



2.2 SVM 支援向量機



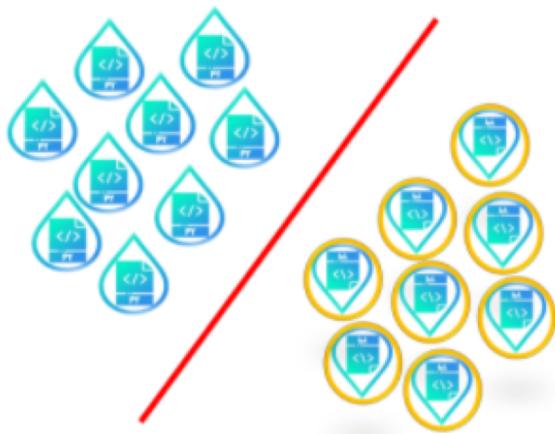
- 是一種知名的二元分類器(binary classifier)
- 基於統計學習理論基礎的機器學習模型
- **針對小樣本、非線性、高維度與局部最小點等問題具有相對的優勢**
- 計算速度快且空間成本低等優勢

1

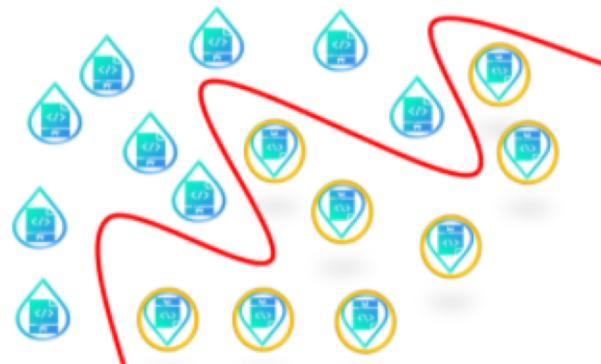
SVM概念



SVM是一種線性分類器
同時卻也可推展到解決**非線性**的分割問題



線性可分

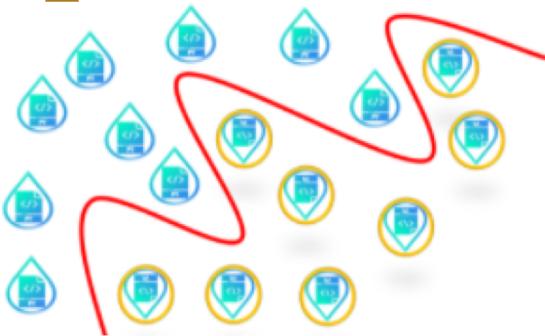


線性不可分



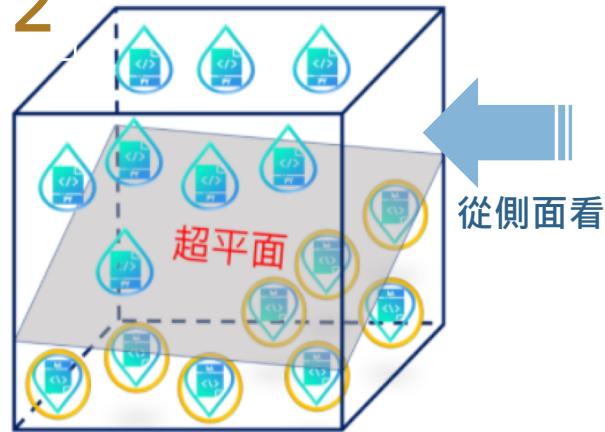
將在低維度空間線性不可分的樣本映射到高維度空間
找到一個超平面將這些樣本做有效的切割
且超平面兩邊的樣本要盡可能地遠離這個超平面(hyperplanes)

