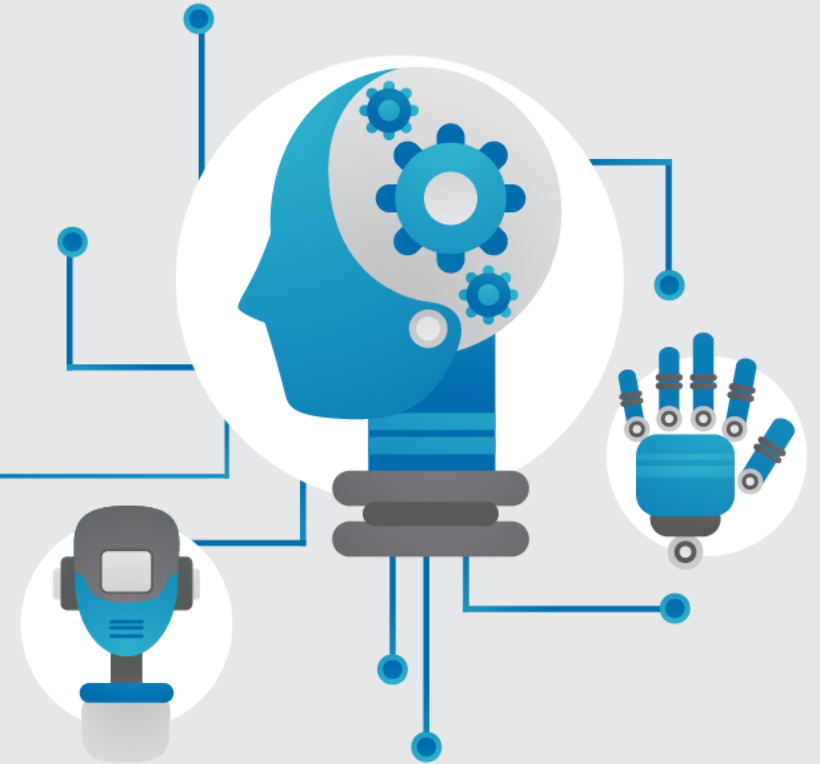


支援向量機





支援向量機



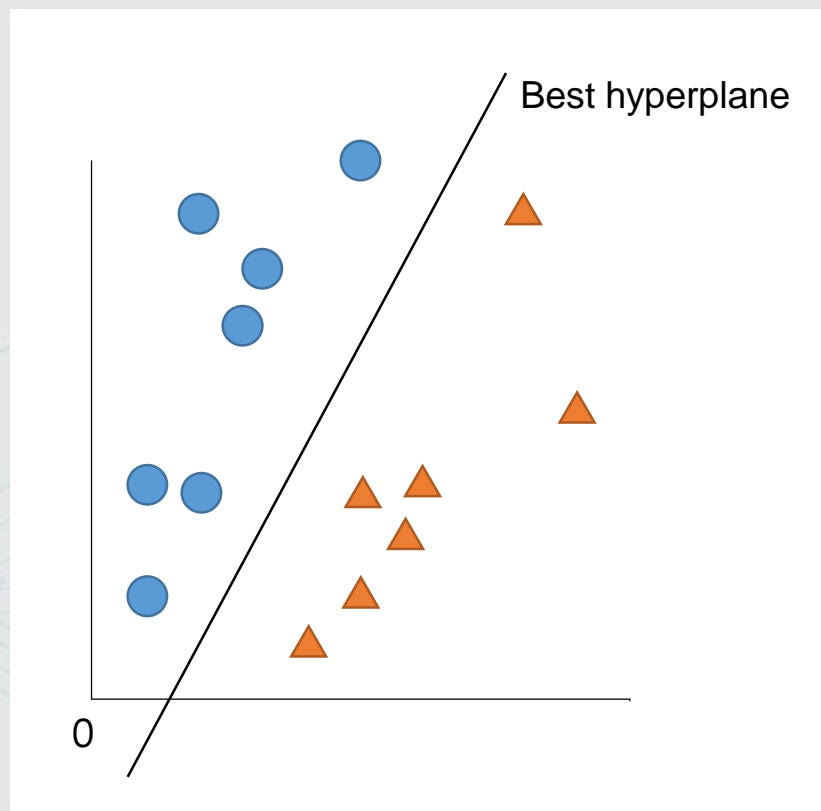
- › 支援向量機最早由V. N. Vapnik 等人於1963 (Support Vector Machine) 年所提出，是一種可以用於分類或迴歸的監督式機器學習演算法，需事先定義各分群的類型，屬於一般化線性分類器，主要用於二元分類的問題。
- › SVM的重要優點是能處理線性不可分的情形
- › SVM模型是將實例表示為空間中的點，在高維度的空間中尋找一個超平面作為類別的分割，以確保最小的分類錯誤率。

