \in \mathbb{R}^{n+1} : y_i(\langle w, x_i \rangle - \theta) > 0 \quad \forall i = 1, \dots, N \} \neq \emptyset$$

ปกติแล้ว separating hyperplane จะมีจำนวนไม่จำกัด

เพื่อที่จะให้คุณค่าของการแบ่งให้แต่ละ (w, θ) เราจะสร้าง functional $\rho : \Gamma \to \mathbb{R}$ เพื่อวัดและเปรียบเทียบ คุณค่าของ (w, θ) ที่แบ่งเซต X_+ และ X_- ออกจากกัน



ระยะระหว่าง X_+ และ X_- นี้ โดยใช้ (ω, θ)

ถ้ายิ่งห่างมาก จะทำให้การแบ่งยังคงถูกได้ ถึงแม้ว่าข้อมูลจะถูกรบกวน (Noise)

ดังนั้น เราจะนิยาม functional ที่ใช้วัดระยะห่าง ระหว่าง X_+ และ X_- นี้ว่า Separation margin ของ (ω, θ) นั้นต่อ.

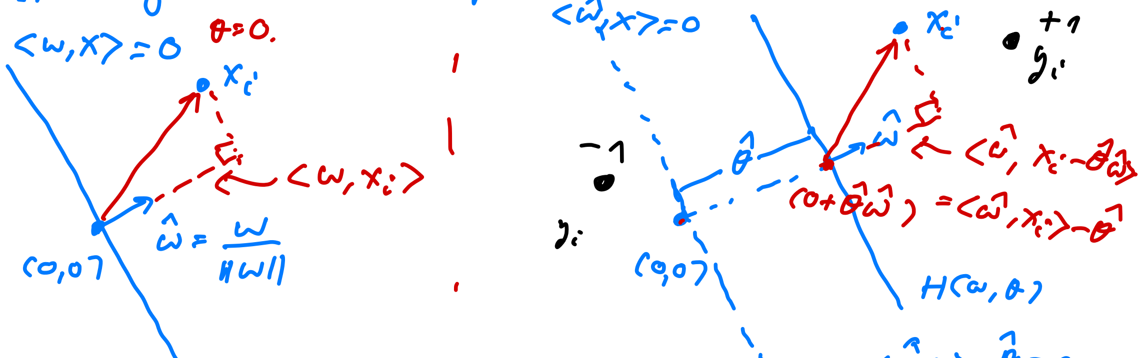
Def: กำหนดให้ $(\omega, \theta) \in \mathbb{R}^{n+1}$, $\omega \neq \vec{0}$ ให้

ระยะห่างน้อยที่สุดเทียบ

$$\rho(\omega,\theta) := \min_{1 \le i \le N} \rho_i(\omega,\theta)$$ โดยที่ $x_i \in X = X_+ \cup X_-$

$$\rho_i(\omega,\theta) = \underline{y_i} \frac{(\langle \omega, x_i \rangle - \theta)}{\|\omega\|}$$ ← ระยะห่างของ x_i กับ $H(\omega,\theta)$

จะเรียก $\rho(\omega,\theta)$ นี้ว่า separation margin ของ (ω,θ).



$$\langle \hat{\omega}, x \rangle - \hat{\theta} = 0$$

$$\langle \hat{\omega}, x \rangle - \hat{\theta}\langle \hat{\omega}, \hat{\omega} \rangle = 0$$

$$\langle \hat{\omega}, x - \hat{\theta}\hat{\omega} \rangle = 0$$

$$\left\langle \frac{\omega}{\|\omega\|}, x \right\rangle - \hat{\theta} \left\langle \frac{\omega}{\|\omega\|}, \hat{\omega} \right\rangle$$

$$\rho_i(\omega,\theta) = y_i \frac{\left(\langle \omega, x_i \rangle - \underbrace{\hat{\theta}\langle \omega, \hat{\omega} \rangle}_{\theta}\right)}{\|\omega\|}$$

สามารถเห็นได้ว่า $\rho_i(w,\theta)$ บอกระยะห่างระหว่างแต่ละจุดกับ $H(w,\theta)$

$$H(w,\theta) = \{x : \langle w,x\rangle - \theta = 0\}$$ separating hyperplane.

