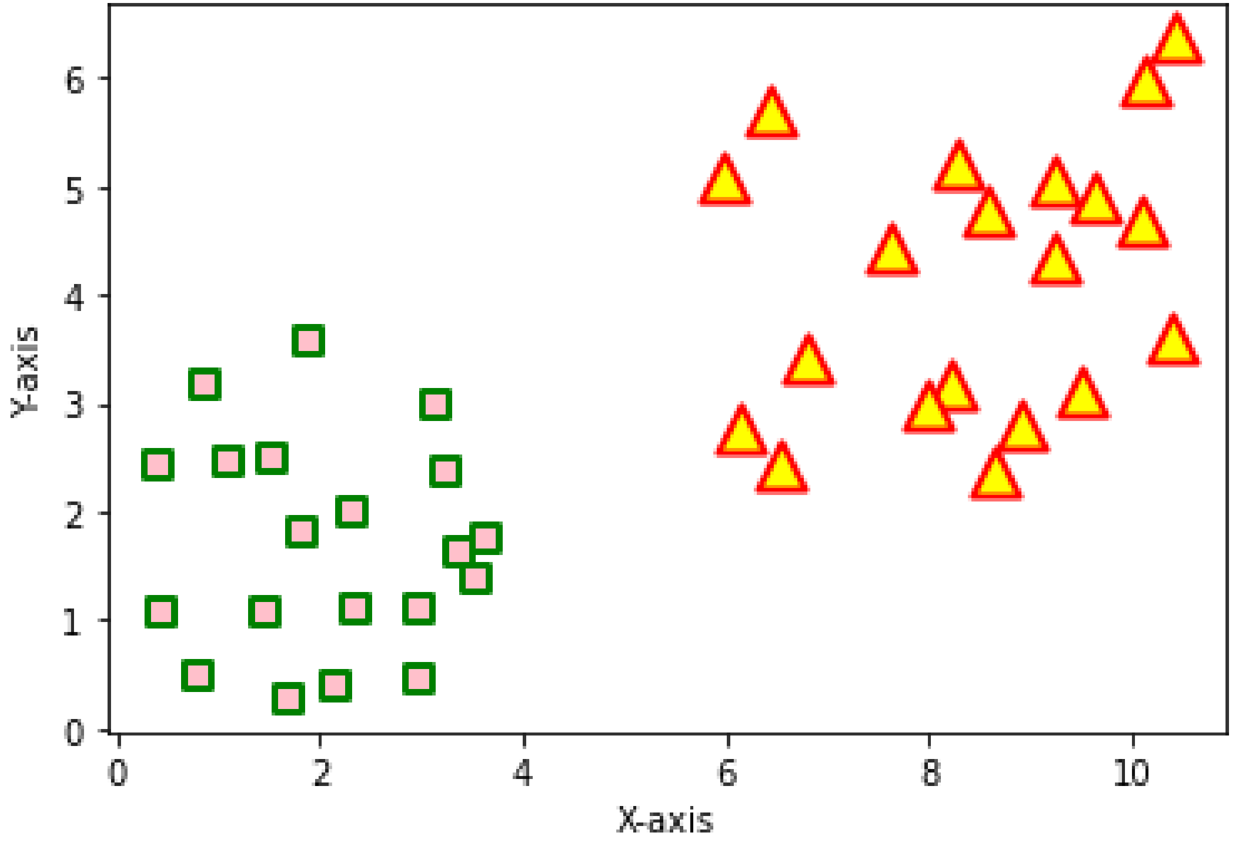
# ML 0106

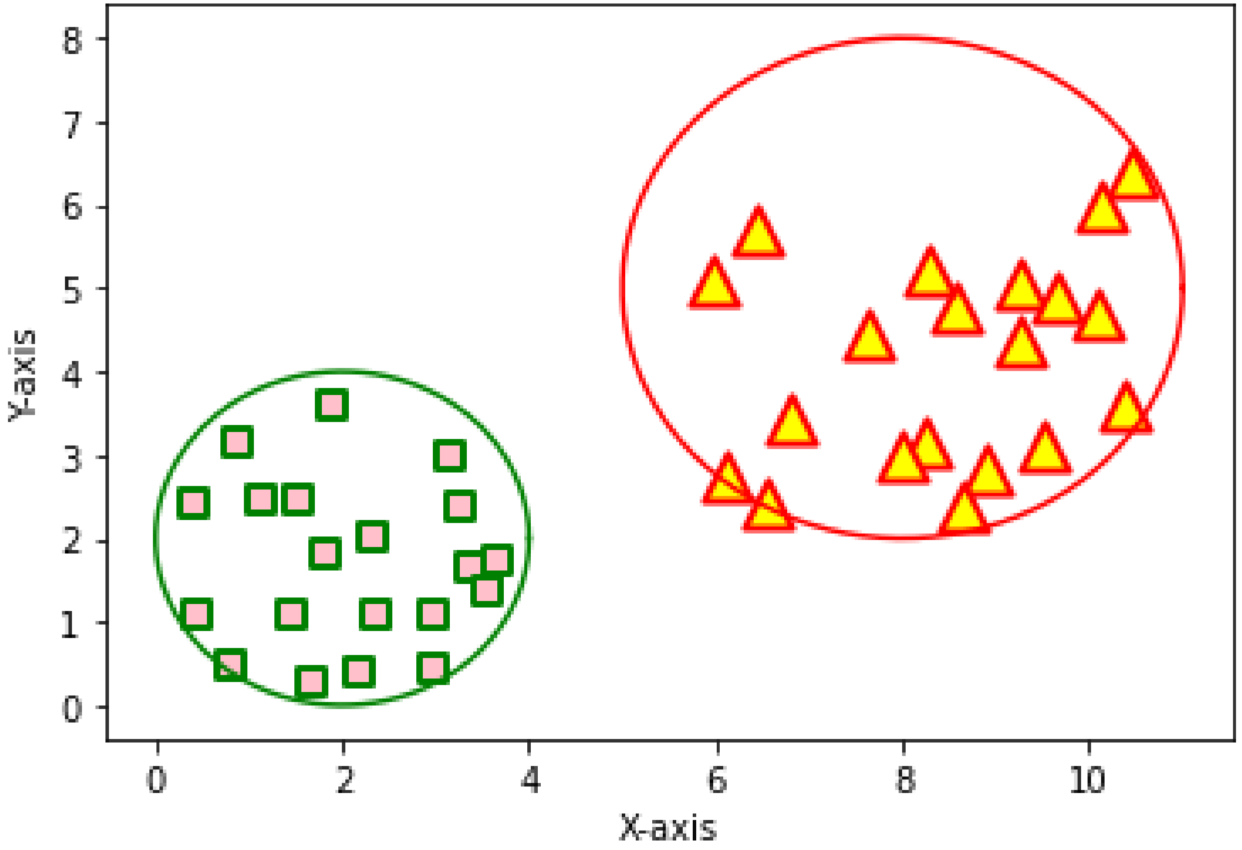
方麒豪

# Introduction

• 假設我們手中有二分類資料並且已有label，我們進行視覺化：



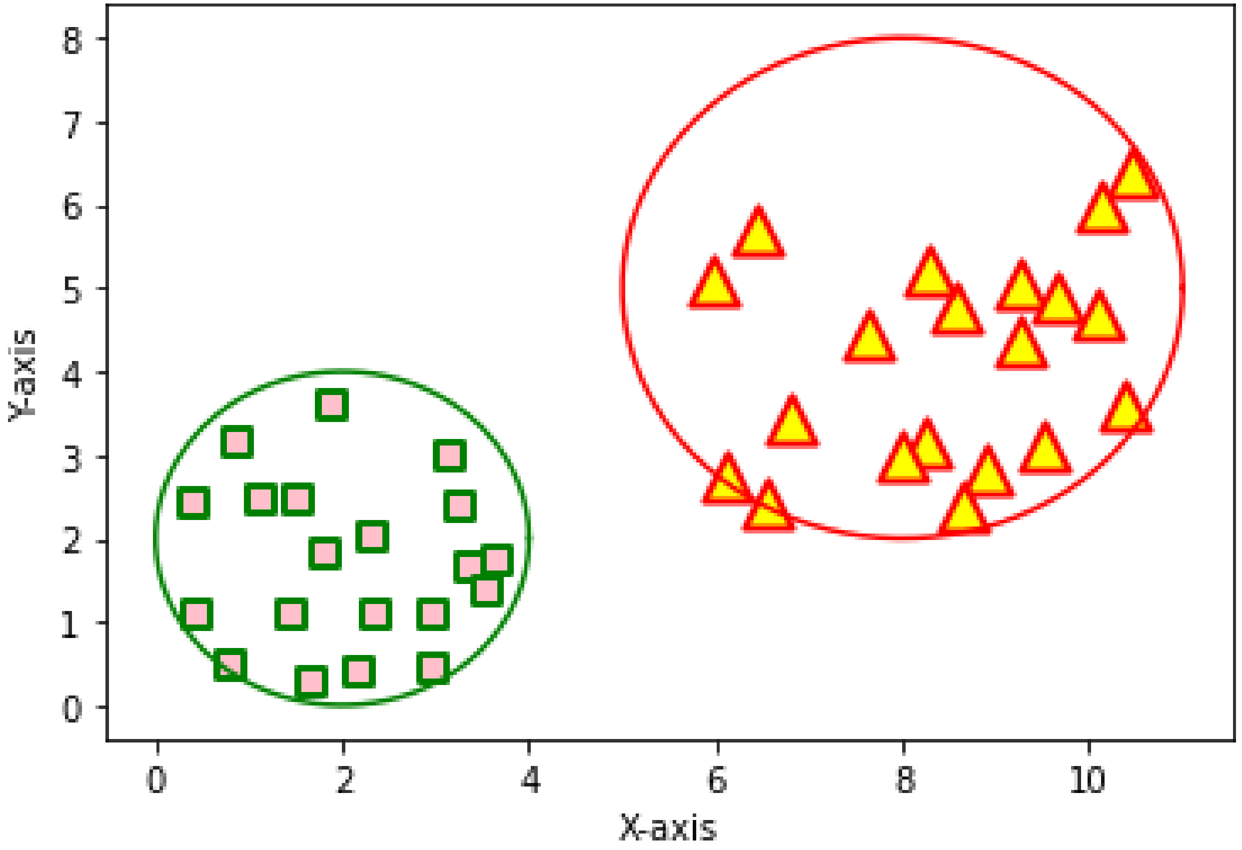
如果我們用貝氏統計來處理，則我們會希望得到資料的類別是什麼分配，其中一種可能性如下圖