1

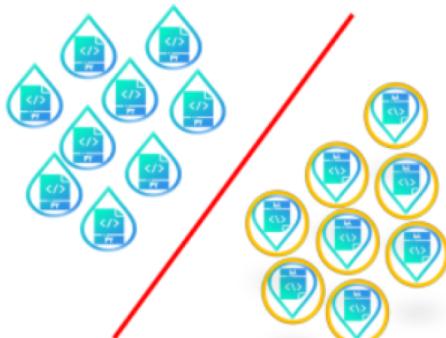


映射到高維度空間

2



3





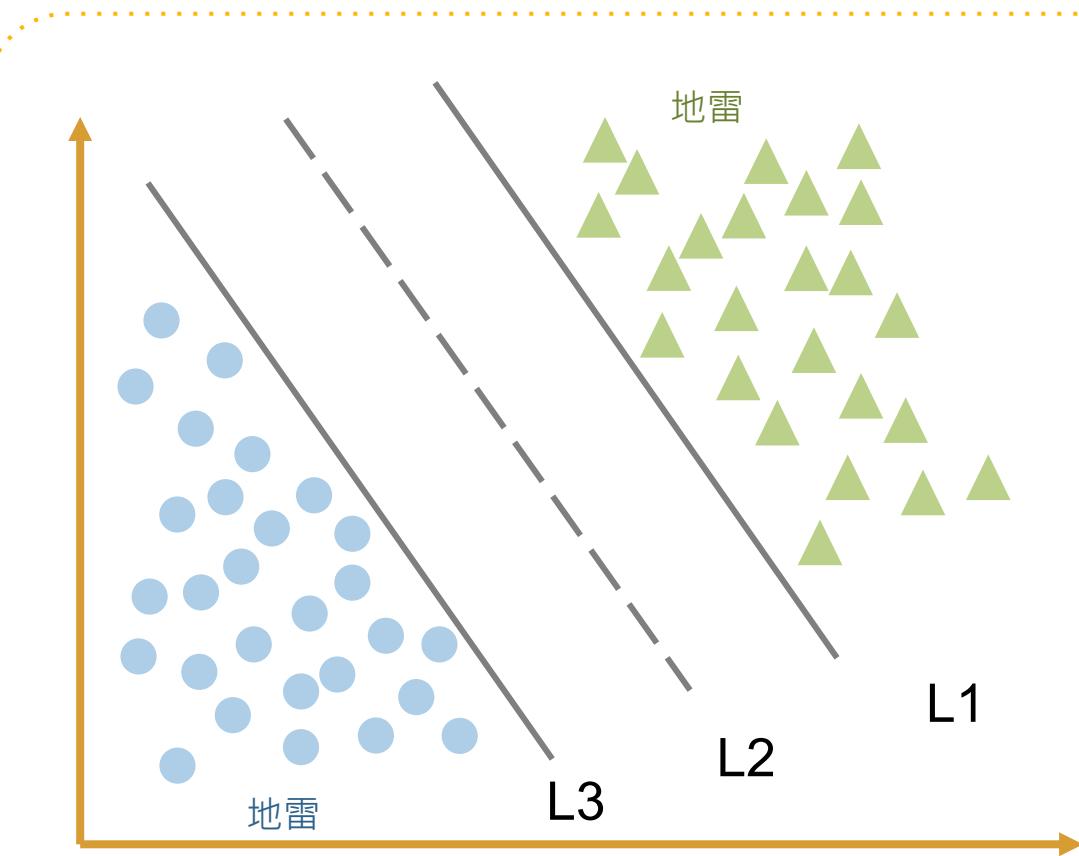
SVM visualization

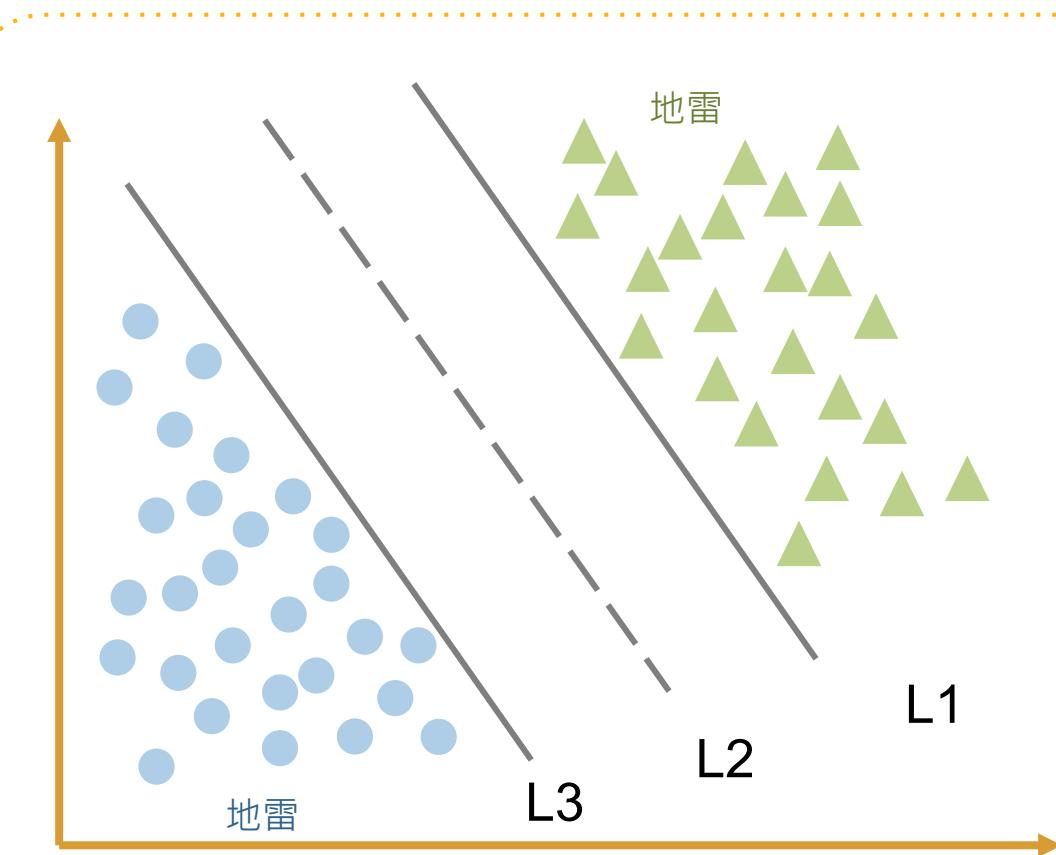