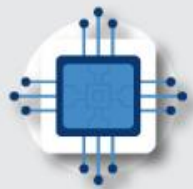


支援向量機概念



- › SVM的基本概念很簡單，就是找到一個決策邊界 (decision boundary) 讓兩類之間的邊界 (margins) 最大化，使其可區隔開來。

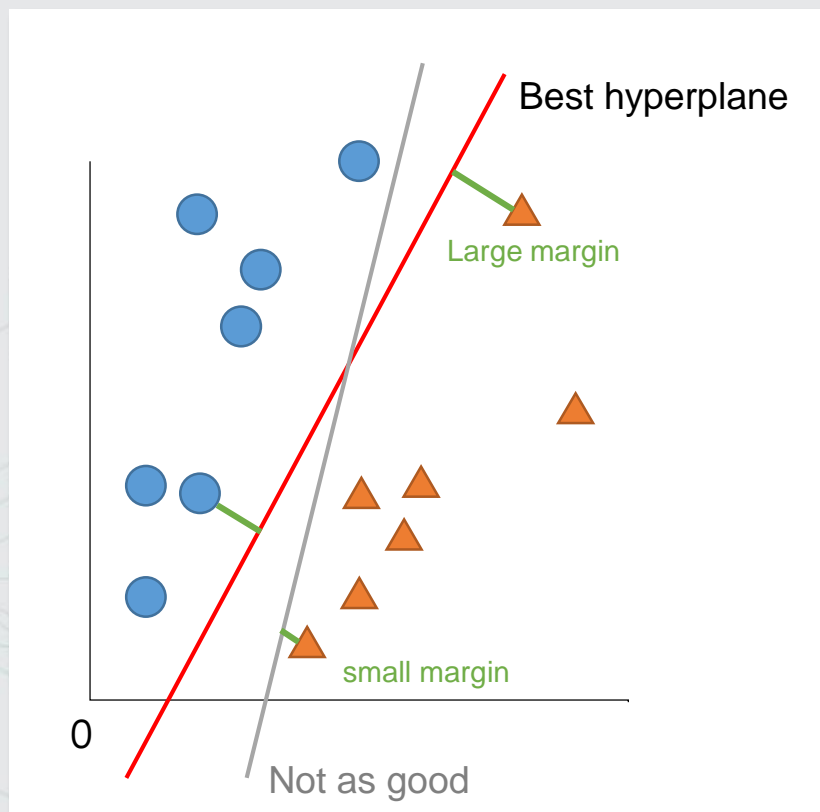


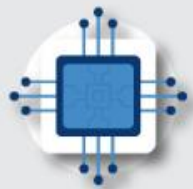


支援向量機概念