我們會去試著得出：

1. 兩個類別可能都長成圓形
2. 這兩個圓形的圓心，半徑等等

如果我們用貝氏統計來處理，則我們會希望得到資料的類別是什麼分配，其中一種可能性如下圖




我們會去試著得出：

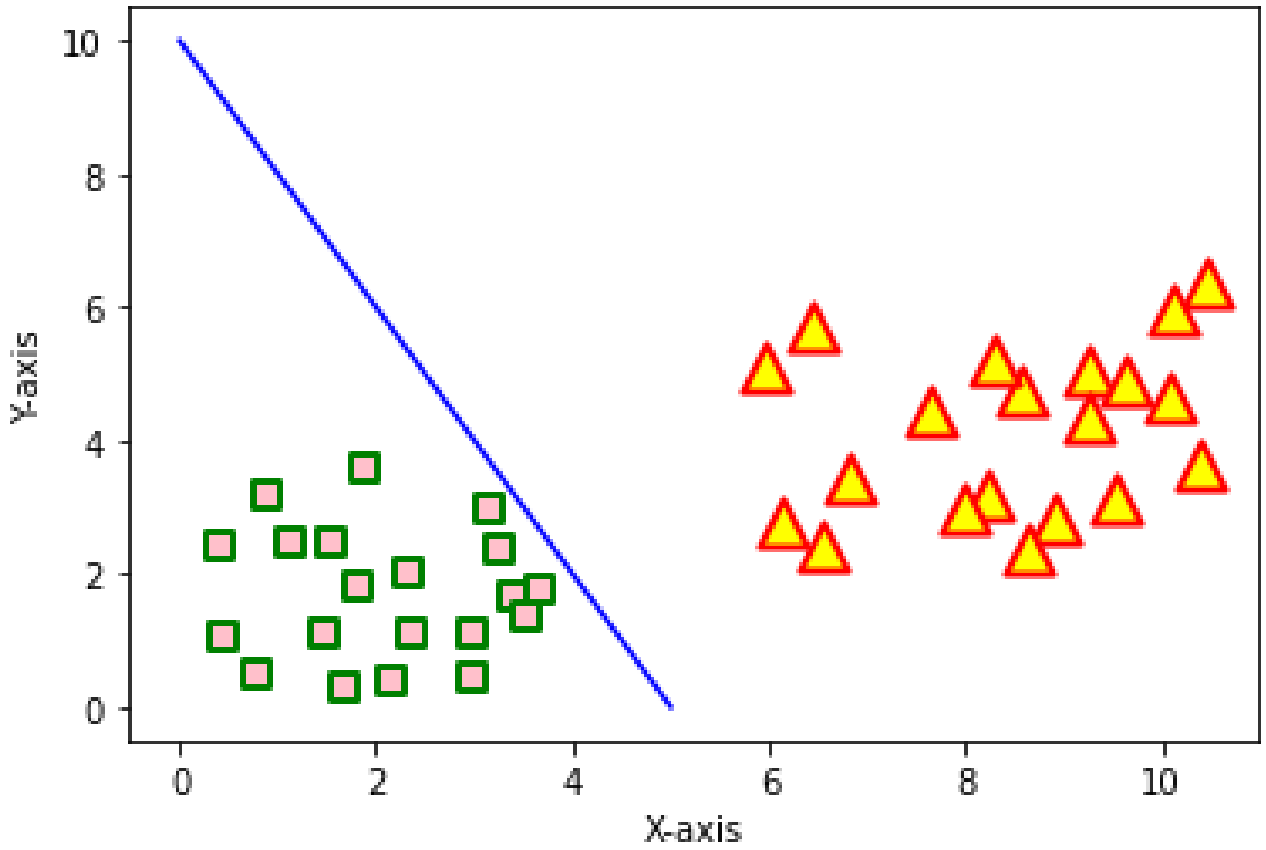
1. 兩個類別可能都長成圓形

2. 這兩個圓形的圓心，半徑等等

我們用此方法去猜測資料的規則或是完整的長相

問：真的需要做到如此嗎?