Margin最大化





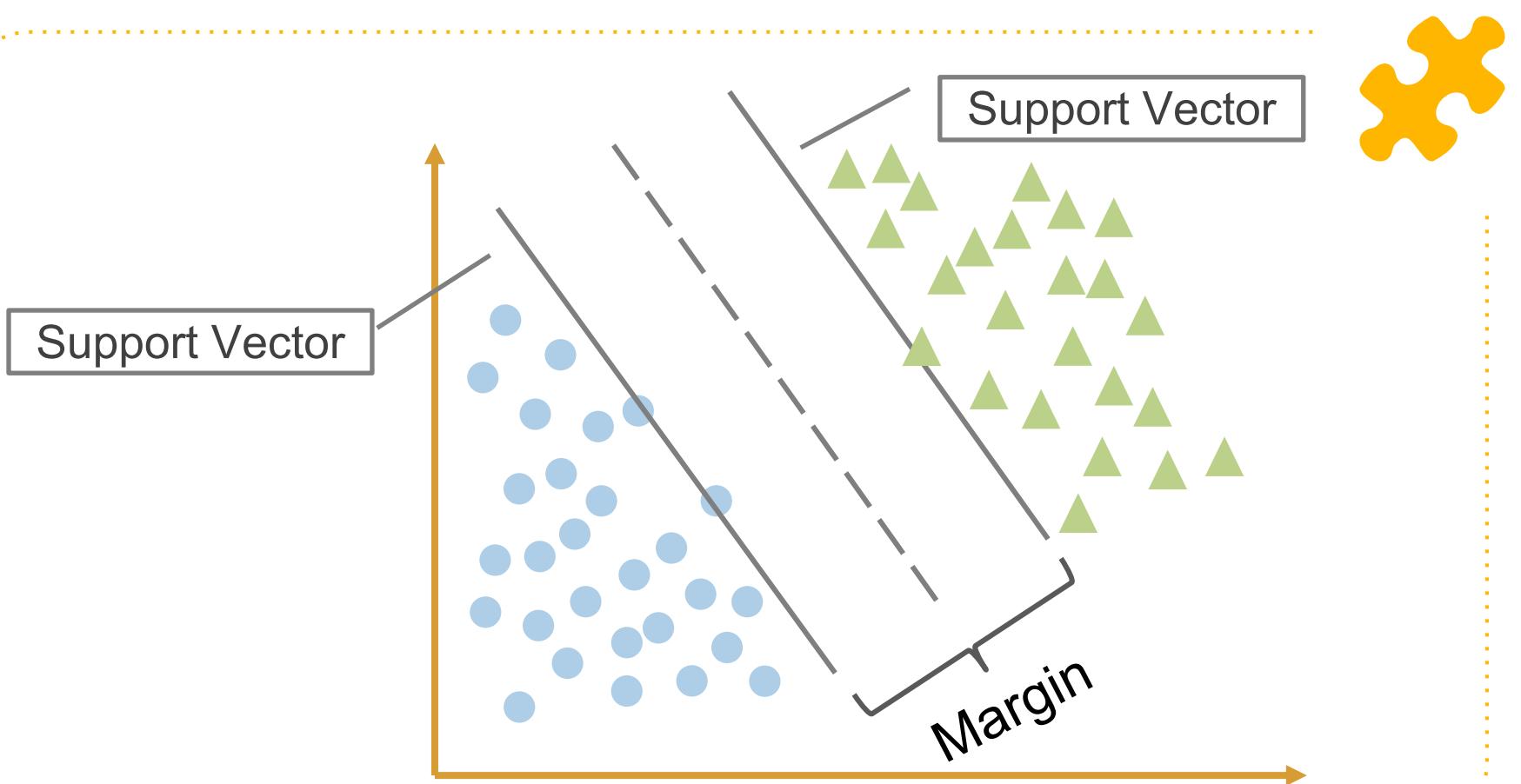
你會選哪條線作為
樣本分割的線？





L1: too close to ▲
L3: too close to ●
L2: just right





“

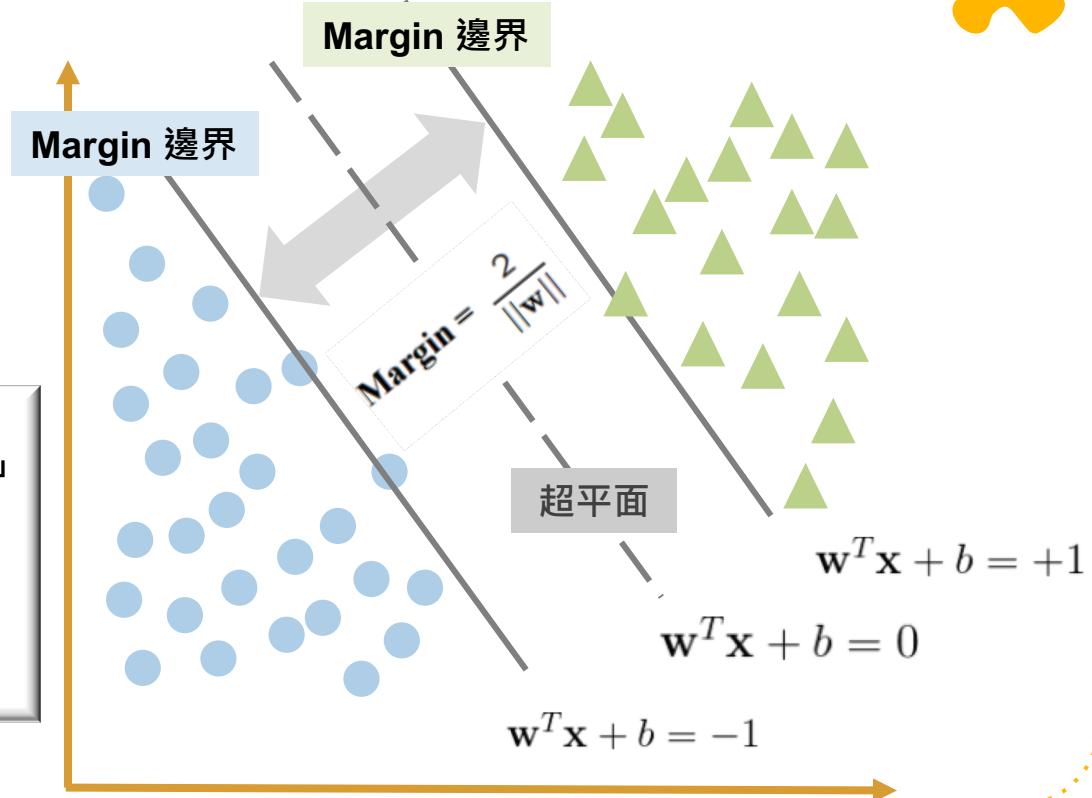
A Max Margin Classifier

如何找到讓Margin最大的超平面



極大化「邊距」就是極小化 $\|w\|$

- w ：法向量，決定超平面的方向
-> 告訴你這條線「朝哪個方向傾斜」
- b ：偏移量，決定超平面距離原點的距離
-> 告訴你這條線「在這個方向上，離原點推多遠」
- x ：資料點



