第四章　心電圖診斷理論基礎

本章節所要介紹的是和心電圖診斷系統的相關理論基礎，首先對資料叢聚的k-mean演算法做介紹，接著對特徵擷取所使用的小波轉換做介紹，最後則是本研究所使用的辨識系統支持向量機的簡介。

4.1　K-mean 分群法

為了辨認某些事物具有相似之特性，並將此等事物依照某些特性，劃分數個群集，使在同一群集內之事物具有高度之均質性，而不同群集之事物則具有相異性；若以幾何圖形表示之，則同一群集內之事物，應相互聚集在一起，而不同群集之事物則彼此遠離。

在使用分割式分群法（partitional clustering）時，我們必須先指定群聚的數目，然後藉著反覆疊代運算，逐次降低一個誤差目標函數的值，直到目標函數不再變化，就達到分群的最後結果。其中介紹最基本也最常被使用到的分群方法，就是所的k-mean分群法(k-mean clustering)，其主要目標是要在大量高維皂資料點中找出具有代表性的資料點，這些資料可以稱為群集中心(cluster centers)，然後在根據這些群中心，進行後續的處理，這些處理可以包含：

1. 資料壓縮：以少數的資料點來代表大量的資料，達到資壓縮的功能。
2. 資料分類：以少數代點來代表特定類別的資料，可以降低資料量及計算量，並可以避免雜訊的不良影響。

分割式分群法的目的是希望盡量減小每個群聚中，每一點與群中心的距離平方誤差(square error)。假設我們現在有一組包含c個群聚的資料，其中第k個群聚可以用集合G_k來表示，假設G_k包含n_k筆資料$\{x_1, x_2, x_3 \cdots x_{nk}\}$，此群聚中心為$y_k$，則該群聚的平方誤差$e_k$可以定義為：

$$e_k = \sum_i |x_i - y_k|^2 \tag{1}$$

其中x_i是屬於群集 k。

而這c個群聚的總和平方誤差E便是每個群聚平方總和，可稱為分群的「誤差函數」(error function)；

$$E = \sum_{k=1\sim c} e_k{}^2 \tag{2}$$

我們分群的方法，也就是一個最佳化的問題，也就是說我們要如何選取c個群聚以及相關的群聚中心，使得E的值為最小。

我們也可以使用另一套數學來描述，給定一組n點資料$X = \{x_1, x_2 \cdots x_n\}$，每一個都有d維，k-mean分群法的任務是找到一組m個代表點$Y = \{y_1, y_2 \cdots y_m\}$，每一點也是d維，以使下列的目標函數愈小愈好：

$$J(X,Y,U) = \sum_{i=1}^{n} |x_i - y_k|^2 \tag{3}$$

其中x_i屬於G_k而且y_k是G_k的代表點。同時在上述目標函數中，目標函數U是一個$m \times n$的分組矩陣，每一個元素值都是0或1，而且每一直行的總和等於1，如果$U(i, j) = 1$，代表第i個資料點是屬於第j組。

若要直接對上述目標函數進行最小化，是一件較因難的事，因為這個目標函數有n個限制條件，很難直接捋用傳統的方式來進行最佳化，但我們可以觀察兩個現象：

1. 當Y(群聚中心)固定時，我們可以很容易找到所對應的U(資料分群方式)，使得 $J(X,Y,U)$ 的值最小。此最佳的U值，即代表最佳的分群方式，因此若將每因此若將每個資料點歸類到距離最近的代表點，即可完成此目標。
2. 當U（資料分群方式）固定實，我們也可以很快地找到對應的Y（群中心），使 $J(X,Y,U)$ 的值為最小。此最佳的Y集合，代表每個代表點至其組員的距離平方和為最小，因此此最佳的Y集合即是每一群的平均點。

在演算法開始進行前，我們必須事先決定好預期分群的群聚數目。假設預期的分群群聚數目為c，則根據上述觀察，我們可以經由下列步驟來進行k-means 分群法：

1. 隨機選取c個資料點，將之分別視為c 個群聚的群中心，這就是Y。
2. 由固定的Y，產生最佳的U。換句話說，對每一個資料點x，尋找與之最接近的群中心，並將x加入該群聚。
3. 計算目標函數 $J(X,Y,U)$，如果保持不變，代表分群結果已經穩定不變，所以可以結束此疊代方法。再由固定的U，產生最佳的Y。跳回第2個步驟。