# 如果我們這樣想




藍線左邊一類、右邊一類。

左邊的資料長成怎樣不管
右邊的資料長成怎樣也不管

我們用此方法就可以只記住這條線就好

# 如果我們這樣想

事情如果能簡單就簡單：
如果我們能夠知道資料類別的分界，就不需要知道資料的類別分配。

註：前提是資料類別能夠劃得出分界

# 問題轉化

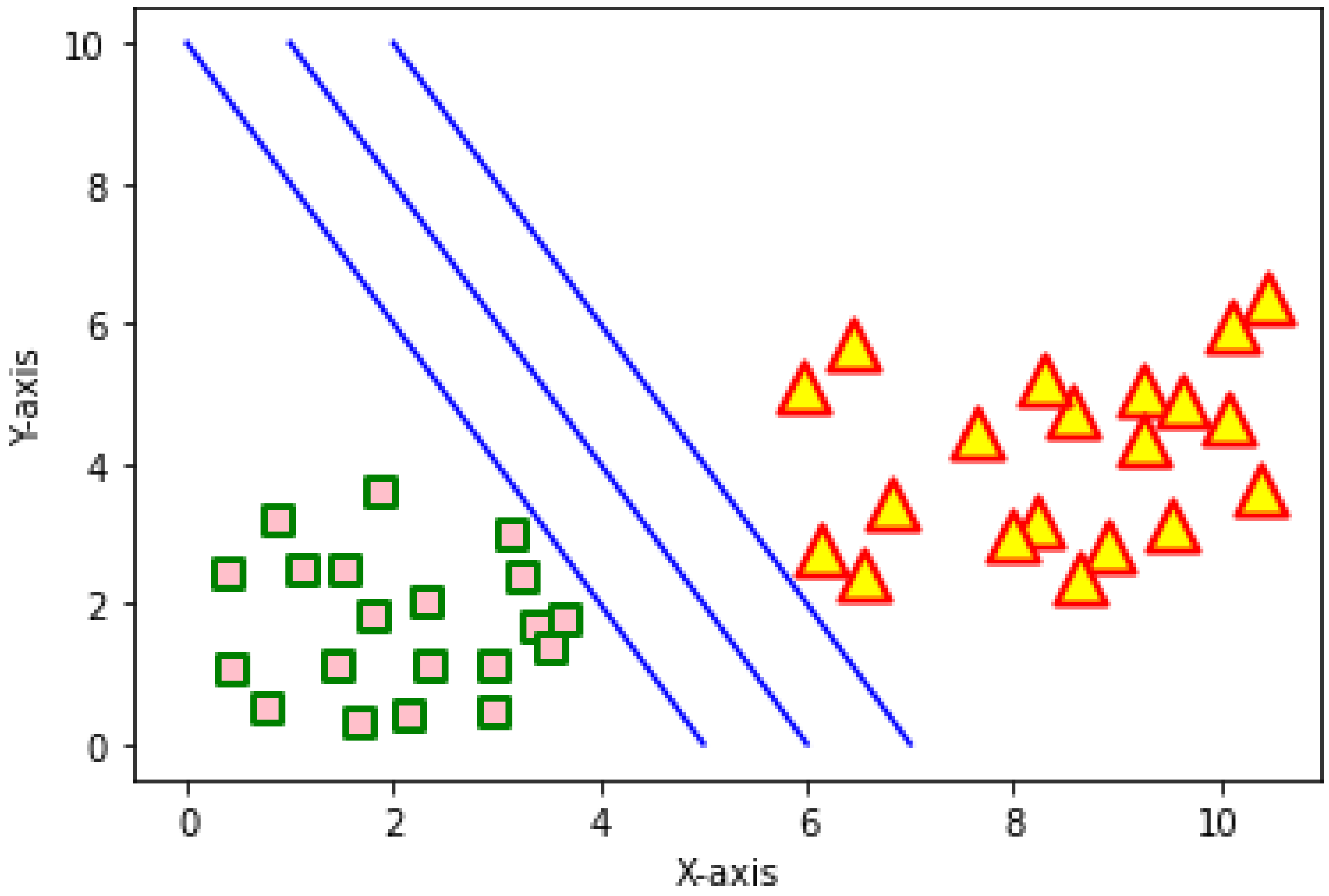
統計方法

學習資料的分配

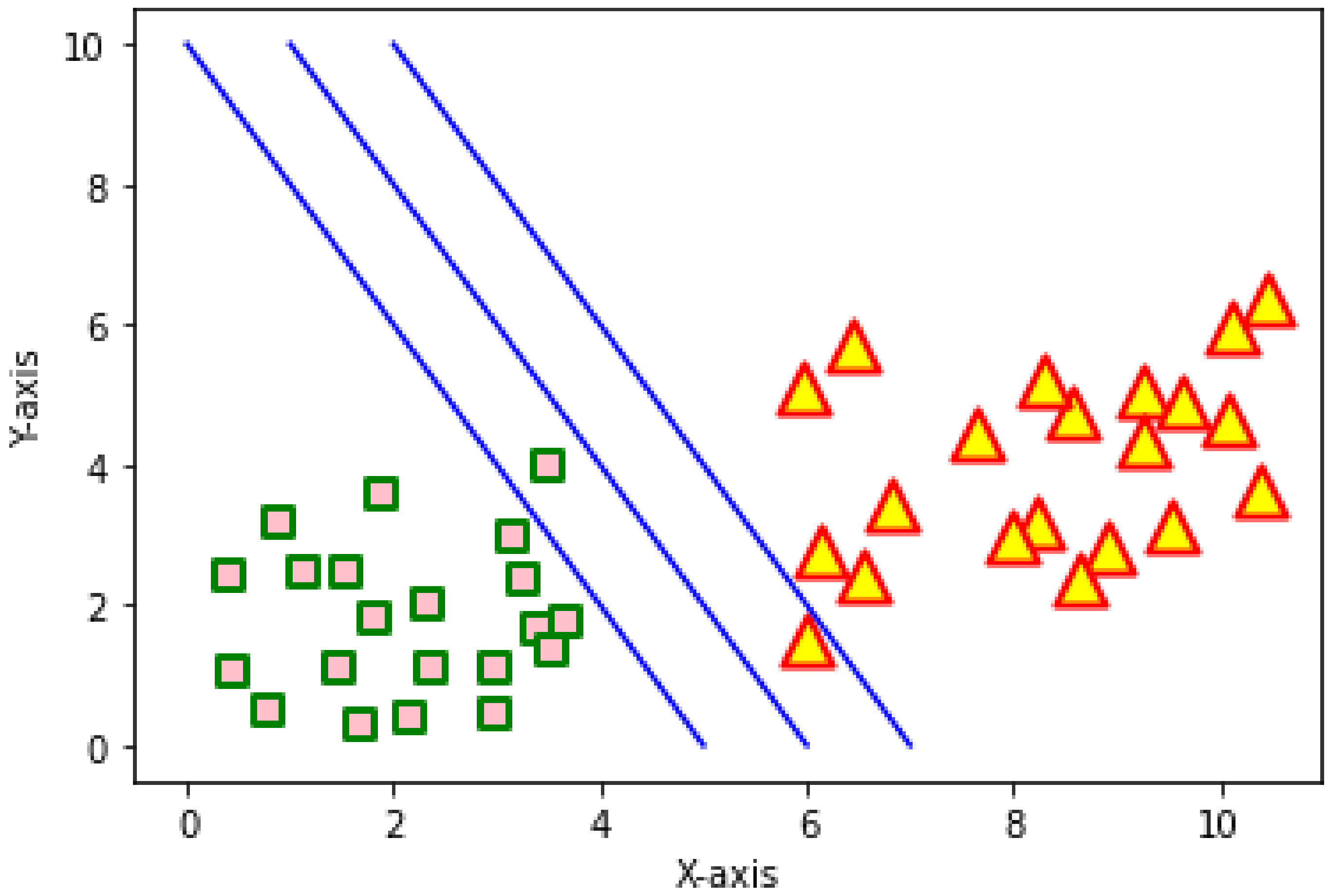
SVM

學習資料的分界

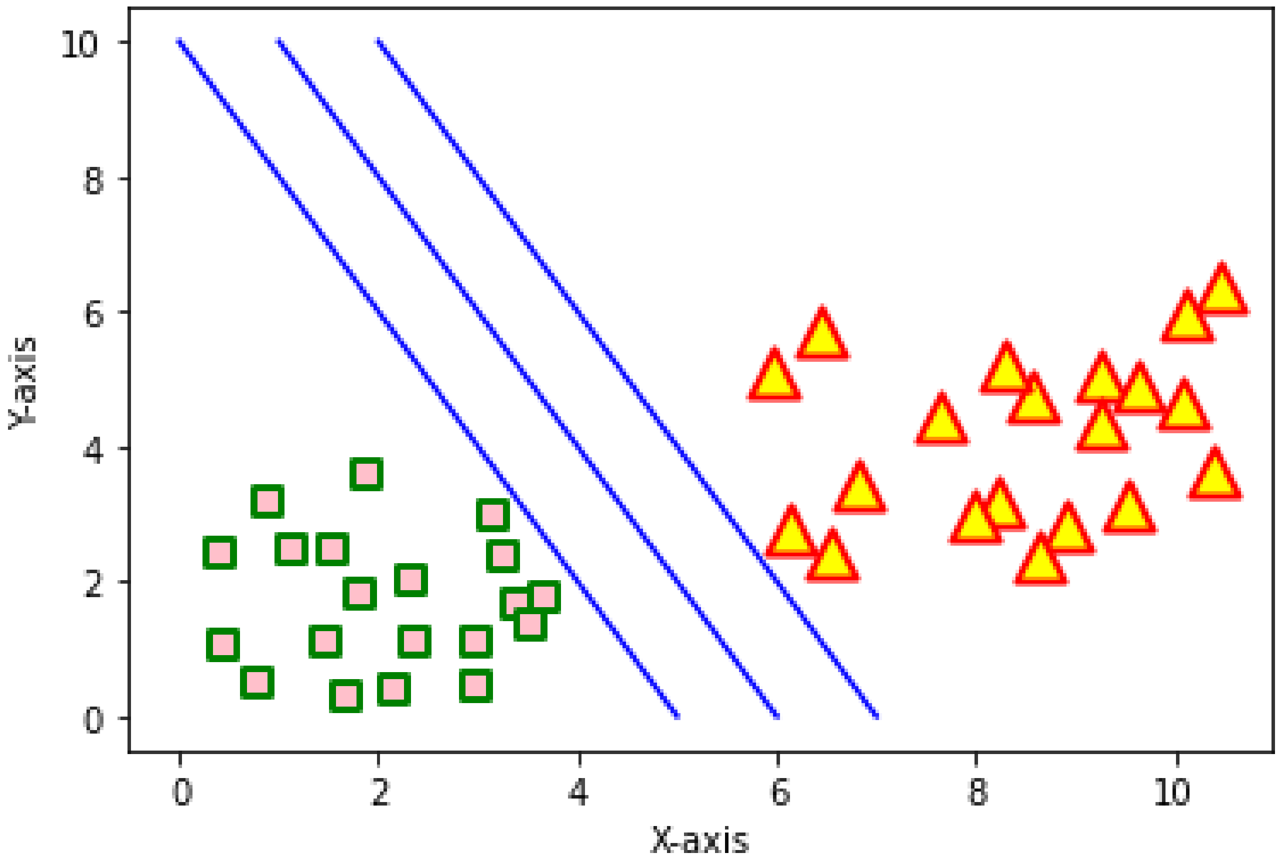
# Q：哪條邊界看起來最好?



