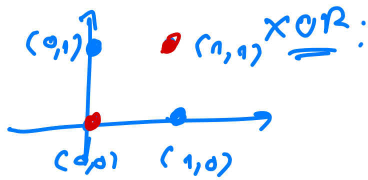
Kernel method for SVM:

ปัญหา: ถ้า X_+, X_- ไม่สามารถแบ่งได้โดย perceptron จะทำอย่างไร?

Idea: Transform X_+, X_- $\subset \mathbb{R}^n$ ไปยัง space H ที่ทำให้ X_+, X_- เป็น affinely separable set นั่นคือ ทำให้ X_+, X_- สามารถแบ่งได้โดย perceptron บน H. แล้ว transform perceptron กลับมาที่ $\mathbb{R}^n$ เพื่อแบ่ง X_+, X_-



Def: ให้ continuous symmetric map

$$k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

จะเรียกว่า possitive kernel ถ้าสำหรับ finite set $\{x_1, \dots, x_N\} \subset \mathbb{R}^n$ เมทริกซ์ที่สร้างโดย

$$K = [k(x_i, x_j)]_{i,j=1,\dots,N}$$

เป็น possitive semi-definite.

$$(\; u^T K u \geq 0 \quad \forall u \in \mathbb{R}^n \;)$$

Thm: ให้ k เป็น possitive kernel แล้วจะมี Hilbert sp H และ map $\Phi : \mathbb{R}^n \to H$ ที่ซึ่ง

$$k(x,y) = \langle \Phi(\underset{\mathbb{R}^n}{x}), \Phi(\underset{\mathbb{R}^n}{y}) \rangle_H$$

สำหรับทุก $x, y \in \mathbb{R}^n$. ■

Hilbert sp: คือ vector space ที่มี inner product $\langle \cdot , \cdot \rangle_H$

โดยมี norm (ระยะห่างระหว่าง elements กับจุด 0) เป็น

$$\|u\|_H^2 = \langle u, u \rangle_H .$$

ที่ complete (ทุก cauchy sequence ลู่เข้าใน H).

และมีระบบมุมตั้งฉาก. คือ. เราเรียก

$$u \perp v \text{ ของ } H \text{ ก็ต่อเมื่อ } \langle u, v \rangle_H = 0$$

โดย Hilbert sp ตามนิยาม. จะสร้างจาก linear combination ของ functional สำหรับ $x \in \mathbb{R}^n$ โดยให้

$$k_x : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad y \longmapsto k_x(y) := k(x,y)$$

โดยที่ ได้ว่า

$$\langle k_x, k_y \rangle = k(x,y) .$$

และถ้าให้ $\Phi(x) = k_x$ จะได้ว่า

$$\langle \Phi(x), \Phi(y) \rangle_H = \langle k_x, k_y \rangle = k(x,y).$$

ตัวอย่าง:

1.) $\boxed{k(x,y) = \langle x, y \rangle^d}$, สำหรับ $d \in \mathbb{N}$, $x, y \in \mathbb{R}^n$.

จะได้

$$k(x,y) = \left(\sum_{i=1}^n x_i y_i \right)^d = \sum_{\alpha \in \mathbb{N}^n, |\alpha| = d} \binom{d}{\alpha} x^\alpha y^\alpha$$

เมื่อ $|\alpha| = \sum_{i=1}^n \alpha_i$, $\binom{d}{\alpha} = \frac{d!}{\alpha_1! \cdots \alpha_n!}$,

$\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$.