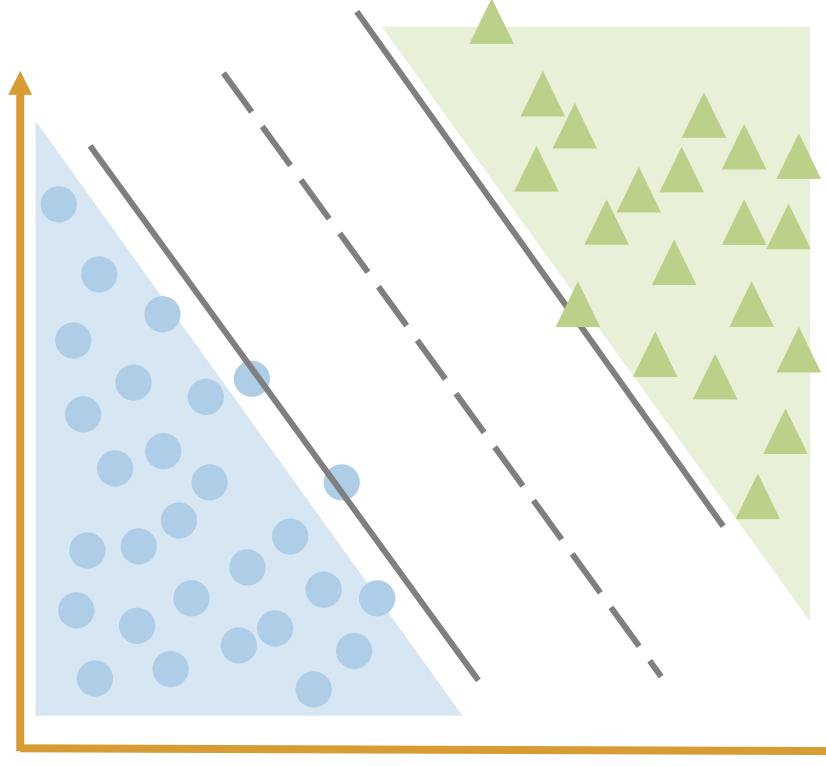
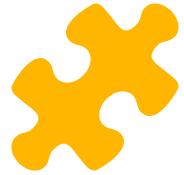


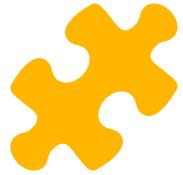
想像你在地上畫一條界線（分左右兩邊）

- w 告訴你這條線該朝哪個方向
- b 告訴你這條線該畫在離原點多遠的地方
- SVM 就是在選擇一組 w 和 b 的組合，讓這條線能把兩邊的樣本分得最開，並且離得最近的樣本還能有最遠的距離（最大 margin）。

Hard-margin SVM v.s. Soft-margin SVM

Hard-margin SVM

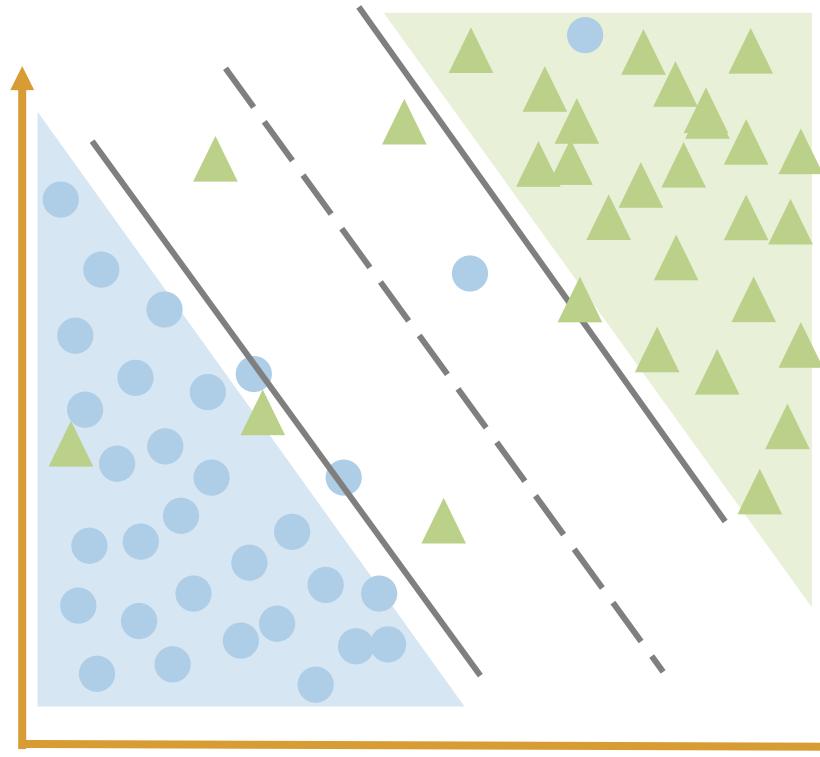




- SVM最初要尋找一個能夠完美將「所有」資料分成兩邊，具有最大margin的超平面，這被稱為“hard-margin SVM”。
- 由於hard-margin SVM，追求要將資料完美分好，因此很容易有overfitting的風險。
- 1995年，Vapnik等人提出了“soft-margin SVM”，讓SVM能容許一些被分錯的資料存在。