เนื่องจาก $x_i \in X$ จะได้ว่า ระยะห่าง จาก x_i ถึง H เป็น

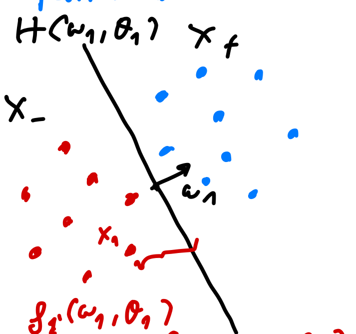
$$\mathrm{dist}(x_i, H) = \min_{x \in H} \| x_i - x \|$$

จากการพิจารณาข้างต้น จะได้ว่า

$$\mathrm{dist}(x_i, H) = \frac{(\langle w, x_i\rangle - \theta)}{\|w\|}$$

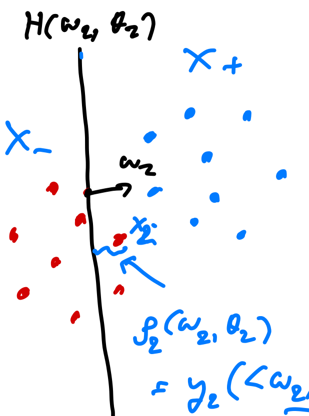
$$= \left| \underbrace{y_i}_{\pm 1} \frac{(\langle w, x_i\rangle - \theta)}{\|w\|} \right| = |\rho_i(w,\theta)|$$

โดยมีกรณีพิจารณา.



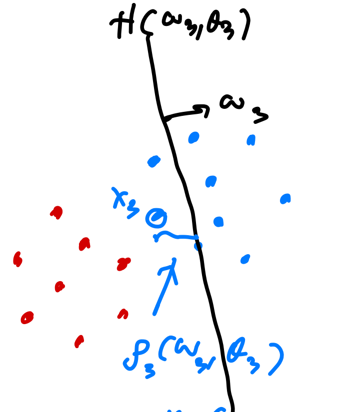
$$\rho_1(w_1,\theta_1) = \underbrace{y_1}_{-1} \underbrace{\frac{(\langle w_1, x_1\rangle - \theta_1)}{\|w_1\|}}_{= -d < 0}$$

$$= (-1)(-d) = d > 0$$



$$\rho_2(w_2,\theta_2) = \underbrace{y_2}_{-1} \underbrace{\frac{(\langle w_2, x_2\rangle - \theta_2)}{\|w_2\|}}_{= +d > 0}$$

$$= (-1)(d) = -d < 0$$



$$\rho_3(w_3,\theta_3) = \underbrace{y_3}_{+1} \underbrace{\frac{(\langle w_3, x_3\rangle - \theta_3)}{\|w_3\|}}_{(-d) < 0}$$

$$= (+1)(-d) = -d < 0$$

จะเห็นได้ว่า H จะแบ่ง X_+ ออกจาก X_- ก็ต่อเมื่อ

$$\rho(w,\theta) = \min_{1\le i\le N} \operatorname{dist}(x_i, H) = \operatorname{dist}(X, H)$$

(และ $\rho(w,\theta) < 0 < \operatorname{dist}(X,H)$ ถ้า H ไม่แบ่ง X_+, X_-)

ดังนั้น เราไม่จำเป็นต้องเลือก $(w,\theta) \in \Gamma$ ← ยังไม่รู้

เราจะเลือก $(w,\theta) \in \mathbb{R}^{n+1}$ และเรียกค่า

optimal separation margin ระหว่างเซต X_+, X_-

เป็น. $$\rho := \sup\{\rho(w,\theta) : (w,\theta) \in \mathbb{R}^{n+1}, w \neq \vec{0}\}$$

(โดยค่านี้จะหาออกมาได้ด้วย ทฤษฎีบทต่อไปนี้:

<u>ทบ:</u> (optimal separation margin).