在上述方法中，我們是先找群中心，再開始反覆疊代的過程。事實上，我們也可以先進行任意分群，然後再進行反覆疊代的過程，得到的結果應該很類似，如圖4-1所示(a)為4筆測試資料，(b)選擇2筆資料作為群聚中心，(c)

計算資料點和群聚中心的距離，並且作類別的歸類，(d)持續疊代，直到可容許的誤差範圍，最後可以發現資料已被分門別類。

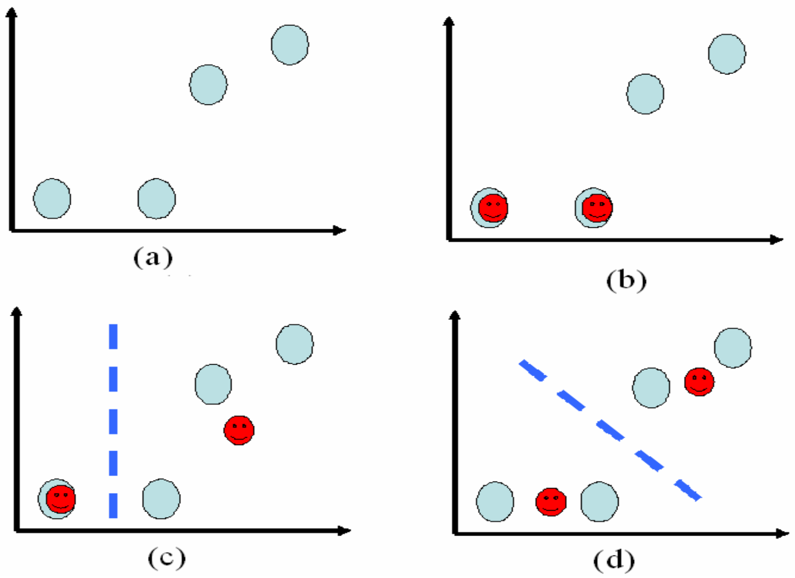

圖 4-1　k-mean 分群示意圖

4.2　小波轉換簡介

約在1986年，一群在調和分析領域的數學家，發現了小波函數。剛開始時，只有屬於數學、物理和地震波的研究者，注意到這個新發現。而小波函數和DSP之間相互的關聯也隨之被發現，很快地，其他領域的專家、科學家與工程師們，都加入了這個發展和研究的行列。也因為這些時領域學者的共同開發，小波函數從一開始就是理論與應用並行，而且發展快速。

廣義來說，小波函數就是那些在時間領域與頻率領域都具有局部性的函數。據海森堡測不準定理可知，任何函數的時間反應與頻率不可能同時都很狹窄。而從應用的角度來說，就是任何濾波器，無法同時在時間領域和頻率領域上都具有高準確度。小波函數也像其它函數一樣受到這種理論的限制。