Soft-margin SVM



3

參調



kernel

當不同類別的資料在原始空間中，無法被線性分類器區隔開來時，經非線性投影後，能在更高維度的空間中可以區隔開。

常用的kernel函數：

$$\text{Linear kernel: } k(x, y) = \langle x, y \rangle$$

$$\text{Polynomial kernel: } k(x, y) = (\langle x, y \rangle + c)^d$$

>>> Gaussian Radial Basis Function kernel (RBF): $k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$

d為正整數(degree)

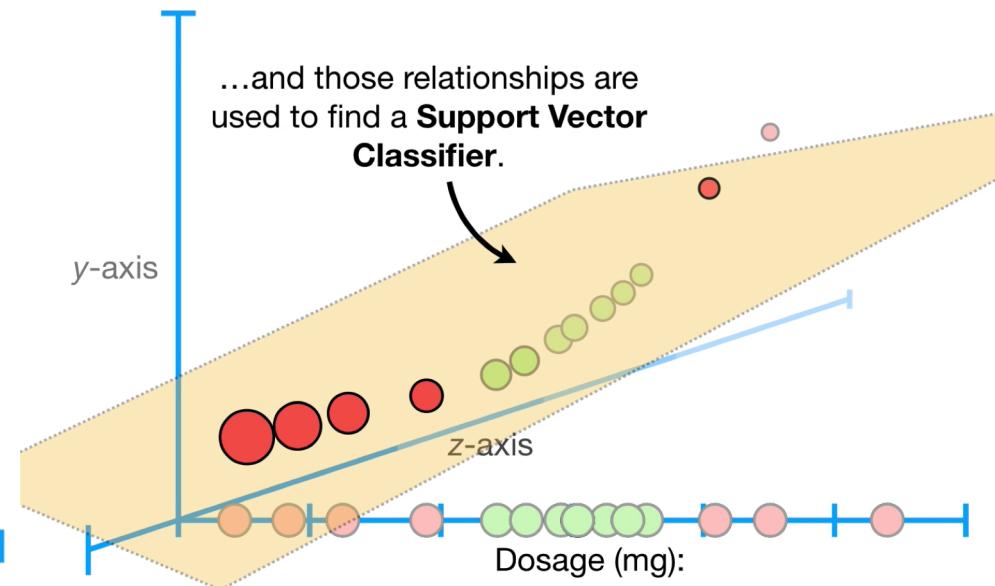
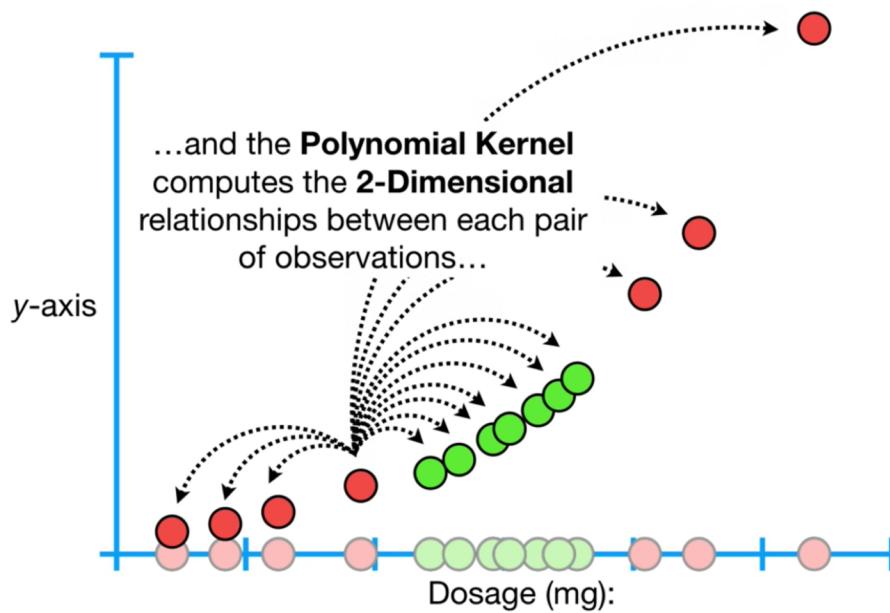
σ 為非0的實數(gamma)，需大於0