ได้. $\boxed{\Phi(x) = \left(\binom{d}{\alpha}^{\frac{1}{2}} x^{\alpha} \right)_{\alpha \in \mathbb{N}^n,\ |\alpha| = d}.}$

<u>Ex:</u> $k(x,y) = \langle x,y \rangle^d$ โดยที่ $n=3,\ d=2$, $x,y \in \mathbb{R}^3$

[illegible] $\Rightarrow$ $k(x,y) = k\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right)$

$$= \left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right\rangle^2$$

$(n=3,\ d=2)$

เขียนใหม่ $k(x,y) = (x_1y_1 + x_2y_2 + x_3y_3)^2$

$$= (x_1y_1 + x_2y_2 + x_3y_3)(x_1y_1 + x_2y_2 + x_3y_3)$$

$$= \left[(x_1y_1)^2 + (x_1y_1)(x_2y_2) + (x_1y_1)(x_3y_3) \right]$$

$$+ \left[(x_2y_2)(x_1y_1) + (x_2y_2)^2 + (x_2y_2)(x_3y_3) \right]$$

$$+ \left[(x_3y_3)(x_1y_1) + (x_3y_3)(x_2y_2) + (x_3y_3)^2 \right]$$

$$= (x_1y_1)^2 + (x_2y_2)^2 + (x_3y_3)^2$$

$\boxed{x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}$

$$+ 2(x_1y_1)(x_2y_2) + 2(x_1y_1)(x_2y_2)$$

$$+ 2(x_2y_2)(x_3y_3)$$

$$k(x,y) = x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2$$

$$+ 2(x_1x_2)(y_1y_2) + 2(x_1x_2)(y_1y_2) + 2(x_2x_3)(y_2y_3)$$

หรือลงท

$$k(x,y) = \left(\sum_{i=1}^{n} x_i y_i\right)^d = \boxed{\sum_{\alpha \in \mathbb{N}^n, |\alpha| = d} \binom{d}{\alpha} x^\alpha y^\alpha} \checkmark$$

ในตัวอย่างนี้ $\alpha \in \mathbb{N}^3$ ที่ $|\alpha| = 2$.

อาทิ. $\alpha = \underset{\alpha_1}{(2,0,0)}, \underset{\alpha_2}{(0,2,0)}, \underset{\alpha_3}{(0,0,2)}$
$\underset{\alpha_4}{(1,1,0)}, \underset{\alpha_5}{(1,0,1)}, \underset{\alpha_6}{(0,1,1)}$

สำหรับเขียนได้ในรูป.

$$k(x,y) = \sum_{\alpha \in \mathbb{N}^3, |\alpha|=2} \binom{d}{\alpha} x^\alpha y^\alpha = \sum_{i=1}^{6} \binom{d}{\alpha_i} x^{\alpha_i} y^{\alpha_i}$$

$(d=2)$

$$= \binom{2}{\alpha_1} x^{\alpha_1} y^{\alpha_1} + \binom{2}{\alpha_2} x^{\alpha_2} y^{\alpha_2} + \binom{2}{\alpha_3} x^{\alpha_3} y^{\alpha_3}$$
$$+ \binom{2}{\alpha_4} x^{\alpha_4} y^{\alpha_4} + \binom{2}{\alpha_5} x^{\alpha_5} y^{\alpha_5} + \binom{2}{\alpha_6} x^{\alpha_6} y^{\alpha_6}$$

$\left(\binom{d}{\alpha} = \frac{d!}{\alpha_1! \cdots \alpha_n!}\right)$

$x^\alpha = x_i^{\alpha_i} \cdots$

$$= \frac{2!}{2!} x_1^2 y_1^2 + \frac{2!}{2!} x_2^2 y_2^2 + \frac{2!}{2!} x_3^2 y_3^2$$
$$+ \frac{2!}{1!\,1!} (x_1 x_2)(y_1 y_2) + \frac{2!}{1!\,1!} (x_1 x_3)(y_1 y_3)$$
$$+ \frac{2!}{1!\,1!} (x_2 x_3)(y_2 y_3)$$

$$k(x,y) = x_1^2 y_1^2 + x_2^2 y_2^2 + x_3^2 y_3^2 + 2(x_1 x_2)(y_1 y_2)$$
$$+ 2(x_1 x_3)(y_1 y_3) + 2(x_2 x_3)(y_2 y_3) \checkmark$$

และ
สำหรับ.

