群組分析：Fisher 與 Mahanlanobis 的距離概念

January 6, 2019

群組分析 (Discriminant Analysis) 的用途非常廣，譬如，網路書店 [1] 想對其一百萬會員促銷一本新書。在成本考量下，希望能針對可能購買該書的會員做促銷的動作。但如何得知哪些會員購買的可能性較高？哪些會員根本不可能購買呢？雖然該網站已有所有會員的資料及過去購買書籍的紀錄，但對於特定的一本書仍缺乏足夠有利的資訊區分出購買群。

於是該網路書店先抽樣選擇 1000 名會員進行銷售，結果有 83 名會員買了這本書。根據這個結果及這 1000 人的相關資料，網路書店便可以進行群組分析，建立群組分析模式，最後對於其他大部分的會員 (999,000) 進行群組預測。當然促銷時，只需針對可能購買的會員進行。如此可以集中「火力」，節省大量成本。

本章將學到關於程式設計
群組資料的繪製技巧。

〈本章關於 MATLAB 的指令與語法〉
指令：biplot, car2pol, pol2cart

1 背景介紹

1.1 Fisher's Approach

進行群組區別時，通常是從自變數 $X_1, X_2, \cdots, X_N$ 的資料去判斷其所屬的群組。之前，必須先進行群組分析，以確立區別的準則。換句話說，從已知的自變數資料及所屬的群組找出之間的關係。如果可以清楚的描述這項關係，就等於掌握了群組的特性。群組與自變數的關係一般稱爲「區別函數」Discriminant Function。自變數間不同的組合可以構成不同的「區別函數」，但區別的效果（或稱區別率或鑑別率）不同。什麼樣的組合才能達到最佳的鑑別程度呢？

首先必須瞭解自變數間線性組合的幾何意義；假設 N 個自變數，線性組合成一個變數，等於從 N 度空間投射到 1 度空間上，或說從 N 個座標軸簡化爲一個座標軸。理論上，問題變簡單了，但也同時因空間的縮減，損失了若干訊息。即便如此，在比較小的空間裡，一樣可以找到最佳的「觀察角度」，諸如此類的區別函數於焉產生。

舉一個簡單的例子，有兩個群組，彼此交錯，其分佈及重疊情形如圖 1 所示。假設鑑別函數

$$f(X_1, X_2) = X_1 \tag{1}$$

即群組鑑別的依據完全取決於變數 X_1 的值。其鑑別能力可以從圖右下角的兩個交錯的常態分配圖清楚看出。如果鑑別函數定義爲

$$f(X_1, X_2) = X_2 \tag{2}$$

即群組鑑別的依據完全取決於變數 X_2 的值，則其鑑別能力可以從圖左上角的兩個交錯更緊密的常態分配圖看出，很明顯的，其鑑別能力更差了。這也說明自變數間的不同組合可以決定群組的鑑別程度。如圖中交錯比較小的兩個常態分配圖，不僅平均數的距離比較遠，連變異數也比較小，自然容易分辨。就好像遠看兩棟相鄰的建築物，從不同的角度可以看到不同的形狀，其間的差距也隨之不同。這也說明看待一件事情，當從不同的角度觀察時，常會呈現出不同的面貌。

Fisher[1] 提出自變數的組合方式（幾何：觀察的角度），希望造成（看到）群組間的「距離」最大，群組內的「分散」程度最小。若不能兩全，則取其比例最大者。以專業的術語來描述「距離」與「分散」如；

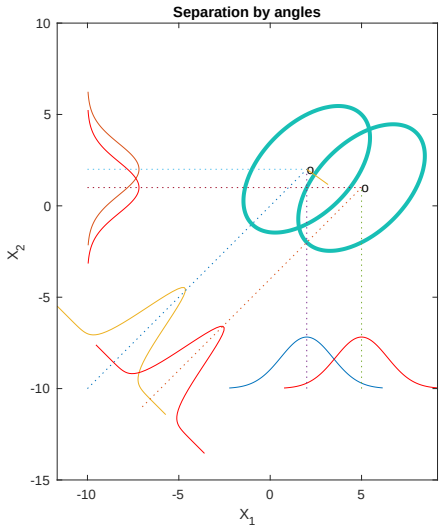

圖 1: 群組區別與觀看角度的關係

Maximize the ratio of the across-group sum of squares to the within-group sum of squares for the combination