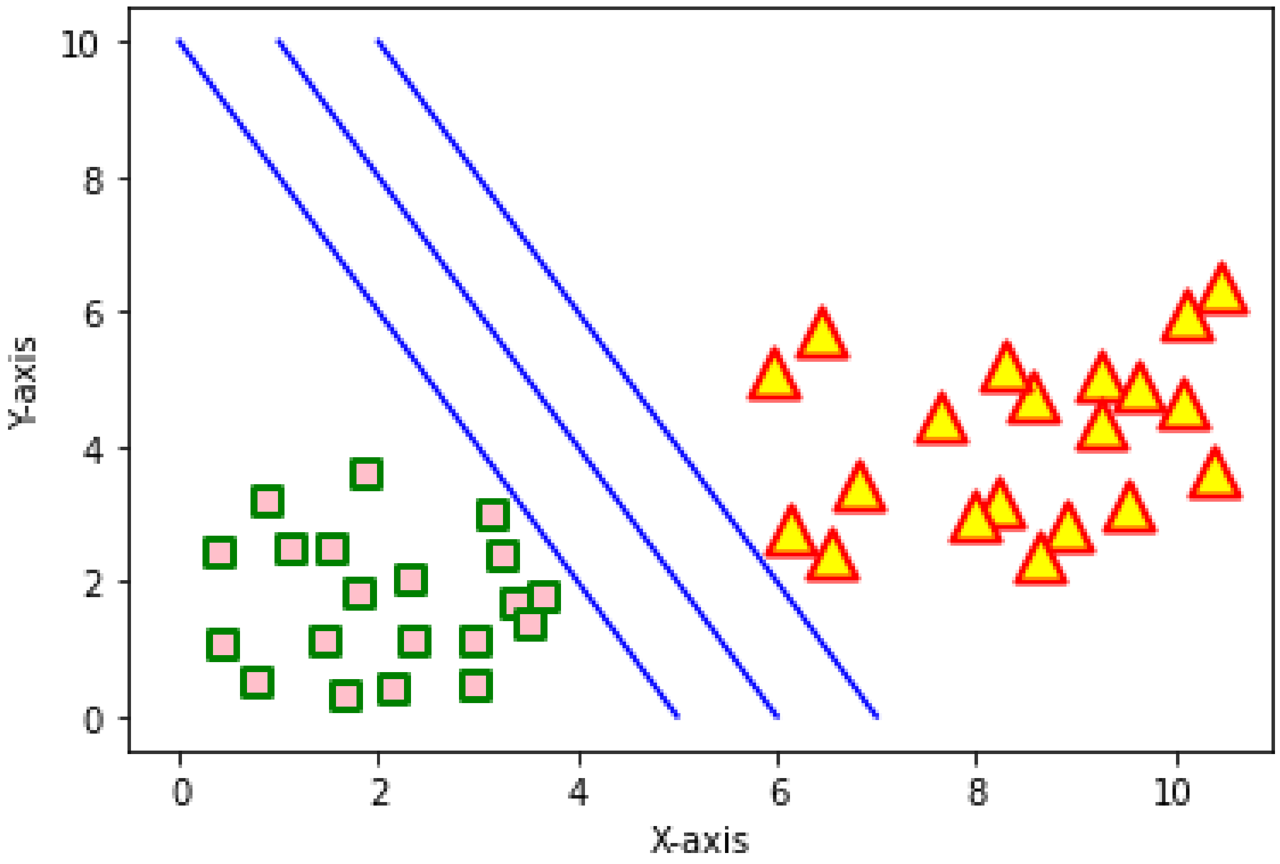
# 如果現在資料變成




中間那條藍色邊界比較能夠處理新資料有浮動的情況




中間的藍色邊界特徵：
讓邊界兩端的資料盡量分開




中間的藍色邊界特徵：
讓邊界兩端的資料盡量分開

Q：如何衡量「盡量分開」？

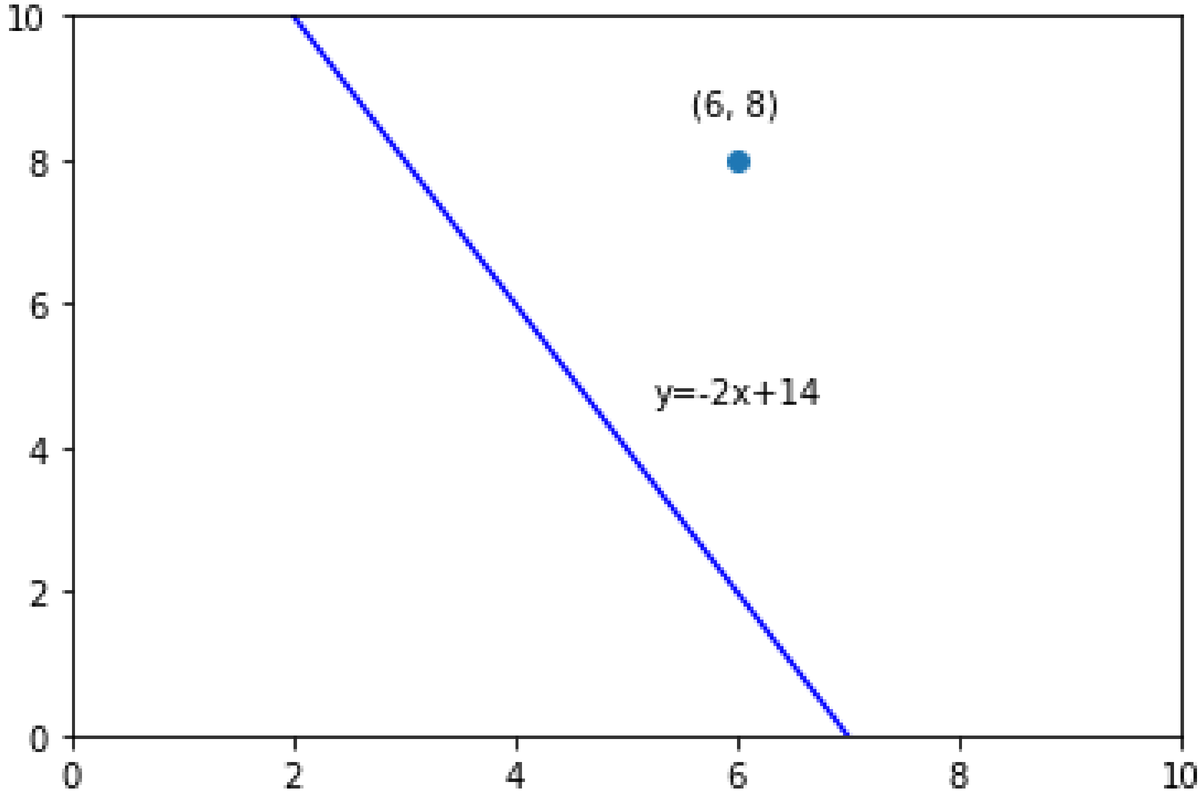
# 目標：離邊界最近的點能離邊界盡量遠

問題：如何衡量點到邊界的距離？

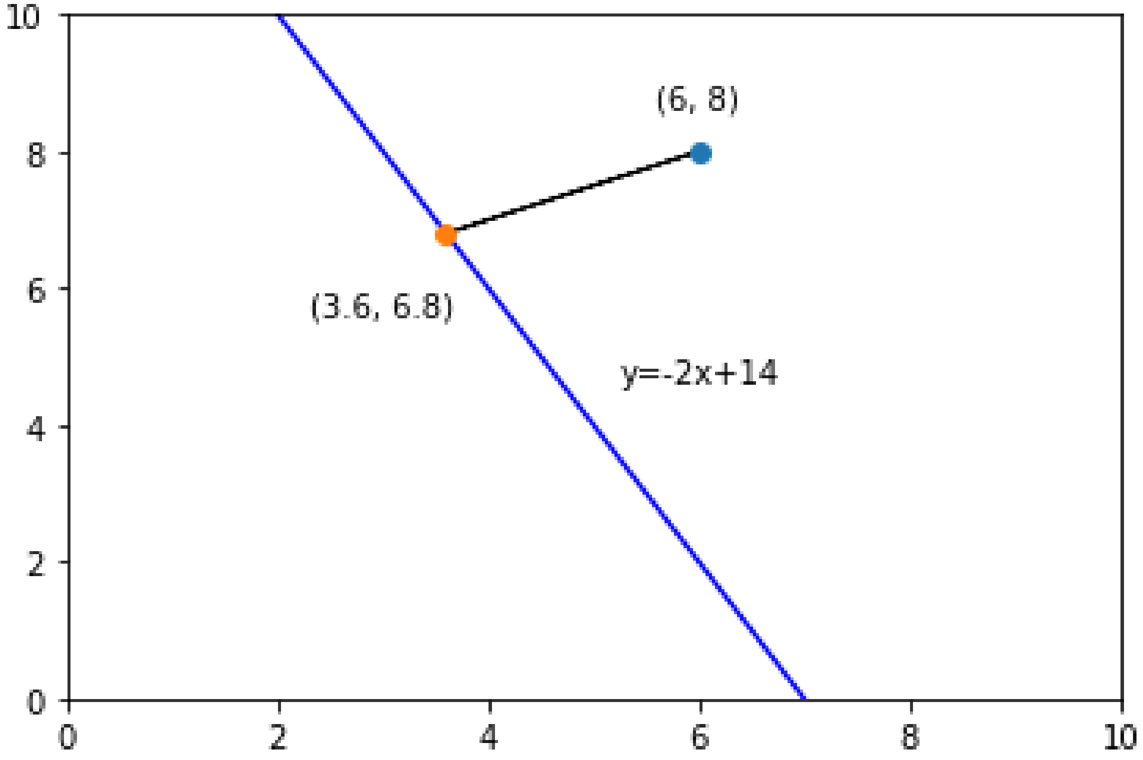
# 目標：離邊界最近的點能離邊界盡量遠

問題：如何衡量點到邊界的距離？

答：投影




假設給定一條線：$y = -2x + 14$