- 對SVM來說，最好的超平面不是下圖的灰線，而是最大化超平面與每個類別的距離的，如下圖紅線。即超平面對於每個類別的最近的元素，距離是最遠的。

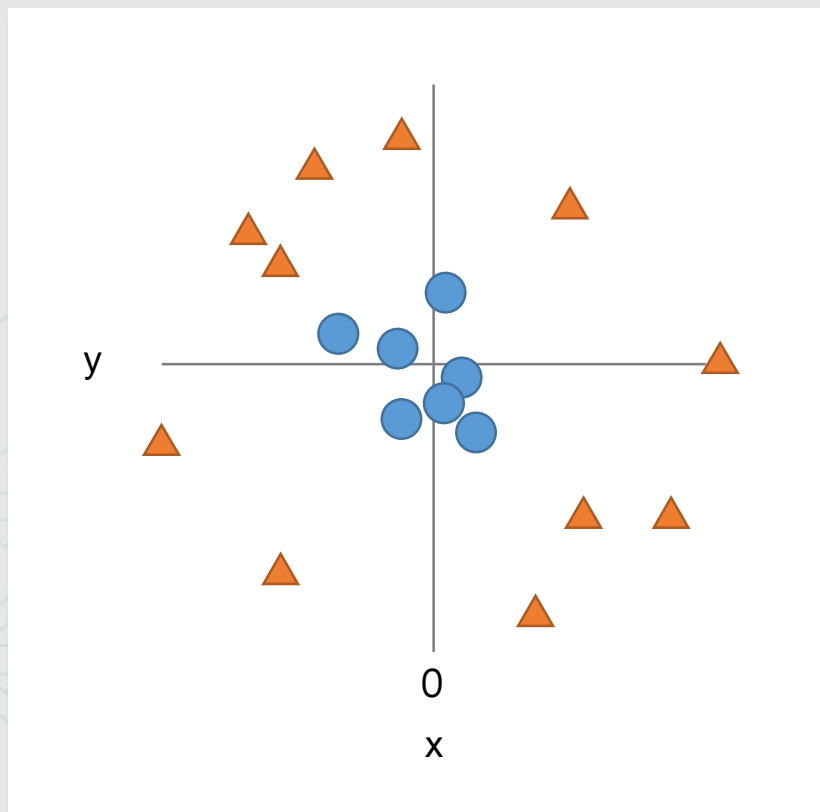




支援向量機概念



- › 下面的例子無法找到一個線性的決策邊界，但兩個類別分的有明顯的間距，看起來應該可以簡單地分開。

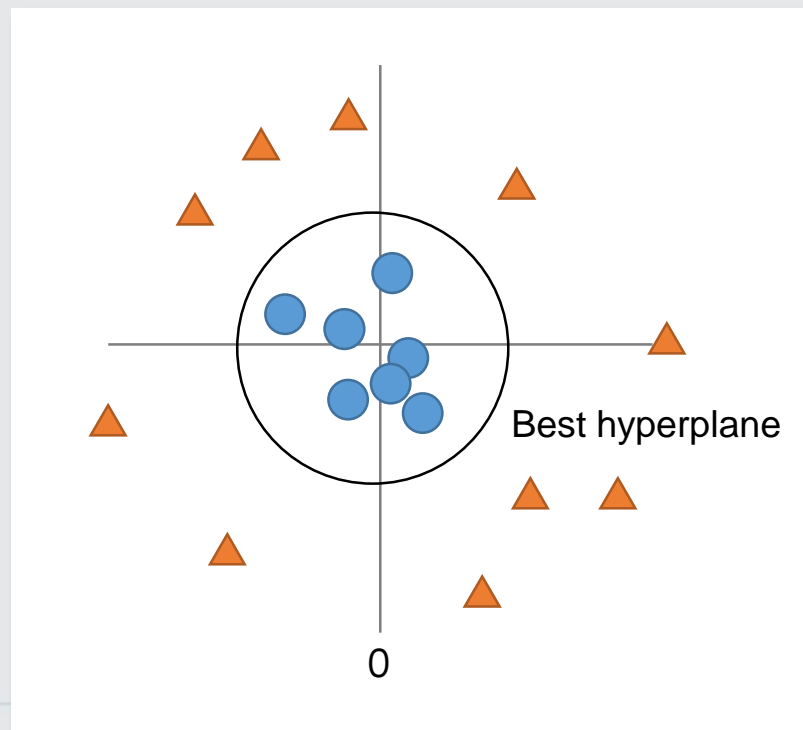
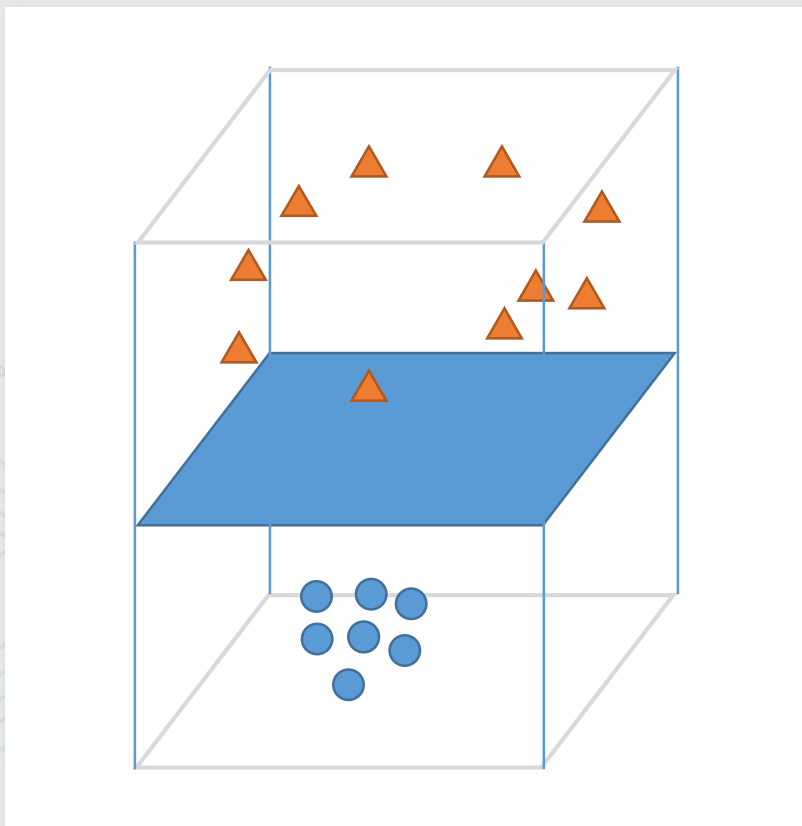