$$\phi(x) = \left(\binom{2}{\alpha_i}^{\frac{1}{2}} x^{\alpha_i}\right)_{i=1,\dots,6}.$$

$$= \binom{2}{\alpha_1}^{\frac{1}{2}} x^{\alpha_1} + \binom{2}{\alpha_2}^{\frac{1}{2}} x^{\alpha_2} + \binom{2}{\alpha_3}^{\frac{1}{2}} x^{\alpha_3}$$

$$+ \binom{2}{\alpha_4}^{\frac{1}{2}} x^{\alpha_4} + \binom{2}{\alpha_5}^{\frac{1}{2}} x^{\alpha_5} + \binom{2}{\alpha_6}^{\frac{1}{2}} x^{\alpha_6}$$

$\big(\alpha_1 = (2,0,0),\ \alpha_2 = (0,2,0),\ \alpha_3 = (0,0,2),\ \alpha_4 = (1,1,0),\ \alpha_5 = (1,0,1),$
$\alpha_6 = (0,1,1)\big)$

$$= \left(\frac{2!}{2!}\right)^{\frac{1}{2}} x_1^2 + \left(\frac{2!}{2!}\right)^{\frac{1}{2}} x_2^2 + \left(\frac{2!}{2!}\right)^{\frac{1}{2}} x_3^2$$

$$+ \left(\frac{2!}{1!1!}\right)^{\frac{1}{2}} x_1 x_2 + \left(\frac{2!}{1!1!}\right)^{\frac{1}{2}} x_1 x_3 + \left(\frac{2!}{1!1!}\right)^{\frac{1}{2}} x_2 x_3$$

$$\Rightarrow \phi(\overset{\in \mathbb{R}^3}{x}) = x_1^2 + x_2^2 + x_3^2 + \sqrt{2}\, x_1 x_2 + \sqrt{2}\, x_1 x_3 + \sqrt{2}\, x_2 x_3$$

(monomial.)

จะเห็นได้ว่า $\phi(x)$ เป็น พหุนามที่มีตัวแปรของแต่ละพจน์เป็น 1 หรือ
ต่างที่มีสัมประสิทธิ์มาก ๆ และแต่ละเทอมมีดีกรีรวมเป็น $d = 2$.
และจะเห็นได้ว่า Hilbert space H ของกรณีนี้ เป็น $\mathbb{R}^6$.
ดังนั้น $\langle \cdot , \cdot \rangle_H$ ก็คือ scalar product ของ $\mathbb{R}^6$
(Euclidean norm.)

2.) $k(x,y) = (1 + \langle x, y \rangle)^d$ สำหรับ $d \in \mathbb{N}$

จะได้

$$k(x,y) = \left(1 + \sum_{i=1}^{n} x_i y_i\right)^d = \sum_{\alpha \in \mathbb{N}^n,\, |\alpha| \le d} \frac{1}{(d-|\alpha|)!} \binom{d}{\alpha} x^\alpha y^\alpha.$$

และ ϕ.

$$\Phi(x) = \left(\left(\frac{1}{(d-|\alpha|)!} \binom{d}{\alpha} \right)^{\frac{1}{2}} x^{\alpha} \right)_{\alpha \in \mathbb{N}^n,\ |\alpha| \le d}.$$

จะได้ H เป็น finite dimension. ซึ่ง

$\Phi(x)$ เป็น โพลิโนเมียล ที่แต่ละเทอม มีดีกรีรวม มากสุดเท่ากับ d.

3.) $k(x,y) = \tanh(\alpha \langle x, y \rangle + \beta)$

สำหรับ α, β ที่เหมาะสม.

4.) $k(x,y) = \exp(-\|x-y\|^2 / 2\sigma^2)$

สำหรับ บางค่า α.

H เป็น inf. dim.

- - - - - - - - -

ด้วยกระบวนการนี้ เราไม่จำเป็นต้องสมมติว่า

X_+, X_- เป็น affinely separable sets.