跟線外一點：(6,8)

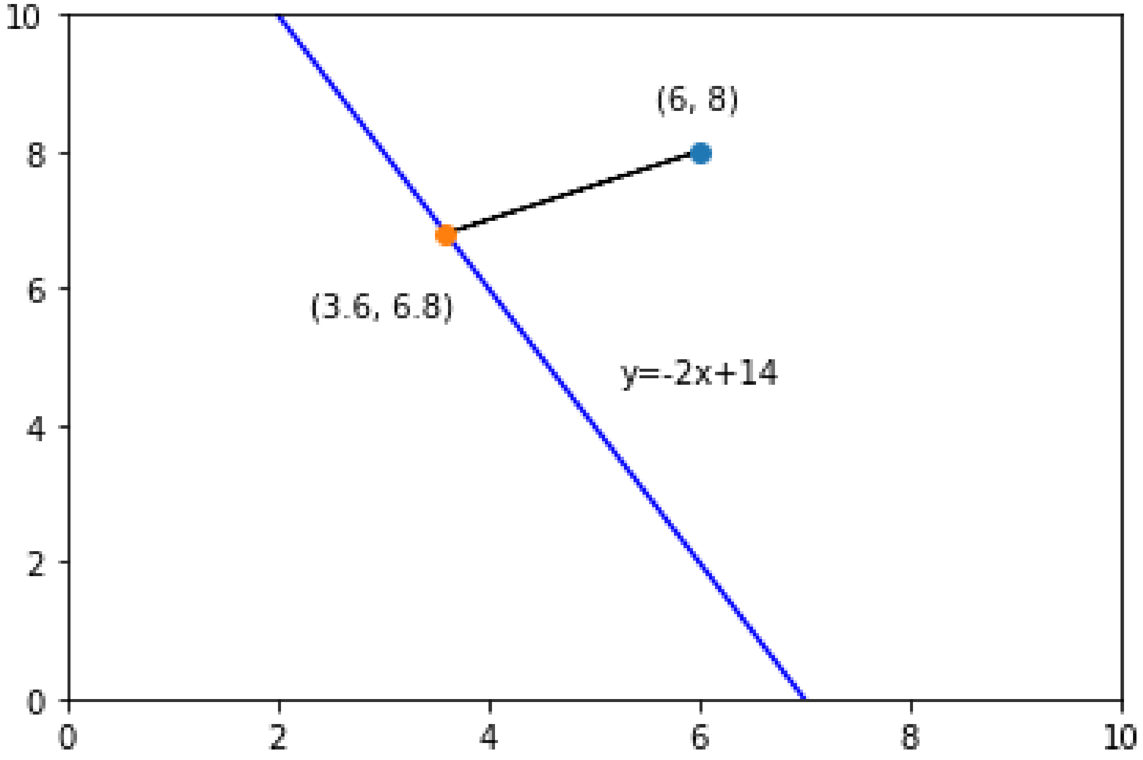



法向量為(1,-2)
假設垂直線方程式為：$x - 2y + b = 0$

帶入(6,8)後得出
垂直線方程式為：$x - 2y + 10 = 0$
並且求出兩條線交點

$$\begin{cases} 2x + y = 14 \\ x - 2y = -10 \end{cases}$$

得出解(3.6,6.8)，即為投影點




而$(6,8) = (3.6,6.8) + \frac{6}{\sqrt{5}} * (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$