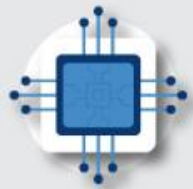


支援向量機概念



- 使用第三個維度 z ，將圖從二維提升為三維空間，即可找到好的超平面將其分開。上方往下看的視角，如右圖。

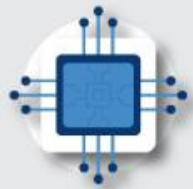




核函數



- › 可透過空間映射找到最佳的分割超平面，解決線性不可分的情形，但這種維度轉換會增加計算成本，維度轉換可以使用**向量積**來進行計算，即**核函數**。
- › 核函數採取的方式是使用**低維度特徵空間上的計算**，來避免在高維度特徵空間中**向量內積**的大量計算。



常用核函數



› 線性核函數 (Linear Kernel)

函數： $K(x,z)=x \cdot z$

用於線性可分SVM

› 多項式核函數 (Polynomial Kernel)

函數： $K(x,z)=(\gamma x \cdot z + r)^d$

是線性不可分SVM常用的核函數之一



常用核函數



- › 高斯核函數 (Gaussian Kernel) , 也稱為徑向基核函數 (Radial Basis Function,RBF)

函數： $K(x, z) = e^{-\gamma \|x-z\|^2}$

是非線性分類SVM最主流的核函數

- › Sigmoid核函數 (Sigmoid Kernel)

函數： $K(x, z) = \tanh(\gamma x \cdot z + r)$

也是線性不可分SVM常用的核函數之一



多分類支援向量機



- › 支援向量機一般以二分類為主，也就是單類別。
若要處理多類別分類，必須結合多個二分類法來達成多分類的目標。

- › SVM處理多分類的方法

One-vs-Rest

處理k類別問題，會產生k個SVM

其中第i個SVM的訓練資料中，第i類標註為 + 1，
其他類別資料都標註為 -1

One-vs-One

處理k類別問題，每兩類資料都會產生一個SVM，
會產生 $k(k-1)/2$ 個二分類SVM



支援向量機優缺點



優

- › 可解決高維問題，即大型特徵空間
- › 解決小樣本機器學習問題
- › 能夠處理非線性特徵的相互作用

缺

- › 一般SVM只支援二分類
- › 當觀測樣本很多時，效率並不是很高
- › 對非線性問題沒有通用解決方案，有時候很難找到一個合適的核函數