或

$$\max_{\text{組合係數}} \frac{\text{across-group sum of squares}}{\text{within-group sum of squares}} \tag{3}$$

從一維的空間座標來看多維度的群組，不同的座標角度會看到不同的群組分佈情況，圖 1 便提供了三個角度。從那個角度可以看到群組間距最大? 同時組內的變異最小? Fisher 用了 across-group sum of squares 來量化群組間距，以 within-group sum of squares 量化組內的變異。數學上的定義如下:

假設有兩個群組，變數 t 爲 N 個自變數組合後的變數（或稱爲鑑別函數），即

$$t = k_1 X_1 + k_2 X_2 + \cdots + k_N X_N = \mathbf{x^T k} \tag{4}$$

在幾何意義上，稱爲投射（mapping），將處在 N 度空間上的點投射到 1 度空間上（變數爲 t），對 N 度空間上的群組的分辨能力，只剩下一個角度，就是與新的座標軸 t 垂直的方向，這個幾何上的意義可以從圖 2 與 3 看出來。

從與向量 **k** 垂直的方向來看群組內的聚散情況，稱爲 within-group sum of squares，定義爲

$$SS_W = \sum_i (t_{i(1)} - \bar{t}_{(1)})^2 + \sum_i (t_{i(2)} - \bar{t}_{(2)})^2 \tag{5}$$

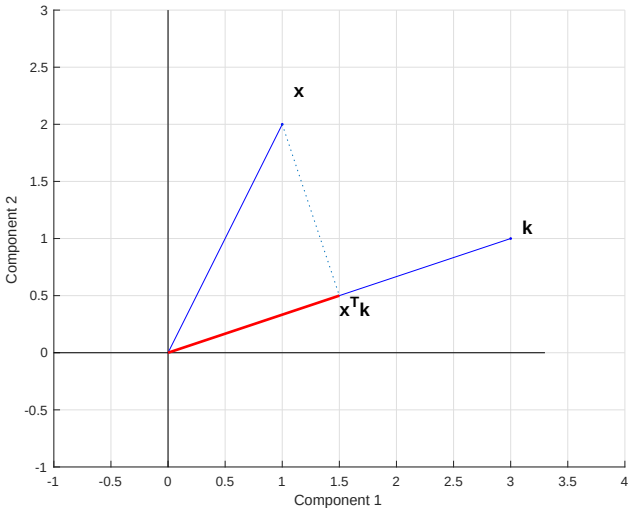

圖 2: 向量間的垂直投射：$\mathbf{x^T k}$ 的幾何意義（$||\mathbf{k}|| = \mathbf{1}$）。

其中 $\bar{t}_{(1)}$ 及 $\bar{t}_{(2)}$ 分別代表組合變數中屬於群組 1 及群組 2 的平均值。$t_{i(1)}$ 及 $t_{i(2)}$ 則分別代表其第 i 個樣本值。假設群組 1 共有 n_1 個樣本，群組 2 共有 n_2 個樣本。上式可以寫成

$$\begin{aligned} SS_W &= \sum_i (\mathbf{k^T x_{i(1)}} - \mathbf{k^T \bar{x}_{(1)}})^2 + \sum_{\mathbf{i}} (\mathbf{k^T x_{i(2)}} - \mathbf{k^T \bar{x}_{(2)}})^2 \\ &= \mathbf{k^T} \left(\sum_{\mathbf{i}} (\mathbf{x_{i(1)}} - \mathbf{\bar{x}_{(1)}})(\mathbf{x_{i(1)}} - \mathbf{\bar{x}_{(1)}})^{\mathbf{T}} + \sum_{\mathbf{i}} (\mathbf{x_{i(2)}} - \mathbf{\bar{x}_{(2)}})(\mathbf{x_{i(2)}} - \mathbf{\bar{x}_{(2)}})^{\mathbf{T}} \right) \mathbf{k} \\ &= \mathbf{k^T (W_1 + W_2) k} \\ &= \mathbf{k^T} \left((\mathbf{n_1 - 1}) \mathbf{C_1} + (\mathbf{n_2 - 1}) \mathbf{C_2} \right) \mathbf{k} \end{aligned} \tag{6}$$

其中 $\mathbf{\bar{x}_{(1)}}, \mathbf{\bar{x}_{(2)}}$ 分別代表群組 1 與群組 2 的樣本平均值，C_1 與 C_2 代表群組 1 與群組 2 的共變異矩陣（又稱爲 within-group covariance matrix），反映了群組內 (within-group) 樣本的分佈情況。在此使用了 unbiased 的估計式。