สำหรับ SVM. เราจะเลือก $k(x,y)$ ที่ $\phi : \mathbb{R}^n \to H$

สามารถแบ่ง X_+, X_- ใน H ได้.

โดยปัญหา SVM. เดิม เป็น หาค่า. λ^* ที่

$$\begin{cases} \text{maximize } \tilde{F}(\lambda) = -\frac{1}{2} \sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle_H \\ \qquad\qquad + \sum_{i=1}^{N} \lambda_i \\ \text{subject to } \sum_{i=1}^{N} \lambda_i y_i = 0 \quad \text{and } \lambda_1, \dots, \lambda_N \ge 0 \end{cases}$$

ซึ่งสามารถเขียนได้ในรูป

$$\tilde{F}(\lambda) = -\frac{1}{2} z^T K z + \sum_{i=1}^{N} \lambda_i$$

โดยที่ $K \in \mathbb{R}^{N \times N}$, $z \in \mathbb{R}^N$ ที่

$$K_{ij} = k(x_i, x_j) \quad \text{และ} \quad z_i = \lambda_i y_i$$

ถ้าเราหาคำตอบ $\lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_N^*)$ จะได้

$$\omega^* = \sum_{i=1}^{N} \lambda_i^* y_i \Phi(x_i) \in H$$

และ

$$\theta^* = \frac{1}{2}\left(\min_{x \in X_+} \sum_{i=1}^{N} \lambda_i^* y_i k(x_i, x) + \max_{y \in X_-} \sum_{i=1}^{N} \lambda_i^* y_i k(x_i, y)\right) \in \mathbb{R}$$

จะได้ optimal separating hyperplane ของ H เป็น

$$H^* = \{ y \in H \mid \langle \omega^*, y \rangle_H = \theta^* \}$$

ถ้าเรา $x \in X_+$ แล้ว $\Phi(x) \in \Phi(X_+)$ ดังนั้น

$$\langle \omega^*, \Phi(x) \rangle_H = \left\langle \sum_{i=1}^{N} \lambda_i^* y_i \Phi(x_i), \Phi(x) \right\rangle_H$$

$$= \sum_{i=1}^{N} \lambda_i^* y_i \underbrace{\langle \Phi(x_i), \Phi(x) \rangle_H}_{= k(x_i, x)} > \theta^*$$

ในทำนองเดียวกัน $y \in X_-$

$$\langle w^*, \Phi(y)\rangle_H = \sum_{i=1}^{N} \lambda_i^* y_i \underbrace{\langle \Phi(x_i), \Phi(y)\rangle}_{k(x_i,y)} < \theta^*$$

ดังนั้น เมื่อพิจารณา $X_+, X_- \subset \mathbb{R}^N$ แล้ว

จะได้ว่า $f(X_\pm) = \pm 1$ โดย

$$f(x) = \mathrm{sign}\Big(\underbrace{\sum_{i=1}^{N} \lambda_i^* y_i k(x_i,x) - \theta^*}_{s(x) \text{ activation fn}}\Big)$$

Transformed perceptron บน $\mathbb{R}^N$

Remark: ในการใช้งาน จะเห็นได้ว่า ในการคำนวณ

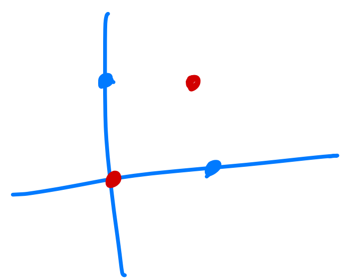
λ^*, θ^* และ activation fn $s(x)$

ใช้เพียงแค่ $k(x,y)$ เท่านั้น ไม่จำเป็นต้องรู้ข้อมูล $\Phi(x)$

Ex: (positive kernel กับ XOR operator.)