其中

$\frac{6}{\sqrt{5}}$ 為兩點距離

(2,1) 為平面向量 (對應$2x + y = 14$)

$\sqrt{5}$為平面向量之長度

若公式化表示的話：

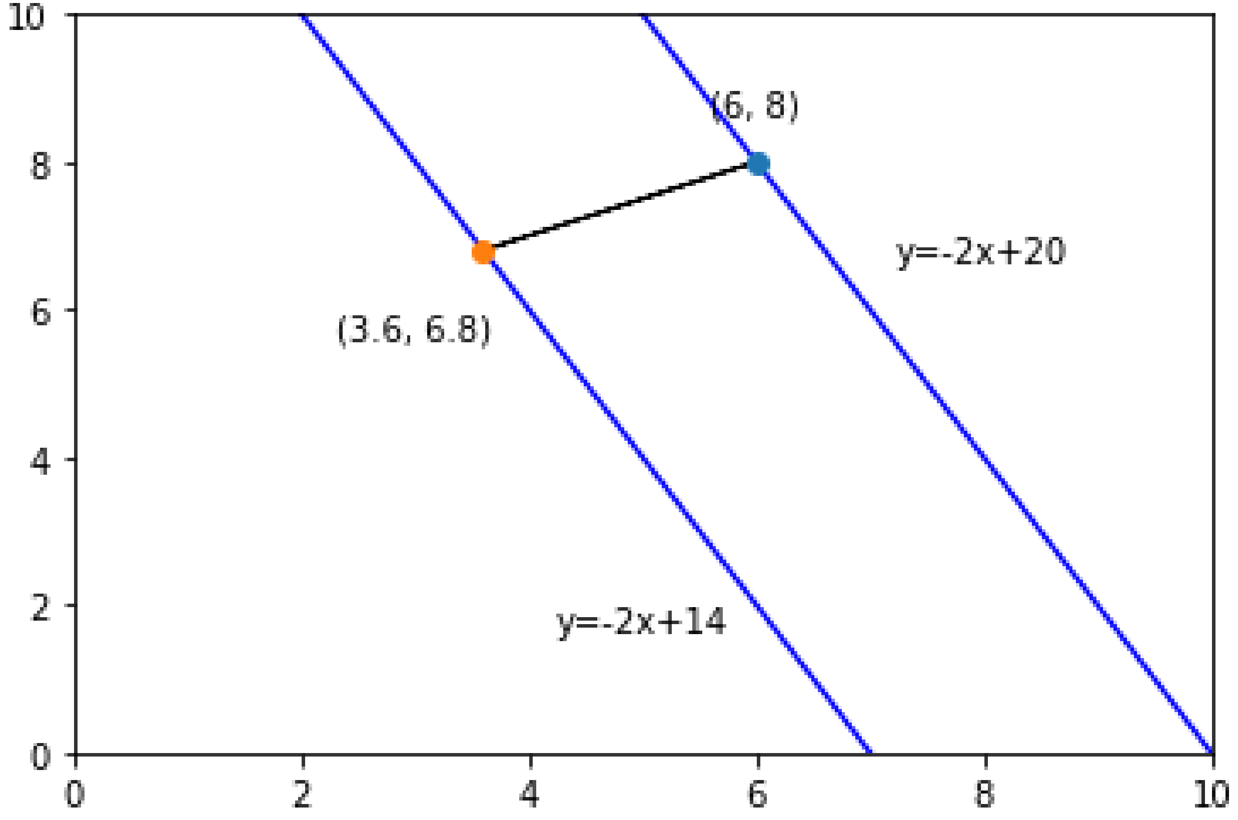
線：$f(x) = wx + b$
線外一點$\hat{x}$
$\hat{x}$在線上之投影點為$x_{\perp}$
點到線的距離為 $r = distance(\hat{x}, x_{\perp})$