สมมติให้ $X_+, X_- \subset \mathbb{R}^n$ เป็น non-empty finite sets

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

แล้ว ค่า optimal separation margin

$$\rho := \sup\{\rho(w,\theta) : (w,\theta) \in \mathbb{R}^{n+1}, w \neq 0\}$$

มีค่าจำกัด ($<\infty$) ที่บางค่า (w^*, θ^*) และ

$$\rho = \frac{1}{2}\operatorname{dist}(C_+, C_-) = \frac{1}{2}\operatorname{dist}(C, 0)$$

โดยที่ $C_\pm := \operatorname{conv}(X_\pm)$

(1)

ขอกล่าวว่า $H = \{ x \in \mathbb{R}^n : \langle w^*, x \rangle - \theta^* = 0 \}$ จะ unique และเรียกว่า optimal hyperplane ②

ในส่วนแรก ของ ทบ. เราต้องการผลจากบทตั้งต่อไปนี้:

<u>บทตั้ง 1</u>: ให้ $\emptyset \neq X_+, X_- \subset \mathbb{R}^n$ เป็น finite sets ที่

ถ้า $C_\pm := \mathrm{conv}(X_\pm)$ แล้ว $C_+ \cap C_- = \emptyset$

ปรับต่อ $C := \mathrm{conv}(X_+ - X_-) = C_+ - C_-$

และ $0 \notin C$ จะได้ว่า

$$\mathrm{dist}(C_+, C_-) = \mathrm{dist}(C, 0) = \max_{\|v\|=1} \min_{z \in C} \langle v, z \rangle$$

<u>proof</u>: สำหรับ $v \in \mathbb{R}^n$ ที่ $\|v\| = 1$ แล้ว $\langle v, z \rangle \leq \underbrace{\|v\|}_{=1} \|z\|$ (Cauchy-Schwarz) $\leq \|z\|$

ดังนั้น

"⇒": $\min_{z \in C} \langle v, z \rangle \leq \min_{z \in C} \|z\| = \mathrm{dist}(C, 0)$

"⇐": เนื่องจาก C เป็น convex set และ $0 \in C$

ดังนั้น จะมี $0 \neq z^* \in C$ ที่ทำให้

$$\|z^*\| = \min_{z \in C} \|z\| = \mathrm{dist}(C, 0)$$

ถ้าให้ $z \in C$ จะมี $0 \leq \lambda \leq 1$ ที่ $\lambda z + (1-\lambda) z^* \in C$

ดังนั้น.

$$\|z^*\|^2 \leq \|\underbrace{\lambda z + (1-\lambda) z^*}\|^2 = \|z^*\|^2 + 2\lambda \langle z^*, z - z^* \rangle + \lambda^2 \|z - z^*\|^2$$

$$\langle \lambda z - z^*, z^* \rangle = \|\lambda (z - z^*) + z^*\|^2$$

นั่นคือ.

$$0 \leq 2 \langle z^*, z - z^* \rangle + \underbrace{\underbrace{\lambda}_{\in [0,1]} \underbrace{\|z - z^*\|^2}_{\geq 0}}_{\geq 0}$$

ทำให้ $0 \leq \langle z^*, z - z^* \rangle$ สำหรับทุก $z \in C$.

$$\Rightarrow \quad \|z^*\|^2 = \langle z^*, z^* \rangle \leq \langle z^*, z \rangle$$

$$\Rightarrow \quad \|z^*\| \leq \Big\langle \underbrace{\frac{z^*}{\|z^*\|}}_{= v \to \|v\| = 1}, z \Big\rangle \quad \forall z \in C.$$