然而小波函數卻容許在此限制下盡可能地設計出高效率的濾波器。

小波轉換與數位信號處理在Mallat[22]及Daubechies[32]結合下，已有所改變。發展至今，小波轉換不論在數學、電腦科學、物理或工程上皆有舉足輕重的地位。就本研究建構的分析工具所採用的小波理論基礎，是由Joseph Fourier在1807年提出的頻域分析理論（Fourier synthesis）開始的，他主張所有的波形都會由sine以及cosine波組成，而後在1910年Haar發展出Haar function，這個函數是由方波所組成，而在1938年的時候Littlewood-Paley對Fourier series建立L-P理論。到了1970年代，主要是Hardy space theory的發展。一直到了1982年，Stromberg提出第一個正交小波，其特性為具有良好的時頻局部定位性，同時也證明小波的存在，因此在1984年的時候Morlet將小波的概念引入訊號分析之中，並與Grossmann定義連續小波轉換，也正式將之命名為「wavelet」，接著就進入了小波轉換研究的黃金時期。於是在1985年，Meyer建立Meyer wavelet，具正交易衰減等特性的平滑小波，其傅利葉轉換為緊湊支撐（compact support），且屬於連續函數空間。在同年，Tchamitchian亦提出雙正交小波。而在1986年的時候，Battle、Lemarie´提出Battle- Lemarie´wavelet，為將B-spline正交後所得，對時領域的支撐為infinity，具有指數衰減的效應。接著在1987年時，Mallat便證實可透過多重解析度分析來產生小波基底。在1988年的時候，Daubechies提出Daubechies wavelet，為有限支撐，且具有良好的時頻局部定位性，而本論文亦採用了Daubechies wavelet為小波壓縮的基礎演算法。隔年，Coifman跟著提出一組包含vanishing moment的正規化尺度函數及小波函數建造方式。而在1991年，Wickerhauser提出小波包（wavelet packet）的算法。隨後到了1993年，Goodman提出多重小波（multi-wavelet）的概念，並與Lee、Tang等拓展Mallat的正交小波，而得出多重小波。在1996年的時候Chui-Lian研究出orthonormal

多重小波。接著在1998年，Jiang提出時頻解析度最佳化的正交多重小波。到了現今，有許多的研究都廣泛的運用小波轉換，例如新一代的JEPG2000壓縮，心電圖與心音圖的應用等等。

小波分析基本理論在於以數學上的函數轉換的方式，將資料分成幾個頻率分量。若傳統的頻譜分析，是以傅利頁轉換（Fourier transform）為數學上的理論依據。如果訊號的性質並未隨時間變化而發生很大變動，則稱這種訊號屬於固定性(stationary)的訊號。然而大多數的情形下，訊號中屬於固定性的分量並非真正希望量測到的，反而是非固定性的分量才是要研究的部分，例如一些不定時產生的尖銳形狀（spike），或是不規則的漂移。這個時候傅利葉分析在應用上就顯得不夠充分。那是因為若以傅利葉分析轉換至頻域，將會失去時域之相關訊息。

D.Gabor為了修正這個缺點，提出將訊號切割成許多小段，逐段作傅利葉分析以得到較多時域上的訊息，這就是所謂的短時間傅利頁轉換（Short-Time Fourier Transform，STFT）。然而這個方法也同樣有其限制，就是切段的長度究竟該如何選擇。如果訊號的頻率很低，則訊號長度必須拉長；如果要分析的頻率很高的訊號，當然必須選擇較短的訊號視窗長度（window length）。但是如果事先無法得知要分析的訊號究竟屬於那個頻帶，這種方法也無法使用。為了解決這個問題小波分析就推導出來了。

所謂小波，是指一個定義為有限長度且平均值為零的波形。藉著拉伸（stretch）或壓縮（compress）小波函數之後，逐段平移（shifting），展開被分析的訊號，如圖4-2所示：