我們可以得出關係式：

$$\hat{x} = x_{\perp} + r * \frac{w}{\|w\|}$$

求出投影點即可對應出點到線之距離

# 換個角度來看點到線的距離




$$r = distance(\hat{x}, x_{\perp})$$

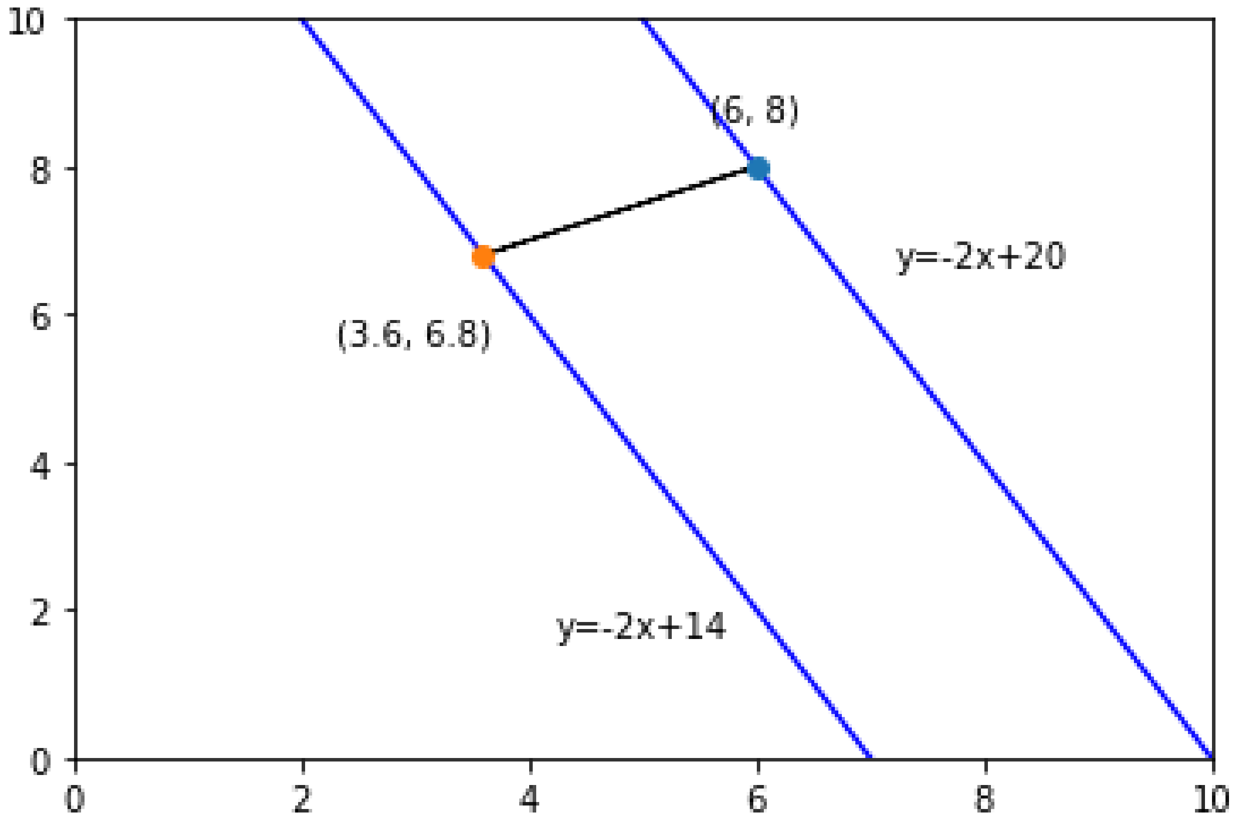
$=$兩平行線距離

$$\begin{cases} 2x + y - 14 = 0 \\ 2x + y - 20 = 0 \end{cases}$$

之距離為

$$\frac{|-14-(-20)|}{\sqrt{2^2+1^2}} = \frac{6}{\sqrt{5}}$$

# 換個角度來看點到線的距離



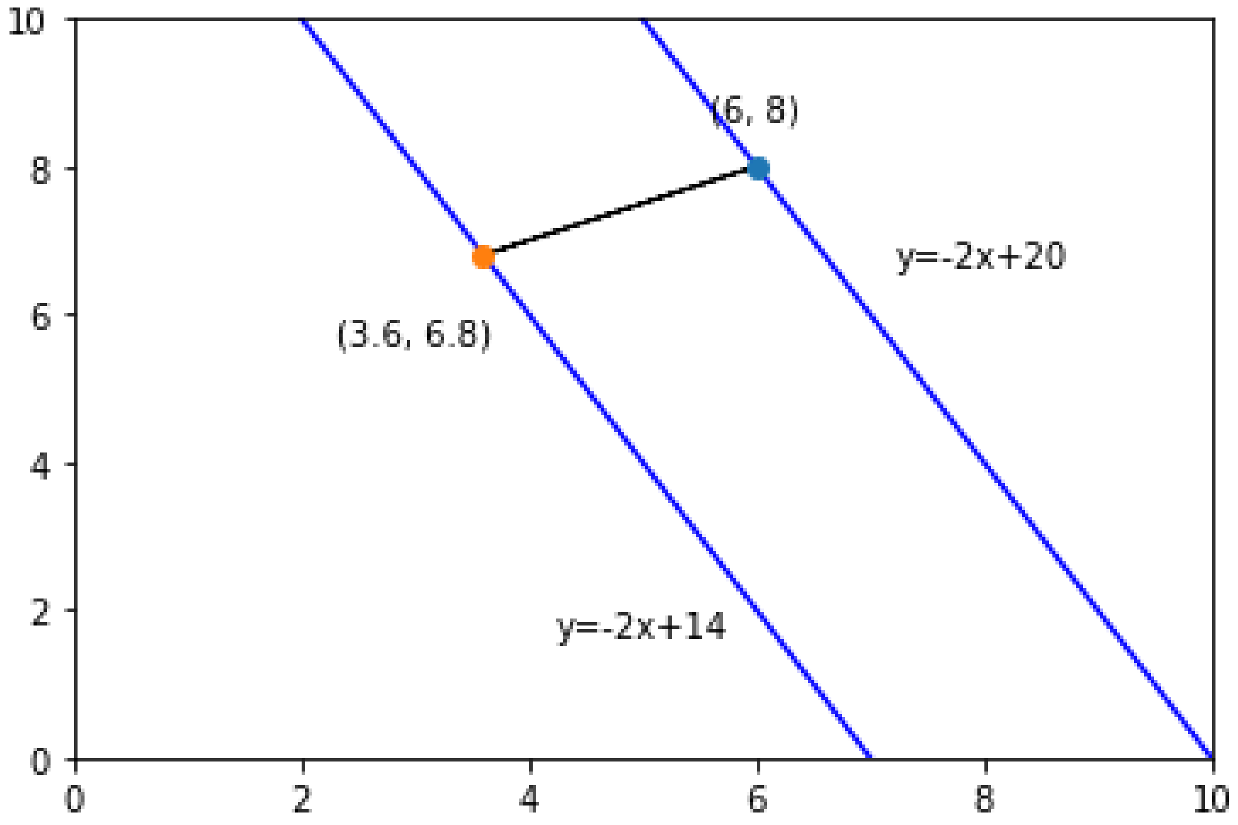

$$r = distance(\hat{x}, x_{\perp})$$
$=$兩平行線距離

$$\begin{cases} 2x + y - 14 = 0 \\ \boxed{2x + y - 20 = 0} \end{cases}$$

之距離為

$$\frac{|-14-(-20)|}{\sqrt{2^2+1^2}} = \frac{\boxed{6}}{\sqrt{5}}$$

# 換個角度來看點到線的距離




$$r = distance(\hat{x}, x_{\perp})$$
=兩平行線距離

$$\begin{cases} 2x + y - 14 = 0 \\ \boxed{2x + y - 14 = 6} \end{cases}$$

之距離為

$$\frac{|-14 - (-20)|}{\sqrt{2^2 + 1^2}} = \frac{\boxed{6}}{\sqrt{5}}$$

$$|2 * 6 + 8 - 14| = 6$$

所以6也等於原先的點帶入原先的線方程式的值之後取絕對值

沿用前面的公式化描述：

面：$f(x) = wx + b,\ \ x$ 為二維空間之一點

面外一點 $\hat{x}$

點到面的距離為 $r$為

$$r = \frac{|w\hat{x} + b|}{\|w\|}$$

對應回前面例子：

$\hat{x} = (6,8); \mathrm{w} = (2,1);\ \mathrm{b} = -14$

沿用前面的公式化描述：

面：$f(x) = wx + b,\ \ x$ 為二維空間之一點

面外一點 $\hat{x}$

點到面的距離為 $r$ 為

$$r = \frac{|w\hat{x} + b|}{\|w\|}$$

SVM使用此方法描述點到邊界之距離

對應回前面例子：

$\hat{x} = (6,8); \mathrm{w} = (2,1);\ \mathrm{b} = -14$

# 二分類問題描述

- 資料集X包含N筆資料$x_1, \ldots, x_N$以及N個相對應label值$l_1, \ldots, l_N$

(亦即資料跟相對應分類)

分類的變數稱為Y，值為$y_1, \ldots, y_N \ \in \{1, -1\}$

label的值也落在$\{1, -1\}$

現在目標：找出線性函數當作類別的分界，並且讓資料跟邊界的距離盡量遠

也就是找出權重$w$跟bias $b$滿足

$$y_i = wx_i + b$$

且

$$\frac{|wx_i + b|}{\|w\|} \geq \rho\,, \qquad i = 1, \ldots, N$$

而我們希望$\rho$ 的值越大越好

$$\frac{|wx_i + b\,|}{\|w\|} \geq \rho$$

$$\Leftrightarrow \frac{l_i(wx_i + b)}{\|w\|} \geq \rho \quad if \quad \begin{cases} wx_i + b > 0\; when\; l_i = 1 \\ wx_i + b < 0\; when\; l_i = -1 \end{cases}$$

$$\Leftrightarrow l_i(wx_i + b) \geq \|w\|\rho \quad if \quad \begin{cases} wx_i + b > 0\; when\; l_i = 1 \\ wx_i + b < 0\; when\; l_i = -1 \end{cases}$$

$$\frac{|wx_i + b\,|}{\|w\|} \geq \rho$$

$$\Leftrightarrow \frac{l_i(wx_i + b)}{\|w\|} \geq \rho \quad if \quad \begin{cases} wx_i + b > 0\ when\ l_i = 1 \\ wx_i + b < 0\ when\ l_i = -1 \end{cases}$$

$$\Leftrightarrow l_i(wx_i + b) \geq \|w\|\rho \quad if \quad \begin{cases} wx_i + b > 0\ when\ l_i = 1 \\ wx_i + b < 0\ when\ l_i = -1 \end{cases}$$

選取$w$跟$b$去最大化$\rho$ 不好處理，也可能得到多組解$(w, b)$

$$\frac{|wx_i + b\ |}{\|w\|} \geq \rho$$

$$\Leftrightarrow \frac{l_i(wx_i + b)}{\|w\|} \geq \rho \quad if \quad \begin{cases} wx_i + b > 0\ when\ l_i\ = 1 \\ wx_i + b < 0\ when\ l_i\ = -1 \end{cases}$$

$$\Leftrightarrow l_i(wx_i + b) \geq \|w\|\rho \quad if \quad \begin{cases} wx_i + b > 0\ when\ l_i\ = 1 \\ wx_i + b < 0\ when\ l_i\ = -1 \end{cases}$$

選取$w$跟$b$去最大化$\rho$ 不好處理，也可能得到多組解$(w, b)$
所以多加入條件 ：$\|w\|\rho$ = 1，亦即$\rho = \frac{1}{\|w\|}$