ดังนั้นถ้า $v = \dfrac{z^*}{\|z^*\|}$ ที่ $\|v\| = 1$ ทำให้

$$\operatorname{dist}(C, 0) \leq \min_{z \in C} \langle v, z \rangle$$

$$\leq \max_{\|v\|=1} \min_{z \in C} \langle v, z \rangle$$

ดังนั้น $\operatorname{dist}(C, 0) = \max\limits_{\|v\|=1} \min\limits_{z \in C} \langle v, z \rangle$ ∎

บทตั้ง 2: ให้ $X_+, X_- \subset \mathbb{R}^n$, $C_\pm = \operatorname{conv}(X_\pm)$

ที่ $C_+ \cap C_- = \phi$, $C = C_+ - C_- = \operatorname{conv}(X_+ - X_-)$

ให้ Γ เป็น weight polyhedron นั่นคือ

$$\Gamma = \{ (w, \theta) \in \mathbb{R}^{n+1} : y_i(\langle w, x_i \rangle - \theta) > 0 , i = 1, \dots, N \}$$

พิจารณา projection ของ Γ บน $\mathbb{R}^n$ นั่นคือ

$$W = \{ w \in \mathbb{R}^n : \exists \theta \in \mathbb{R} : (w, \theta) \in \Gamma \}$$

สามารถเขียนได้ในรูป.

$$W \overset{\subseteq \,\Rightarrow}{\underset{\supseteq \,\Leftarrow}{=}} \{ w \in \mathbb{R}^n : \langle w, z \rangle > 0 \quad \forall z \in C \}$$

จะได้ว่า ถ้าให้ $w \in W$ แล้ว ให้

$$\theta_w := \frac{1}{2} \left(\min_{x \in C_+} \langle w, x \rangle + \max_{y \in C_-} \langle w, y \rangle \right)$$

ทำให้ $(w, \theta_w) \in \Gamma$ เงื่อนไขเพิ่มเติมคือได้ว่า ถ้าให้ $w \neq \vec{0}$

แล้ว $\rho(w, \theta_w) = \frac{1}{2} \min_{z \in C} \langle \frac{w}{\|w\|}, z \rangle$

proof: จากโครงสร้างของ W จะได้ว่า

"$\Rightarrow$": $w \in W \Leftrightarrow \exists \theta : \begin{cases} \langle w, x \rangle - \theta > 0 & \forall x \in X_+ \\ \langle w, x \rangle - \theta < 0 & \forall x \in X_- \end{cases}$

$[(-1)(\langle w, x \rangle - \theta) > 0]$

$\langle w, -x \rangle + \theta > 0$

ดังนั้น.

$$w \in W \Rightarrow \langle w, z\rangle > 0 \quad \forall\, z = x - y \in X_+ - X_-$$

จะได้ $\langle w, z\rangle > 0$ สำหรับทุก $z \in C = \mathrm{conv}(X_+ - X_-)$

ให้ $z \in C = \mathrm{conv}(X_+ - X_-)$

$\Rightarrow \exists\, x, y \in X_+ - X_-$ และ $\lambda \in [0,1]$

ซึ่ง $z = \lambda x + (1-\lambda) y$

แล้ว. $\langle w, z\rangle = \langle w, \lambda x + (1-\lambda) y\rangle = \lambda \underbrace{\langle w, x\rangle}_{>0} + (1-\lambda)\underbrace{\langle w, y\rangle}_{>0}$

> 0

จะได้ $W \subseteq \{ w \in \mathbb{R}^n : \langle w, z\rangle > 0 \ \forall z \in C\}$

"$\Leftarrow$": ให้ $w \in \mathbb{R}^n : \langle w, z\rangle > 0 \ \forall z \in C$

แล้ว. จาก.

$$\theta_w := \frac{1}{2}\left(\min_{x \in C_+} \langle w, x\rangle + \max_{y \in C_-} \langle w, y\rangle\right)$$

