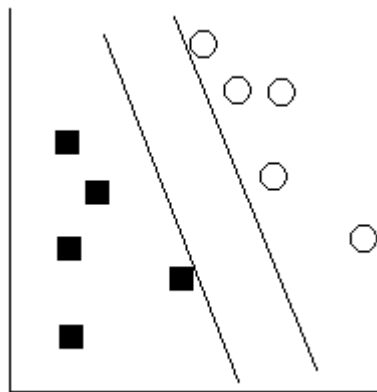


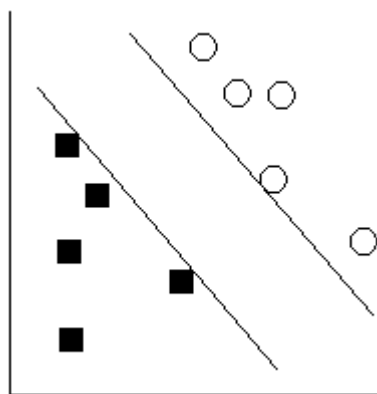
附錄二 支持向量機 Support Vector Machine (SVM) 簡介

SVM 是較新的 machine learning 技術 (Boser, Guyon, and Vapnik (1992), Cortes and Vapnik (1995)) 它使用一些策略來最大化具有不同特徵的資料中間的界限, 並針對未知資料的特徵來判斷它屬於哪個類別。SVM 已在文件分類 (Joachims (1998) Taira and Haruno (1999)) 以及名詞組標示 (Kudo and Matsumoto (2000, 20001)) 取得超越其它作法的準確性, 而近幾年應用在自然語言處理的各個議題的研究更是方興未艾, 如未知詞辨識 (unknown word guessing) (Nakagawa, Kudo, and Matsumoto (2001)) 詞性標注 (part of speech tagging) (Nakagawa, Kudo, and Matsumoto (2002), Giménez Jesús and Márquez Lluís (2004)) 句法依存關係辨識 (dependency analysis) (Kudo and Matsumoto (2000)) 詞義辨別與標注 (word sense disambiguation and sense tagging) (Cabezas, Resnik, and Stevens (2001)) 語意剖析 (semantic parsing) (Pradhan et al. (2004) Sun and Jurafsky (2004)) 等都取得不錯的成果。

SVM 是一個分類用的 machine。請參照圖 (一, 二),



圖一



圖二

SVM 找出兩種資料 (黑色方形與白色圓形) 中間的界限, 圖一, 圖二顯示出可能的兩種分割方式, 顯然的, 後者的切割方式是較佳的 (兩種資料的界線為兩平行線之中線), 而 SVM 以滿足下面條件

$$\min \Phi(\omega) = (1/2) \|\omega\|^2$$

找出最佳平面（即在線性可分的情況下，可視為解二次規畫的問題），而此可由拉格朗日乘子法（Lagrange multiplier）求解。

由於很多的問題常常並不是線性可分的（如我們的詞組切割），這個時候 SVM 在比現有資料更高的向量空間 H 使用線性分類函數 $\Phi: R^d \rightarrow H$ 將 x 對應到高維空間，便可

在此以不破壞資料特徵亦不增加複雜度的方式對其進行分類。在轉換的過程中，我們會使用一 kernel function: $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ 來實現非線性變換後的線性分類，而使用不同的 kernel function 對不同的資料會有不同的效果。以下為一個簡單的 SVM 運作方式

給定一個訓練的資料集合：

$$(x_i, y_i) \{ i = 1, 2, \dots, l; x_i \text{ 屬於 } R^n; y_i \text{ 屬於 } \{ 1, -1 \} \}$$

其中 l 為訓練之資料數， x_i 為一個 n 維向量， y_i 則是其類別（分為正類別 1 與負類別 -1 ）SVM 找到正類別與負類別中之最大的界限，即解決下面的最佳化問題的解答

$$\min_{w, b, c} (1/2) w^T w + C \sum_{i=1}^l e_i \text{ 使得} \\ y_i (w^T \Phi(x_i) + b) \geq 1 - e_i, e_i \geq 0$$

x_i 經由 Φ 函數被對應到一個更高維的向量空間 H 之後 SVM 於此找到不同類別之間最大的界限； $K(x_i, x_j)$ 為 Kernel function.