最大化$\rho$ 變成相當於最小化$\|w\|$

運用優化理論的方式，最小化$\|w\|$也相當於最小化$\frac{1}{2}\|w\|^2$

所以整個問題變成：

$$\min \frac{1}{2}\|w\|^2 \;\; subject\ to\ \ l_i(wx_i + b) \geq 1$$

- 接著將有限制式的優化問題轉換為無限制式的優化問題：透過 Lagrange multipliers $\lambda_i$ 方式

得到loss function

$$L(\lambda_i, w, b) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i (l_i(wx_i + b) - 1)$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i l_i(wx_i + b) + \sum_{i=1}^{N} \lambda_i$$

此問題為二次規劃問題(quadratic programming)，透過優化理論的技巧可以得出結論。

訓練完之結果會得到部分Lagrange multipliers $\lambda_i$為0之現象，涵義為我們不需要這些$x_i$的資料來幫助我們劃邊界。

主觀解釋可能原因：

離邊界太遠的點可以丟掉不考慮。

Lagrange multipliers $\lambda_i \neq 0$ 對應的資料$x_i$稱為支撐向量(support vector)

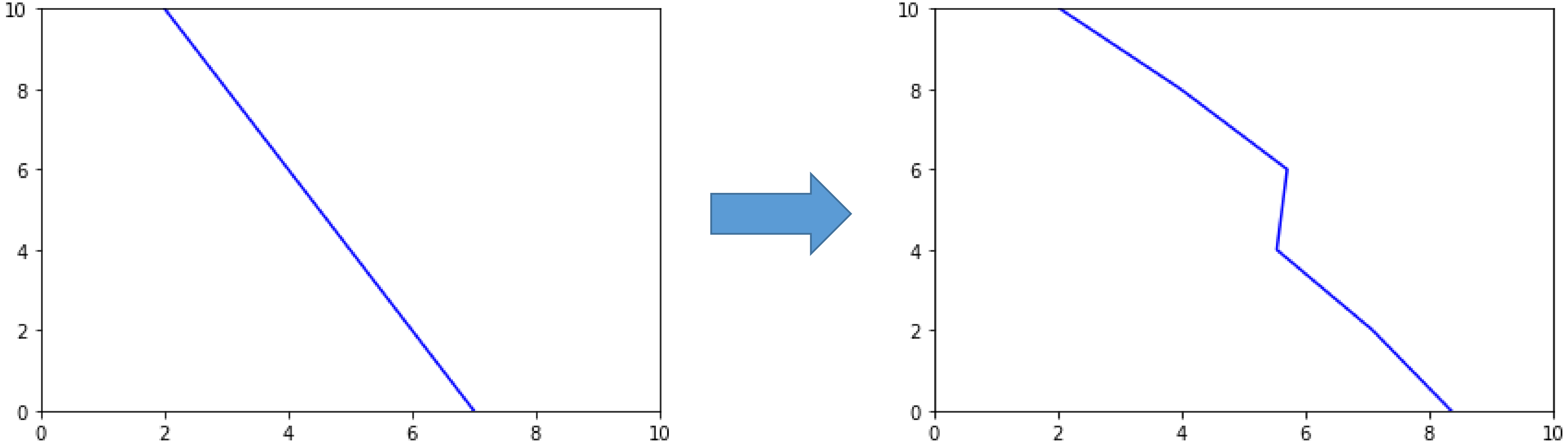
而整體分類流程則稱為支撐向量機(support vector machine)

# 進階技巧

問題：如果資料並非線性可分時該如何處理?

舉例：台北市行政區邊界劃分

# 解決方法1：讓邊界是可以彈性變化的



# 解決方法1：讓邊界是可以彈性變化的

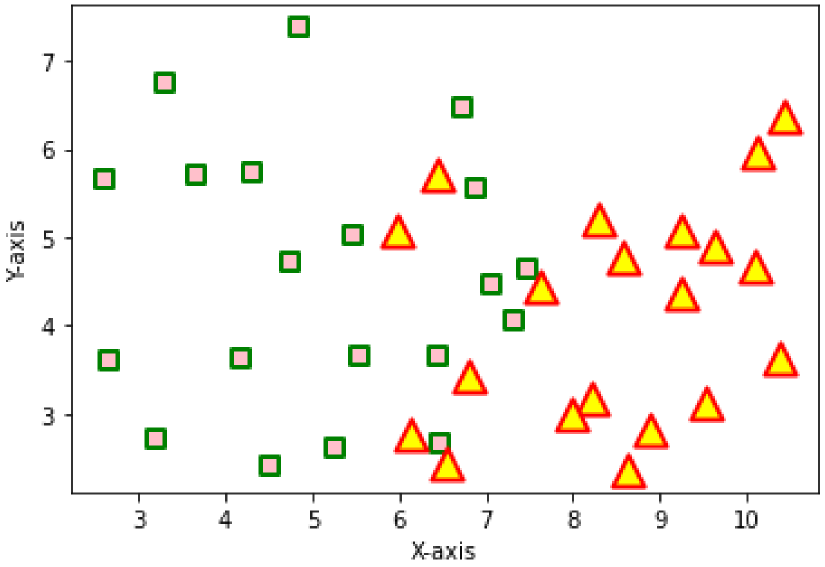
- 具體作法：

$$\min \frac{1}{2}\|w\|^2 \ \ subject\ to\ \ l_i(wx_i + b) \geq 1 - \xi_i$$

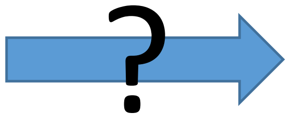
1. 如果$\xi_i = 0$ ，則回到線性可分的情況
2. 如果$0 < \xi_i < 1$ ，$x_i$的分類正確，但是離邊界太近
3. 如果$\xi_i \geq 1$，$x_i$的分類錯誤

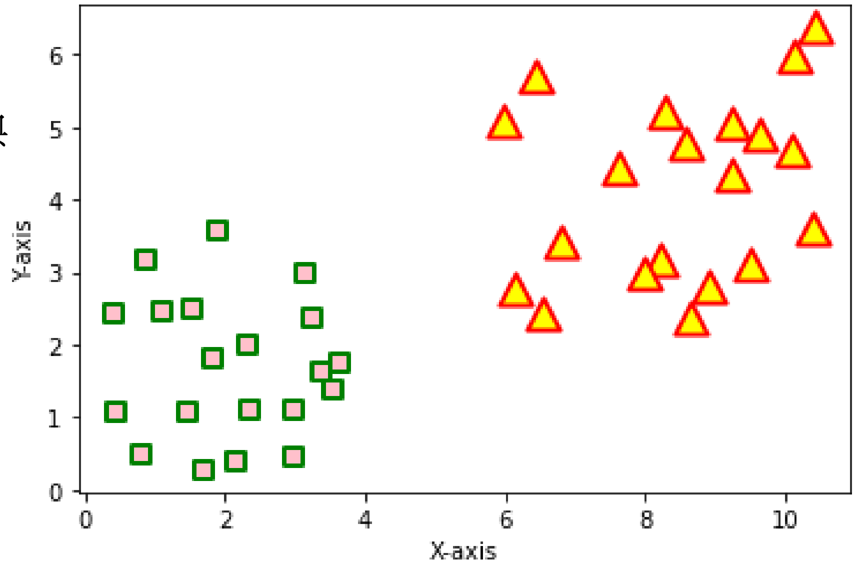
# 解決方法2：將資料做轉換後再嘗試線性分類

- 想法：




非線性函數轉換




原資料

# 解決方法2：將資料做轉換後再嘗試線性分類

- 具體作法：

設定轉換函數$\phi$，稱為basis function。
之後將資料點$x_i$轉換成$\phi(x_i)$
接著嘗試

$$\min \frac{1}{2}\|w\|^2 \;\; subject\ to \;\; l_i(w\phi(x_i) \; + b) \geq 1$$

# 解決方法2之延伸：類神經網路

- 問題：如果非線性轉換後還是無法線性可分的話呢?
- 答：一次轉換不夠就做兩次，兩次不夠就三次...