另一方面，式 (4) 代表群組間聚散的情況的 across-group sum of squares，定義爲

$$SS_A = n_1(\bar{t}_{(1)} - \bar{t})^2 + n_2(\bar{t}_{(2)} - \bar{t})^2 \tag{7}$$

其中 $\bar{t}$ 代表組合變數的整體樣本平均值，樣本大小 n_1 與 n_2 反應其比例上的權重。上式可以繼續推演爲（習題 1）

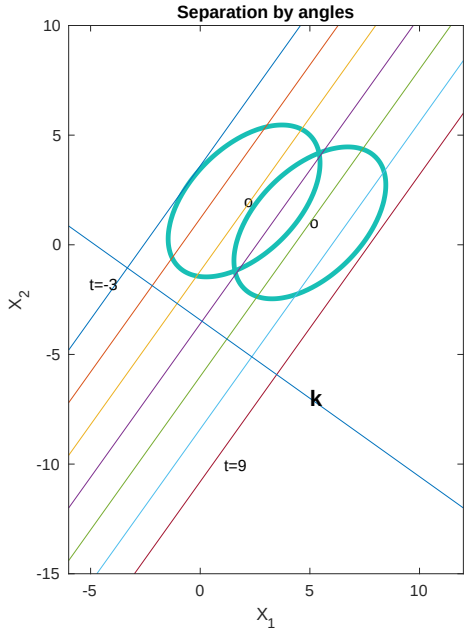

圖 3: 2 度空間到 1 度空間的投射

$$\begin{aligned} SS_A &= n_1\mathbf{k^T}(\bar{\mathbf{x}}_{\mathbf{(1)}} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\mathbf{(1)}} - \bar{\mathbf{x}})^{\mathbf{T}}\mathbf{k} + \mathbf{n_2}\mathbf{k^T}(\bar{\mathbf{x}}_{\mathbf{(2)}} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\mathbf{(2)}} - \bar{\mathbf{x}})^{\mathbf{T}}\mathbf{k} \\ &= \mathbf{k^T}\left(\mathbf{n_1}(\frac{\mathbf{n_1}}{\mathbf{n_1}+\mathbf{n_2}})^{\mathbf{2}}\mathbf{d}\mathbf{d^T} + \mathbf{n_2}(\frac{\mathbf{n_2}}{\mathbf{n_1}+\mathbf{n_2}})^{\mathbf{2}}\mathbf{d}\mathbf{d^T}\right)\mathbf{k} \\ &= \lambda\mathbf{k^T}\mathbf{d}\mathbf{d^T}\mathbf{k} \end{aligned} \tag{8}$$

其中 $\bar{\mathbf{x}}$ 代表整體樣本的平均數，而 $\mathbf{d} = \bar{\mathbf{x}}_{\mathbf{(1)}} - \bar{\mathbf{x}}_{\mathbf{(2)}}$ 。Fisher 對於自變數的最佳組合來自下列的最佳化問題:

$$\max_{\mathbf{k}} \frac{\mathbf{k}^T\mathbf{d}\mathbf{d}^T\mathbf{k}}{\mathbf{k}^T C_W \mathbf{k}} \tag{9}$$

目標函數爲 SS_A 與 SS_W 的比例，其中 C_W 一般稱爲 Pooled within-group covariance matrix，其不偏估計 (unbiased estimate) 定義爲

$$C_W = \frac{(n_1-1)C_1 + (n_2-1)C_2}{n_1+n_2-2} \tag{10}$$

透過目標函數一次導數爲零的必要條件，上式的最佳解爲（習題 2）:

$$\mathbf{k}^o \propto C_W^{-1}\mathbf{d} \tag{11}$$

$\mathbf{k}^0$ 代表在所有可能的一度空間裡，提供同時兼顧群組間距與群組內聚合性的最佳角度。值得注意的是，Fisher 的觀念只提出最佳的鑑別視野，並未明確的指出群組的分界線在哪裡，所以式 (11) 只是個方向，還不能當作群組的鑑別條件。Mahalanobis 直接從群組分界線切入來看這個問題，其結果不但與 Fisher 的鑑別觀念不謀而合，並且也得到一組分界線的方程式。