จะได้ว่า สำหรับ $x \in X_+$ แล้ว.

$$\langle w, x\rangle - \theta_w = \underbrace{\langle w, x\rangle}_{\geq \min_{x \in C_+ \supset X_+} \langle w, x\rangle} - \frac{1}{2}\min_{x \in C_+}\langle w, x\rangle - \frac{1}{2}\max_{y \in C_-}\langle w, y\rangle$$

$$(*) \; - \quad \geq \frac{1}{2}\min_{x \in C_+}\langle w, x\rangle \underbrace{- \frac{1}{2}\max_{y \in C_-}\langle w, y\rangle}_{= +\frac{1}{2}\min_{y \in C_-}\langle w, -y\rangle}$$

$$= \frac{1}{2}\min_{x \in C_+,\, y \in C_-}\langle w, \underbrace{x - y}_{z \in C}\rangle > 0 \quad \forall x \in X_+$$

เช่นเดียวกับ $y \in X_-$ แล้ว.

$$\langle w, y\rangle - \theta_w = \underbrace{\langle w, y\rangle}_{< \max\limits_{y\in C_-} \langle w, y\rangle} - \frac{1}{2}\min_{x\in C_+}\langle w, x\rangle - \frac{1}{2}\max_{y\in C_-}\langle w, y\rangle$$

$$(**) \quad \text{—} \quad \le -\frac{1}{2}\min_{x\in C_+}\langle w, x\rangle + \overbrace{\frac{1}{2}\max_{y\in C_-}\langle w, y\rangle}^{-\frac{1}{2}\min\limits_{y\in C_-}\langle w, -y\rangle}$$

$$= -\frac{1}{2}\min_{x\in C_+,\, y\in C_-}\langle w, \overbrace{x-y}^{z\in C}\rangle < 0 \quad \forall y \in X_-$$

ดังนั้น

$$\begin{cases} \langle w, x\rangle - \theta_w > 0 & \forall x \in X_+ \\ \langle w, y\rangle - \theta_w < 0 & \forall y \in X_- \end{cases} \Rightarrow w \in W$$

นั่นคือ

$$\{ w \in \mathbb{R}^n : \langle w, z\rangle > 0 \quad \forall z \in C \} \subseteq W$$

นอกจากนี้ เมื่อพิจารณา. Convex set C_+ แล้วจะมี.
x, y ที่ทำให้เงื่อนไข $(*)$, $(**)$ มีค่า "=" นั่นคือ

$$\min_{x\in X_+}(\langle w, x\rangle - \theta_w) = \min_{y\in X_-}(-\langle w, y\rangle + \theta_w)$$

$$= \frac{1}{2}\min_{z\in C}\langle w, z\rangle$$

ดังนั้น ถ้าเลือก $w \in W$, $(w, \theta_w) \in \Gamma$ และ

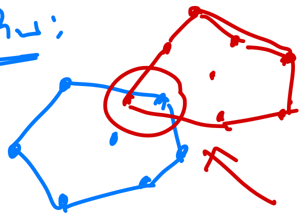
$$\rho(w,\theta_w)\|w\| = \min_{1\le i\le N} y_i\left(\langle w, x_i\rangle - \theta_w\right)$$

$(x_i \in X = X_+ \cup X_-)$

$$= \min\left(\min_{x\in X_+}\left(\langle w,x\rangle - \theta_w\right),\ \min_{y\in X_-}\left(-\langle w,y\rangle + \theta_w\right)\right)$$

$$= \frac{1}{2}\min_{z\in C}\langle w, z\rangle \qquad \blacksquare$$

สรุปผล:



มี 2 sections ที่ตัดกัน.

⇒ ไม่มีการซ้อนรอบ

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

(ไม่มีการซ้อนรอบ ของตัดกันเพราะ ส่วนของการเรียงตรง !

จุดตัด