Linear Algebra: Linear geometry and linear classifier

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}$$

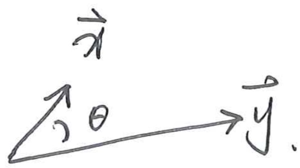
空間　　向量.

Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

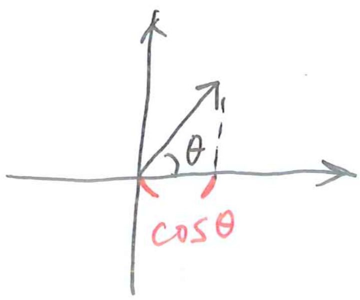
Then $\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$.

Fact: $|\vec{x}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\vec{x} \cdot \vec{x}}$.

Fact: $\vec{x} \cdot \vec{y} = |\vec{x}| \cdot |\vec{y}| \cdot \cos\theta$



θ	0	45°	90°	135°	180°
$\cos\theta$	1	$\frac{1}{\sqrt{2}}$	0	$-\frac{1}{\sqrt{2}}$	-1



垂直 $\Leftrightarrow$ $\vec{x} \cdot \vec{y} = 0$.

Fact: $(\vec{x} \pm \vec{y}) \cdot (\vec{x} \pm \vec{y}) = |\vec{x} \pm \vec{y}|^2$

$= |\vec{x}|^2 \pm 2\vec{x}\cdot\vec{y} + |\vec{y}|^2$

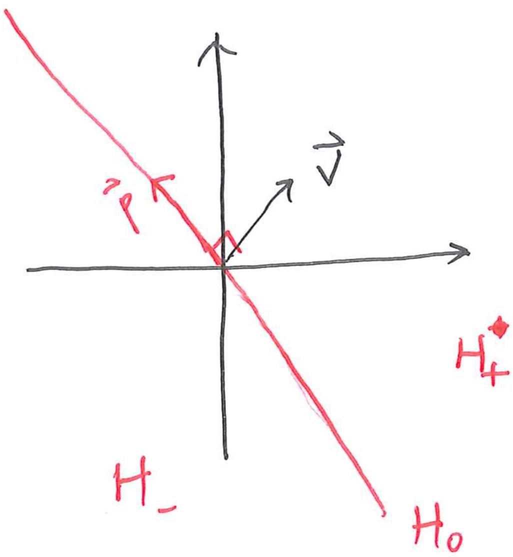
Fact: a normal vector $\vec{v}$ defines a "hyperplane".

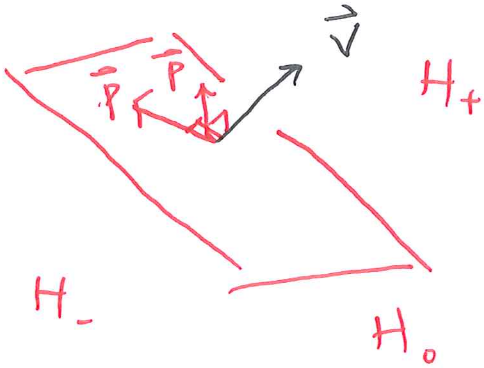
$\mathbb{R}^n$ 中的 $\mathbb{R}^{n-1}$

$$H_0 = \{ \vec{p} \in \mathbb{R}^n : \vec{p} \cdot \vec{v} = 0 \}.$$

Exmp. $\vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{R}^2$.

Exmp: $\vec{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in \mathbb{R}^3$





Fact: H_0 is a hyperplane and it passes through the origin.

$\boxed{\{\vec{p} \in \mathbb{R}^n : \vec{p} \cdot \vec{v} = d\}$ 是平移過的平面. d 愈大, 離原點愈遠$}$

Fact: H_0 cuts $\mathbb{R}^n$ into two "halfspace".

$$H_+ = \{ \vec{p} \in \mathbb{R}^n : \vec{p} \cdot \vec{v} > 0 \}$$

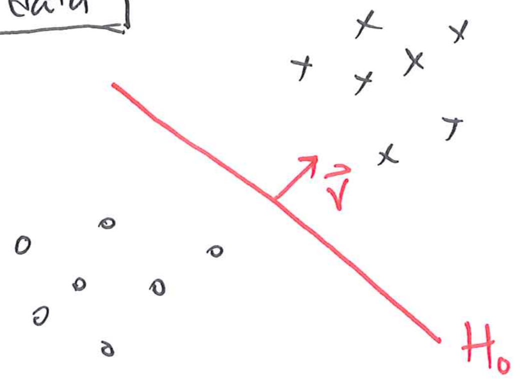
← 跟 $\vec{v}$ 同側

$$H_- = \{ \vec{p} \in \mathbb{R}^n : \vec{p} \cdot \vec{v} < 0 \}$$

← 跟 $\vec{v}$ 相反邊

Linear classifier (supervised).

兩類資料 labeled data $\longrightarrow \vec{v} \longrightarrow$ 預測新資料的類別



$$\begin{cases} \vec{p}\cdot\vec{v} > 0 \longrightarrow \times \\ \vec{p}\cdot\vec{v} < 0 \longrightarrow \circ \end{cases}$$

Math model:

Labeled data:

data: $\vec{x}_1, \cdots, \vec{x}_N \in \mathbb{R}^d$

label: $y_1, \cdots, y_N \in \{1, -1\}$.

$\vec{v} \in \mathbb{R}^d$ is a linear classifier

線性分類器、感知器 (perceptron)

if $\begin{cases} \vec{x}_i \cdot \vec{v} > 0 \Longleftrightarrow y_i = 1 \\ \vec{x}_i \cdot \vec{v} < 0 \Longleftrightarrow y_i = -1 \end{cases}$ for all i.