Linear kernel

- Linear核函數對數據不做任何變換。 $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- 使用時機：特徵較豐富，樣本量大，需進行實時計算。
- 不需設置任何參數，可直接使用。

Polynomial kernel





類型	項目	意義
線性	x_1, x_2	各看各的
平方	x_1^2, x_2^2	單一特徵變強。 這個特徵本身很大時，影響是不是會加速放大？
交叉	x_1x_2	兩個特徵同時出現時（組合），效果是不是不一樣？

- 例子：貸款風險
 x_1 ：收入
 x_2 ：負債
- Linear 學到：
收入高 → 風險低
負債高 → 風險高
- degree = 2 能多學到：
收入高 \times 負債低 → 非常安全
收入低 \times 負債高 → 非常危險
- ✓ 「又高又低」的判斷，就是交叉項在做的事。

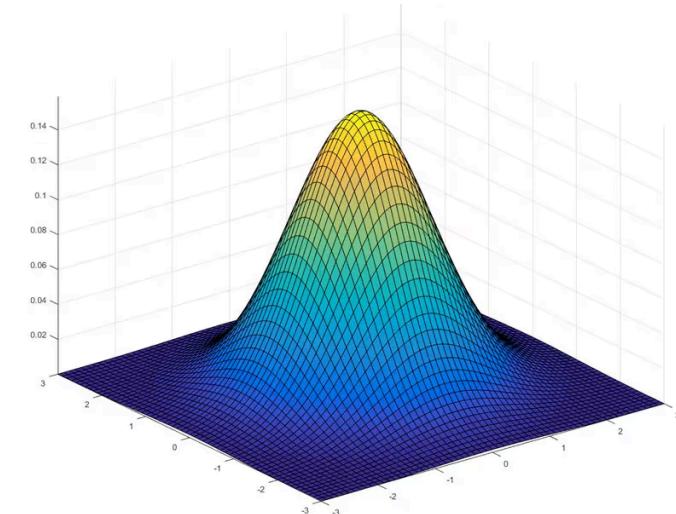


RBF kernel (高斯核函數)

$$k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

- $x-y$ 表示兩個樣本之間的距離衡量。
- 若 x, y 很相似，結果是1，很不相似為0。

“如果 A 跟 B 很像，B 跟 C 也很像
那 A 跟 C 應該也會被當成一類
但如果 D 離大家都很遠，它應該屬於另一類”

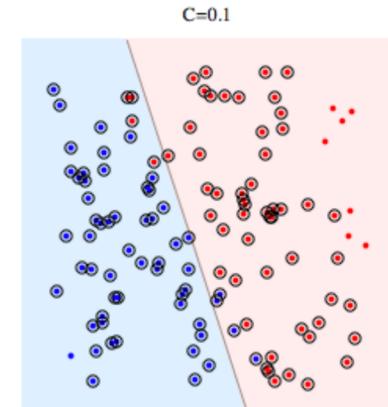
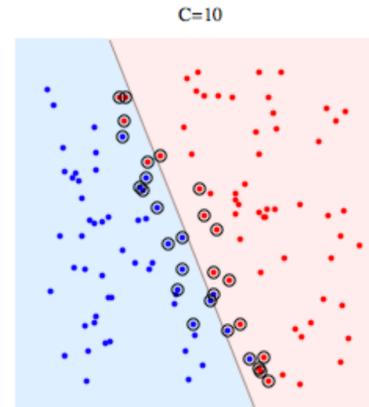
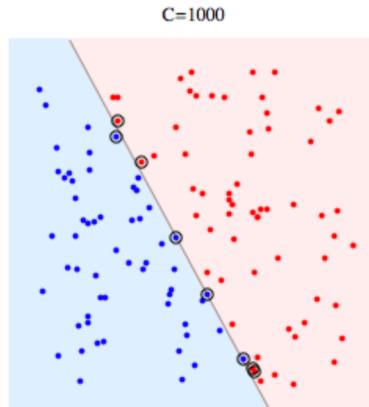




C (cost)

在soft-margin SVM的損失函數(loss function)中，C的存在就是容錯項，藉由C給予那些被分錯的資料懲罰值，控制support vectors(用來決定超平面的那些資料點)的影響力。