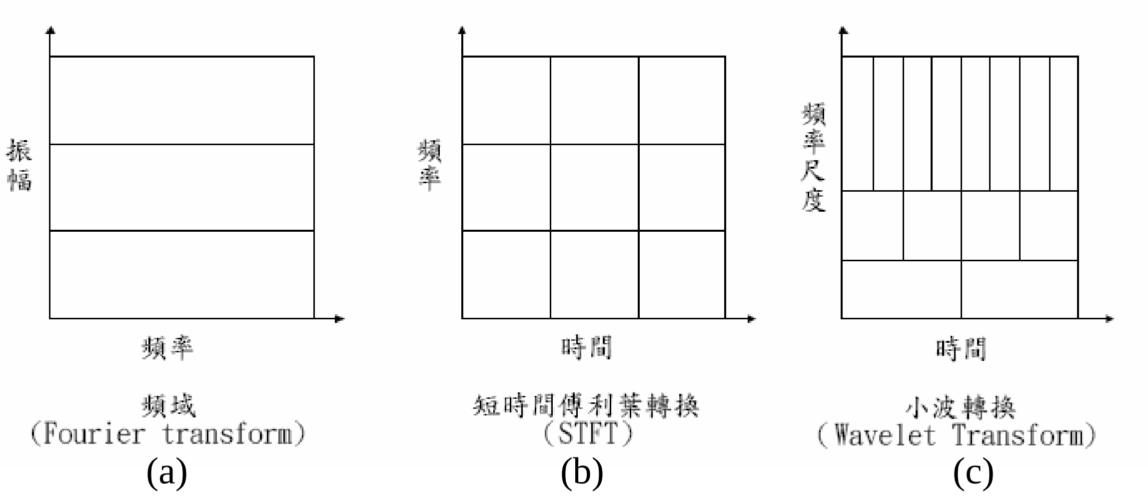

圖 4-2　(a)傅利頁轉換、(b)短時間轉換、(c)小波轉換

小波轉換的原理和傅利頁轉換很相似，差異在於小波轉換只是以小波函數經過一定比例之拉伸或壓縮之後，所形成的一系列函數作為展開之基底函數。可以數學式表示如下：

$$W_\psi f(b,a) \equiv \int_{-\infty}^{\infty} f(t)\overline{\psi_{b,a}(t)}dt \tag{1}$$

其中，$\psi(t)$為小波函數：

$$\psi_{b,a} = a^{-1/2}\psi\left(\frac{t-b}{a}\right) \qquad a > 0 \tag{2}$$

其中 a 為壓縮或拉伸的比例，b 為整段訊號的平移。

小波轉換是利用多重解析度(Multi-resultion)對訊號作分解及運算，實際應用在很多的領域像是訊號壓縮、影像壓縮，雜訊消除、偵測自我的相似性、訊號的不連續點或轉折點等，其中訊號的不連續點或轉換點的地方，經過小波轉換後就會呈現明顯不一樣的特徵，其多重解析度的觀念如圖4-3所示，先對訊號S作小波一階段濾波得 A_1 和 D_1，再對訊號 A_1 作小波濾波得 A_2 和 D_2，其餘依此類推。這邊注意的是以後 D_2 的細節係數其實是訊號S經過一個帶通濾波器（Bandpass Filter），例如 D_2 是訊號S先經過低通濾波器得 A_1，A_1 再經過高通濾波器得 D_2，依此類推得 D_2 以後更高尺的 D_n。

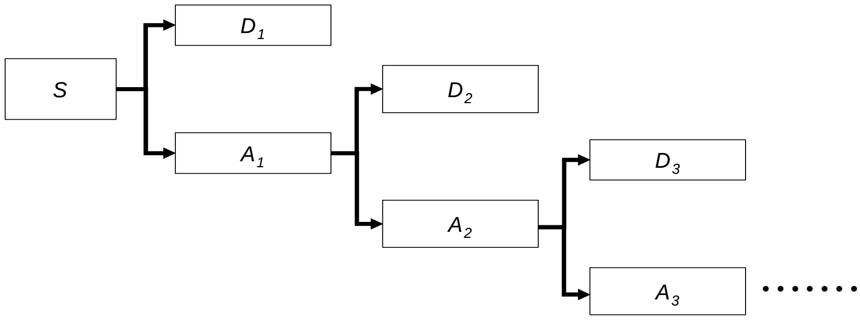

圖 4-3　訊號 S 作小波轉換多重分解示意圖

4.3　支持向量機簡介

SVM是1998年以統計理論為基礎所提出的機器學習理論[33][34][45]，它不像是傳統的類神經網路方法是基於經驗風險最小化（Empirical Risk Minimization Principle, ERM）而是一種結構風險最小化原理（Structural Risk Minimization Principle, SRM）的統計學習理論，用於分類與回歸問題。SRM使 VC（Vapnik Cherovnenkis）維數的上限最小化，這使得SVM方法比類神經網路的具有更好的泛化能力。這套演算法可以歸納成如下：他是在透過核函數的映射，將原本非線性的特徵空間mapping到一個高維的核特徵空間中去訓練一個線性學習器。（而且在這個高維空間中我們並不需要知道資料當中的特徵向量有什麼性質。）而由於Generalisation Theory 而有對偶表示的形式而成為有效率的演算法。而其中有幾個條件和理論更完善了『支持向量機』的架構。