พิจารณา $f: \{0,1\}^2 \to \{\pm 1\}$ ที่แทน XOR operator

	XOR	
$x_1 = (0,0)$	$y_1 = -1$	(0)
$x_2 = (0,1)$	$y_2 = 1$	(1)
$x_3 = (1,0)$	$y_3 = 1$	(1)
$x_4 = (1,1)$	$y_4 = -1$	(0)



ให้ $\boxed{k(x,y) = \langle x, y\rangle^2}$ กับ $x, y \in \mathbb{R}^2$ จะได้

$$\Rightarrow K = \left[k(x_i, x_j)\right]_{i,j=1,\dots,4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 4 \end{bmatrix}$$

$$\left(\begin{array}{l} k(x_1,x_1) = \left\langle \binom{0}{0}, \binom{0}{0} \right\rangle^2 = 0 \\ k(x_2,x_2) = \left\langle \binom{0}{1}, \binom{0}{1} \right\rangle^2 = 1 \end{array}\right)$$

$$k(x_2,x_3) = \left\langle \binom{0}{1}, \binom{1}{0} \right\rangle^2 = 0$$

$$\vdots$$

$$k(x_4,x_4) = \left\langle \binom{1}{1}, \binom{1}{1} \right\rangle^2 = 4$$

$$\Rightarrow z = \begin{bmatrix} y_1\lambda_1 \\ y_2\lambda_2 \\ y_3\lambda_3 \\ y_4\lambda_4 \end{bmatrix} = \begin{bmatrix} -\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ -\lambda_4 \end{bmatrix}$$

$$\Rightarrow \tilde{F}(\lambda) = -\frac{1}{2} z^T K z + \sum_{i=1}^{4} \lambda_i$$

$$= -\frac{1}{2} \begin{bmatrix} -\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} -\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ -\lambda_4 \end{bmatrix} + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$$

$$= -\frac{1}{2}\left[\lambda_2^2 + \lambda_3^2 + 4\lambda_4^2 - 2\lambda_2\lambda_4 - 2\lambda_3\lambda_4\right] + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$$

$$\tilde{F}(\lambda) = -\frac{1}{2}(\lambda_2^2 + \lambda_3^2) - 2\lambda_4^2 + \lambda_2\lambda_4 + \lambda_3\lambda_4 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$$

ภายใต้เงื่อนไข $\sum_{i=1}^{4} y_i \lambda_i = 0 \Rightarrow -\lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 = 0$

$$\Rightarrow \boxed{\lambda_1 = \lambda_2 + \lambda_3 - \lambda_4}$$

แทนใน $\hat{F}(\lambda)$ จะได้

$$\hat{F}(\lambda_2,\lambda_3,\lambda_4) = -\frac{1}{2}(\lambda_2^2+\lambda_3^2) - 2\lambda_4^2 + \lambda_2\lambda_4 + \lambda_3\lambda_4 + 2\lambda_2 + 2\lambda_3$$

เงื่อนไข $\nabla\hat{F} = \vec{0}$

$$\Rightarrow \nabla\hat{F} = \begin{pmatrix} -\lambda_2 + \lambda_4 + 2 \\ -\lambda_3 + \lambda_4 + 2 \\ -4\lambda_4 + \lambda_2 + \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{matrix} — ① \\ — ② \\ — ③ \end{matrix}$$

$$\left.\begin{aligned} ① &\Rightarrow \lambda_2 = \lambda_4 + 2 \\ ② &\Rightarrow \lambda_3 = \lambda_4 + 2 \end{aligned}\right\} \Rightarrow \lambda_2 = \lambda_3 = \lambda_4 + 2$$

แทนใน ③ $\Rightarrow -4\lambda_4 + (\lambda_4+2) + (\lambda_4+2) = 0$

$$\Rightarrow \lambda_4 = \frac{4}{2} = 2$$

จะได้ $\lambda_2 = \lambda_3 = \lambda_4 + 2 = 2+2 = 4$

$$\lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 = 4+4-2 = 6$$

$\therefore \lambda^* = (6, 4, 4, 2)$

ทบทวน: ได้กำหนดเองว่า

$$\vec{\phi}(x) = (x_1^2, x_2^2, \sqrt{2}\,x_1x_2)^T$$

หาค่า w^* และ θ^*

จาก. $$w^* = \sum_{i=1}^{4} \lambda_i^* y_i \vec{\phi}(x_i)$$

$$= 6(-1)\vec{\phi}\left(\begin{pmatrix}0\\0\end{pmatrix}\right) + 4(+1)\vec{\phi}\left(\begin{pmatrix}0\\1\end{pmatrix}\right)$$