Prediction

When new data $\vec{p}$ comes ~~that~~ in,

$$\begin{cases} \text{if } \vec{p}\cdot\vec{v} > 0 \Rightarrow \text{guess } \vec{p} \text{ in group } 1. \\ \text{if } \vec{p}\cdot\vec{v} < 0 \Rightarrow \text{guess } \vec{p} \text{ in group } -1 \end{cases}$$

How to find a linear classifier?

If $\vec{v}$ is not a classifier,

there is ① $\vec{x}_i$ such that $\vec{x}_i \cdot \vec{v} > 0$ but $y_i < 0$.

or

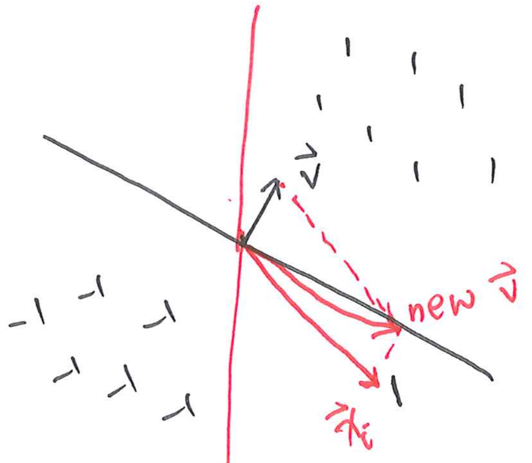
② $\vec{x}_i$ such that $\vec{x}_i \cdot \vec{v} < 0$ but $y_i > 0$.

$\Leftrightarrow (x_i \cdot \vec{v}) \times y_i < 0$

If ①, $\vec{v}$ 太靠近 $x_i \Rightarrow \vec{v} \leftarrow \vec{v} - \vec{x}_i$

If ②, $\vec{v}$ 離 $\vec{x}_i$ 太遠 $\Rightarrow \vec{v} \leftarrow \vec{v} + \vec{x}_i$

$\vec{v} \leftarrow \vec{v} + y_i \cdot \vec{x}_i \Leftrightarrow$



Algorithm [Perceptron Learning Algorithm (PLA)].

Input: Data: $\vec{x}_1, \cdots, \vec{x}_N \in \mathbb{R}^d$

Labels: $y_1, \cdots, y_N \in \{1, -1\}$

Output: a classifier (if exists)

① $\vec{v}^{(0)} = \vec{0}$

② If $\vec{v}^{(k)}$ is not a classifier because $(\vec{x}_i \cdot \vec{v}) \times y_i < 0$ for some i,

then $\vec{v}^{(k+1)} = \vec{v}^{(k)} + y_i \vec{x}_i$

③ Repeat ② until $\vec{v}^{(k)}$ becomes a classifier.

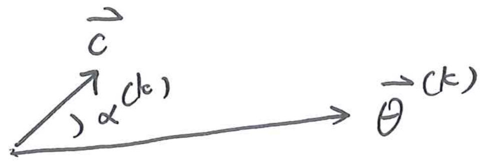
Thm (Convergence of PLA).

If a set of labeled data has a classifier $\vec{c}$, then PLA can find a classifier in finite steps.

Parameters:

$$R = \max_i |x_i| \quad , \quad \gamma = \min_i |\vec{c} \cdot \vec{x}_i|$$

pf: We will show



ⓐ $\vec{v}^{(k)} \cdot \vec{c} \geq k\gamma$

ⓑ $\cancel{\theta}\, |\vec{v}^{(k)}|^2 \leq kR^2$

If so, then

$$\underbrace{\cos \alpha^{(k)}}_{\text{介在 }-1 \sim 1\text{ 之間}} = \frac{\vec{c} \cdot \vec{v}^{(k)}}{|\vec{c}| \cdot |\vec{v}^{(k)}|} \geq \cancel{\frac{k\gamma}{\ }}\ \cancel{\frac{k\gamma}{|\vec{c}| \cdot \theta}}\ \frac{k\gamma}{|\vec{c}| \cdot \sqrt{kR^2}}$$

$$= \sqrt{k} \cdot \frac{\gamma}{|\vec{c}| \cdot R} \longrightarrow \infty$$

不可能 k 一直變大.

Claim ⓐ:

$$\vec{v}^{(k)} \cdot \vec{c} = (\vec{v}^{(k-1)} + y_i \vec{x}_i) \cdot \vec{c}$$

$$= \vec{v}^{(k-1)} \cdot \vec{c} + y_i (\vec{x}_i \cdot \vec{c}) \geq \vec{v}^{(k-1)} \cdot \vec{c} + \gamma$$

y_i: $+$ or $-$; $(\vec{x}_i \cdot \vec{c})$: $+$ or $-$

$$\vec{v}^{(0)} \cdot \vec{c} = \vec{0} \Rightarrow \vec{v}^{(k)} \cdot \vec{c} \geq k\gamma.$$

Claim ⓑ:

$$|\vec{v}^{(k)}|^2 = |\vec{v}^{(k-1)} + y_i \vec{x}_i|^2 \leq |\vec{v}^{(k-1)}|^2 + R^2$$

$$= |\vec{v}^{(k-1)}|^2 + \underbrace{2 y_i \vec{v}^{(k-1)} \cdot \vec{x}_i}_{\leq 0} + \underbrace{|\vec{x}_i|^2}_{\leq R^2}$$

y_i: $+$ or $-$; $\vec{v}^{(k-1)} \cdot \vec{x}_i$: $-$ or $+$

$$|\vec{v}^{(0)}|^2 = 0 \Rightarrow |\vec{v}^{(k)}|^2 \leq kR^2$$