- C越大，代表容錯越小，越少support vectors，越接近 hard-margin SVM 的概念，卻容易 overfitting
- C越小，代表容錯越大，越多support vectors，margin的範圍越來越大





Gamma

控制模型複雜程度(控制Radial basis function kernel的大小)

- gamma越大：模型越複雜，映射維度越高，容易勾勒出擬合近點的超平面，也容易造成overfitting。
- gamma越小：模型越精簡，映射維度偏低，因此能勾勒出平滑、近似直線的超平面。



Kernel 優缺點

Linear Kernel

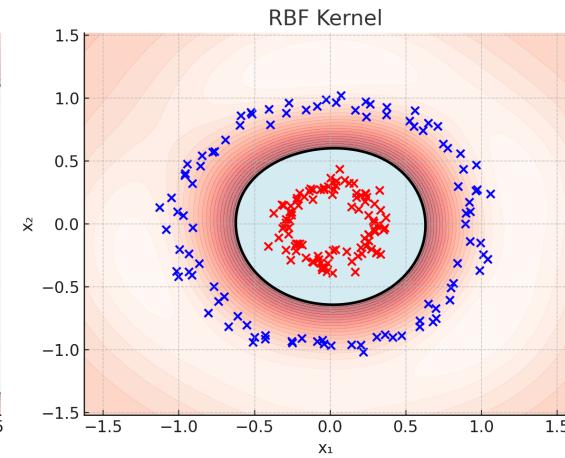
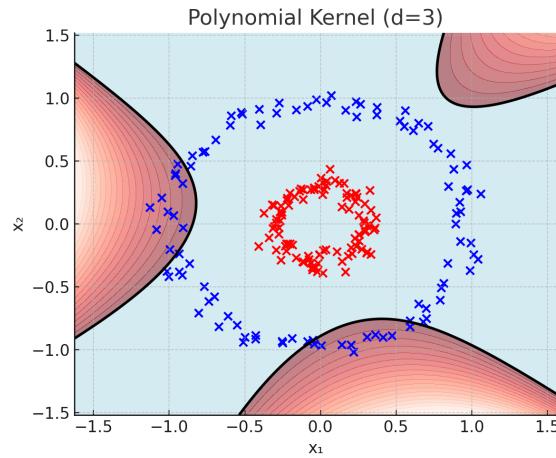
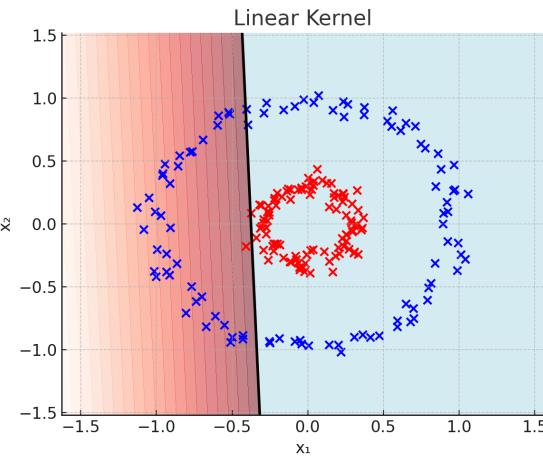
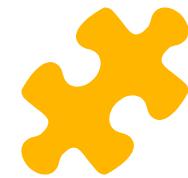
- 優點：模型較為簡單，也較安全，不容易 overfit。
- 缺點：限制較多，如果資料點非線性可分就沒用。

Polynomial Kernel

- 優點：可以進行d次轉換，分類能力會比 Linear Kernel 好。
- 缺點：高次轉換可能會造成計算結果怪異。因此會用在比較低次轉換的 SVM 問題上，但這樣就可以用 Linear SVM 取代。

Gaussian Kernel

- 優點：無限多維的轉換，分類能力更好，需選擇參數少。
- 缺點：較沒有解釋性，也較易發生 overfitting。



4

SVM的優缺點



優點：

1. 可解決小樣本下的問題
2. 可以解決多特徵空間
3. 透過不同核函數的選擇，可處理不同資料
4. 泛化能力強



缺點：

1. 樣本量太大，映射維度過高容易造成運算上的負擔
2. 特徵遠大於樣本的情況下，易造成過度擬合問題
3. 非線性問題的核函數選擇沒有通用標準，難以確定一個適合的核函數
4. 對缺失值敏感



Thanks!