例如VC維的理論，使得有限集合的分析可以拓展到無限假設集的領域。Lagrange理論則可以將學習問題轉化成沒有不等式約束的最佳化問題。再經由Kuhn-Tucker條件的補充，將這個方法進一步推廣到具有不等式條件亦可使用。而Mercer條件使得對應的最優化問題變成一個沒有局部最小值的

凸二次函數問題。

通過以上的完整學習演算法架構，SVM可以自動尋找出那些對分類有較好區分能力的支援向量，並由此構造出的分類器可以最大化類與類的間隔，因而有較好的適應能力和較高的分辨率。該方法只需要由各類域的邊界樣本的類別來決定最後的分類結果。而支持向量機由於其各種快速演算法（例如：SMO）的出現，使得其在圖像人臉識別、語音/文字辨識、以及生物基因工程、回歸分析等領域都有很大的進展與應用。

SVM可由最簡單的二元分類開始做起如圖4-4所示，然後推廣至多類分類。其原理都是根據二分法的原理開始的，故介紹的時候以二分法為主。

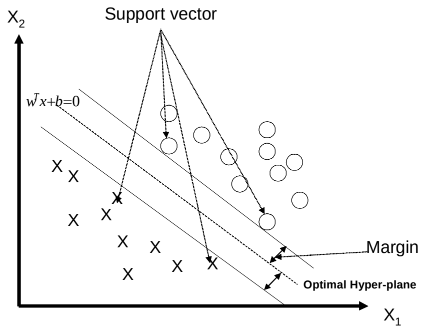

圖 4-4　　SVM 二元分類圖

此時輸入向量 ($\vec{x}_i$) 為 d 個維度，

$\vec{x}_i$ =（x_1, x_2, . . . ， x_d），　　(1)

而由於是二分法，所以會有兩個 lavel(1 or －1)。

y_i = {+1, –1} (i =1, 2, ..., N)。　　(2)

而由於我們要決定支持向量的最佳分類面（optimal separating

hyperplane），所以會有權重向量weight vector 和一個偏移值bias分別表示為$\vec{w}$和b,

而最佳分類面此時就可以表示為$\vec{w}\cdot\vec{x}_i$+b = 0，而此時所有的分類就可以表示為$y_i(\vec{w}\cdot\vec{x}_i+b)\geq 1$ (i =1, 2, ..., N) ，而為了使得分類間距(2/$\|\vec{w}\|$)為最大，所以我們必須使得$\|\vec{w}\|$/2 為最小，根據Lagrange理論(引入Lagrange乘子Lagrange multiplier α_i和Lagrange函數Ld(α)和Kuhn-Tucker條件的補充可得：

$$L_d(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(\vec{x}_i\cdot\vec{x}_j) \quad (i=1, 2, ..., N) \tag{3}$$

在約束條件(constraint) $\sum_{i=1}^{N}\alpha_i y_i = 0$ 和 $\alpha_i \geq 0$。下，所有滿足 $\alpha_i > 0$ 的 input vector 就稱為支援向量（Support Vectors）。經由 Training 就可以找到理想的 Lagrange multiplier α_0 和偏移值 b_0。

而若是輸入資料的噪擾很高（high noise level）則我們必須增加一個參數，稱為間隔鬆弛變量(Margin Slack Variable) ξ_i (i =1, 2 ..., N)，來調整期誤差值能夠小於某個範圍以下。此時，原本的式子可表示為 $y_i(\vec{w}\cdot\vec{x}_i+b)\geq 1-\xi_i$。而最佳分類面可表示為如下：

$$J(\vec{w}, \xi_i) = \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{N}\xi_i \quad (i=1, 2, ..., N)， \tag{4}$$

其中 C 是一個衡量 Training Error 和 VC 維度大小的損失參數（penalty parameter)，所以 Lagrange 函數為：

$$\mathrm{Ld}(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(\vec{x}_i\cdot\vec{x}_j) \tag{5}$$

此時

$$0 \le \alpha_i \le C, \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \ (i = 1, 2, ..., N)。 \tag{6}$$