$$+\,4(+1)\,\Phi\left(\begin{pmatrix}1\\0\end{pmatrix}\right) + 2(-1)\,\Phi\left(\begin{pmatrix}1\\1\end{pmatrix}\right)$$

$$= \begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}0\\4\\0\end{pmatrix} + \begin{pmatrix}4\\0\\0\end{pmatrix} - 2\begin{pmatrix}1\\1\\\sqrt{2}\end{pmatrix}$$

$$w^* = 2\begin{pmatrix}1\\1\\-\sqrt{2}\end{pmatrix} \in \mathbb{R}^3$$

ในกรณีนี้เลือกค่า $\theta^* = 1$

$\boxed{\text{หมายเหตุ: ในกรณีของเรา } \theta^* = 1.}$

ดังนั้น $H^* = \left\{ y \in \mathbb{R}^3 \;\middle|\; \left\langle \underbrace{2\begin{pmatrix}1\\1\\-\sqrt{2}\end{pmatrix}}_{w^*}, \begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix} \right\rangle = 1 \right.$

$$= \left\{ y \in \mathbb{R}^3 \;\middle|\; \underbrace{2(y_1 + y_2 - \sqrt{2}\,y_3) - 1}_{S(y_1, y_2, y_3)} = 0 \right\}$$

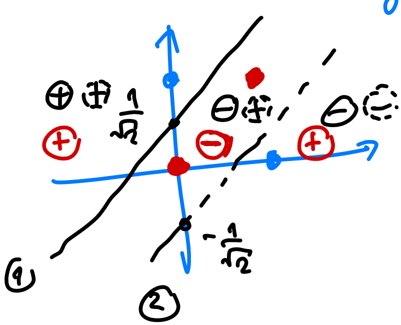
จาก $\Phi\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = \begin{pmatrix}x_1^2\\x_2^2\\\sqrt{2}\,x_1x_2\end{pmatrix} \begin{matrix}\leftarrow y_1\\\leftarrow y_2\\\leftarrow y_3\end{matrix}$

ดังนั้น

$$f(x) = \operatorname{sign}\Big(2\big(x_1^2 + x_2^2 - \overbrace{\sqrt{2}(\sqrt{2}\,x_1x_2)}^{2x_1x_2}\big) - 1 \Big)$$

แยกตัวประกอบ

$$= \operatorname{sign}\Big(\underbrace{\left(x_2 - x_1 - \tfrac{1}{\sqrt{2}}\right)}_{①} \underbrace{\left(x_2 - x_1 + \tfrac{1}{\sqrt{2}}\right)}_{②} \Big)$$



Thm: ถ้าเรา. Boolean fn $f : \{0,1\}^n \to \{\pm 1\}$

จะมี polynomial ที่มีแต่ละเทอมดีกรีรวมเป็น n. นั่นคือ

$$p(x) = \sum_{\alpha \in \mathbb{N}^n, |\alpha| = n} a_\alpha x^\alpha \quad : \mathbb{R}^n \to \mathbb{R}$$

โดยที่ $x \in \mathbb{R}^n$, $a_\alpha \in \mathbb{R}$ ที่ซึ่ง

$$f(x) = \text{sign}(p(x) - \theta) \quad \text{สำหรับทุก } x \in \{0,1\}^n$$

(พิสูจน์ ใช้ข้อสังเกตแบบนี้: $k(x,y) = \langle x, y \rangle^n$

แยก X_+, X_- ได้ ที่ $X_\pm \subset \{0,1\}^n$

และ $X_+ \cap X_- = \emptyset$

ทำได้เสมอ) ∎