而求解Lagrange函數（Lagrangian）Ld(α) 之最大值就可以得到最佳分類面（optimal separating hyperplane）。但是由於input data在原本的空間中式呈現非線性分佈的，所以我們此時必須透過核函數(kernel functions) 將輸入資料映射到一個高維度的線性空間，然後再來求取最佳化平面的最佳化參數值。而這些核函數比較著名的如下：

linear: $$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j \ , \tag{7}$$

polynomial: $$K(\vec{x}_i, \vec{x}_j) = (\gamma\ \vec{x}_i \cdot \vec{x}_j + coef0)^d, \tag{8}$$

RBF: $$K(\vec{x}_i, \vec{x}_j) = exp(-\gamma(\vec{x}_i - \vec{x}_j)^2), \tag{9}$$

sigmoid: $$K(\vec{x}_i, \vec{x}_j) = tanh(\gamma\ \vec{x}_i \cdot \vec{x}_j + coef0), \tag{10}$$

其中的 d， γ ， coef0 皆為常數參數值。

支持向量機理論只考慮高維特徵空間的點積運算$K(x_i,x_j)=\Phi(x_i) \cdot \Phi(x_j)$，而不直接使用函數Φ，從而巧妙地解決了因Φ未知而W無法顯式表達的問題，稱$K(x_i,x_j)$為核函數，已經証明，只要滿足Mercer條件的對稱函數即可作為核函數。

對於兩類問題分類,存在線性可分和線性不可分兩種支持向量機,但是在實際中，為了將兩類模式儘可能分類開來，一般要構造非線性可分的支持向量機。一個複雜的模式識別分類問題，在高維空間比低維空間更容易線性可分。支持向量機就是首先透過核函數把訓練樣本中的低維數據映射到高維特徵空間，然後在高維特徵空間構造一個最佳分類平面。由於構造的核函數滿

足Mercer條件，所以在訓練中只需考慮核函數K，而不必明確知道低維向高維的映射函數Φ。

一個二元函數K(x,y)通常稱為是一個核函數(簡稱核)，給定K(x,y)，若有實數λ和非零函數Ψ(x)成立

$$\int_a^b K(x,y)\Psi(x)dx = \lambda\Psi(y) \tag{11}$$

稱λ為核的一個特徵值，稱Ψ(x)為核的關於特徵值λ的一個特徵函數。
對稱半正定的連續核稱為 Mercer 核。

關於Mercer核有如下定理：
Mercer 定理：Mercer 核 K(x,y)可以展開成一致收斂的函數

$$K(x,y) = \sum_i \lambda_i \Psi_i(x)\Psi_i(y) \tag{12}$$

其中 λ_i，Ψ(x)分為核 K(x,y)的特徵值和特徵向量，它們的個數可能無限或無窮。

Mercer 核很多，如徑向基函數核，雙曲正切函數核等。由已知的Mercer核經過某些運算可以生成新的Mercer核，特別是，由點積定義的核必是Mercer核：

$$K(x,y) = (\phi(x) \bullet \phi(y)) \tag{13}$$

特徵映射和特徵空間稱由 Mercer 核的特徵函數張成的函數集為特徵空間，記為 F。原樣本空間記為 X。如果我們做如下樣本空間 X 到特徵空間 F 的非線性映射Φ：

$$\Phi(x) = (\sqrt{\lambda_1}\Psi_1(x), \sqrt{\lambda_2}\Psi_2(x), \cdots\cdots\cdots \sqrt{\lambda_3}\Psi_3(x)\cdots\cdots) \tag{14}$$

則有：

$$K(x,y)=\sum_{i}\lambda_i\Psi_i(x)\Psi_i(y)=\left(\phi(x)\bullet\phi(x)\right) \tag{15}$$

從此可以看出：當我們樣本空間通過非線性映射映入特徵空間時，如果只用映射點積，則可以用相對應的核函數來代替，而不需要知道映射的顯式表述式。這是從線性支持向量機到非線性支持向量機的關鍵一步。

非線性支持向量機在特徵空間F中應用線性支持向量機的方法，分類決策函數式變為：

$$y=Sgn(w*\phi(x)+b)=Sgn(\sum_{sv}\alpha_i y_i(\phi(x)\phi(x))+b) \tag{16}$$

這裡 φ(x) φ(x)代替了 x 和 xi 因此計算過程相同。

考慮 Mercer 定理，可以化簡為

$$y=Sgn(\alpha_i y_i K(x,x_i)+b) \tag{17}$$

這就是非線性支持向量學習機的最終分類決策函數。雖然用到了特徵空間及非線性映射，但實際計算中並不需要知道他們的顯示表述。只需求出支持向量及其支持的 α 和b值，通過核函數的計算，即可得到原空來樣本空間的非線輸出值。

$$y(x)=\sum_{i=1}^{\infty}\lambda_i\psi_i\Phi_i(x)+b=\sum_{j=1}^{n}\alpha_j y_j K(x,x_j)+b \tag{18}$$

$$\psi=\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i)=(\psi_1,\psi_2,\cdots,\psi_{i,\cdots}) \tag{19}$$

其中 K 可以展開成

$$K=(K(x_i,x_j))=V\Lambda V^t=(v_1\cdots\cdots v_n)\begin{bmatrix}\lambda & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \lambda_n\end{bmatrix}\begin{bmatrix}v_1^t\\ \vdots\\ v_n^t\end{bmatrix} \tag{20}$$

其中特徵向量為：

$$\phi(x_i) = (\sqrt{\lambda_1}v_{1i}, \sqrt{\lambda_2}v_{2i}, \cdots\cdots\cdots, \sqrt{\lambda_n}v_{ni})^T, \lambda_t \geq 0 \quad for \quad \forall k \tag{21}$$

利用兩個常用的 kernel 來證明 Mercer 定理：

範例一：多項式展開

$$x, z \in R^2, \phi(x), \phi(z) \in R^6$$

$$x = [x_1, x_2], z = [z_1, z_2]$$

$$\phi(x) = \left[1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right]$$

$$\phi(z) = \left[1, \sqrt{2}z_1, \sqrt{2}z_2, z_1^2, z_2^2, \sqrt{2}z_1z_2\right]$$

$$\phi^T(x)\phi(z) = x_1^2z_1^2 + x_2^2z_2^2 + 1 + 2\left(x_1z_1 + x_2z_2 + x_1z_2x_2z_1\right)$$

$$\begin{aligned}
\left(x^Tz+1\right)^2 &= \left[\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + 1\right]\left[\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + 1\right] \\
&= \left[x_1z_1 + x_2z_2 + 1\right] \cdot \left[x_1z_1 + x_2z_2 + 1\right] \\
&= \left(x_1z_1 + x_2z_2 + 1\right)^2 \\
&= x_1^2z_1^2 + x_2^2z_2^2 + 1 + 2\left(x_1z_1 + x_2z_2 + x_1z_2x_2z_1\right)
\end{aligned}$$

範例二：RBF 展開

$$e^{-\frac{\|x-z\|^2}{2\sigma^2}} = e^{-\frac{x^2-2xz+z^2}{2\sigma^2}} = e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{z^2}{2\sigma^2}} \cdot e^{\frac{xz}{\sigma^2}}$$

$$= e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{z^2}{2\sigma^2}} \cdot \left[1 + \frac{xz}{\sigma} + \frac{1}{2!}\left(\frac{xz}{\sigma^2}\right) + \cdots\right]$$

$$= e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{z^2}{2\sigma^2}} \cdot \{[1, \frac{x}{\sigma}, \frac{1}{\sqrt{2!}}\left(\frac{x}{\sigma}\right)^2, \cdots\cdots] \cdot [1, \frac{z}{\sigma}, \frac{1}{\sqrt{2!}}(\frac{z}{\sigma})^2, \cdots\cdots]\}$$

$$= e^{-\frac{x^2}{2\sigma^2}}[1, \frac{x}{\sigma}, \frac{1}{\sqrt{2!}}\left(\frac{x}{\sigma}\right)^2, \cdots\cdots] \cdot e^{-\frac{z^2}{2\sigma^2}}[1, \frac{z}{\sigma}, \frac{1}{\sqrt{2!}}(\frac{z}{\sigma})^2, \cdots\cdots]$$

$$= \Phi(x) \cdot \Phi(z)$$