範例 1 式 (6) 的 C_1 與 C_2 分別代表群組 1 與群組 2 的共變異矩陣，計算式 (10) 的 Pooled within-group covariance matrix。從網站上下載資料 Book_1.txt [1，chap 12]，根據資料的描述計算 C_W。

通常資料處理前需經過前置作業（Pre-processing），將屬於同組的資料集合在一起，方便後續的處理，一方面減輕重複處裡的負擔，一方面也讓程式簡潔些。譬如這個資料檔的第四欄是購買與 (1) 否 (0)，也就是群組的屬性。

```
D=load('BOOKS_1.txt'); % 第 2, 3 欄爲自變數資料
X1=D(D(:,4)==0, 2:3); % Group 1
X2=D(D(:,4)==1 ,2:3); % Group 2
n1=size(X1,1); n2=size(X2,1); % Group size
Cw=((n1-1)*cov(X1)+(n2-1)*cov(X2))/(n1+n2 - 2)
```

計算得到的 C_W，大約是

$$C_W = \begin{bmatrix} 63.2366 & 0.1644 \\ 0.1644 & 0.4308 \end{bmatrix}$$

1.2 Mahalanobis's Method

Mahalanobis 對於自變數的組合方式有不一樣的看法；找出兩群組間等距離的軌跡 (locus) 函數，認爲這是對群組做最適當的切割。所謂「群組的距離」必須先定義。假設 $\mathbf{x}$ 爲任意一點，其與第 k 個群組的距離定義爲:

$$D_k^2 = (\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k})^\mathbf{T} \mathbf{C}_\mathbf{W}^{-\mathbf{1}} (\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k}) \tag{12}$$

其中 $\bar{\mathbf{x}}_k$ 代表第 k 個群組的中心點，而 C_W^{-1} 的介入考慮了群組內的變異情形（在此假設所有群組的共變異矩陣相同，都等於群組間的 Pooled within-group covariance matrix C_W，

即 $C_1 = C_2 = \cdots = C_k = C_W$）。這也可以解釋爲「加權的歐幾理德距離（Weighted Euclidian Distance）」。在只有兩個群組的條件下，等距離的軌跡函數必須滿足以下的條件:

$$D_1^2 = D_2^2 \tag{13}$$

經過推導後，上式變爲（習題 3）

$$2\mathbf{x^T C_W^{-1} d} = \mathbf{\bar{x}_{(1)}^T C_W^{-1} d} + \mathbf{\bar{x}_{(2)}^T C_W^{-1} d} \tag{14}$$

其中 $C_W^{-1}\mathbf{d}$ 與 Fisher 提出的組合係數 $\mathbf{k}$ 成正比（參考式 (11)），直接將 $\mathbf{k} = \mathbf{C_W^{-1} d}$ 代入，上式變爲

$$\mathbf{x}^T\mathbf{k} = \frac{\bar{\mathbf{x}}_{(1)}^T\mathbf{k} + \bar{\mathbf{x}}_{(2)}^T\mathbf{k}}{2} = \frac{t_{(1)} + t_{(2)}}{2} \tag{15}$$

這就是分界線方程式。等式右邊提供了鑑別函數 $\mathbf{x^T k}$ 的判斷準則。

範例 2 以 Book_1.txt 的資料爲例，利用式 (11), (15) 分別計算 Fisher 所提出的最佳組合係數 $\mathbf{k^o}$ 及 Mahalanobis 所定義的等距方程式，並畫出如圖 4 所示的圖形。請注意 Book_1.txt 的資料都是整數，繪圖前每筆資料都加上些許的變動值，使得資料的分佈看起來更接近眞實。另外，資料 Book_1.txt 之群組大小懸殊，本範例先探討群組對等的情況，因此從大群組中隨機取出與小群組等量的樣本進行計算。至於針對大小樣本不等之兩群組的問題，留至習題請讀者解決。[1]

圖 4 中的虛線代表 Mahalanobis 所定義的等距方程式，垂直於 Fisher 提出的 k vector。[2]程式碼如下:

[1]讀者將會發現群組規模大小相差太大時所產生的困擾。

[2]從圖 4 上來看，Mahalanobis 所定義的等距方程式並不垂直於 Fisher 提出的 k 向量，這是由於座標軸不等距所致。若將座標軸設爲 axis equal 便能看到垂直的了條線，只不過中間的符號會因座標軸配置的關係，將擠成一團，不適合呈現群組的關係。

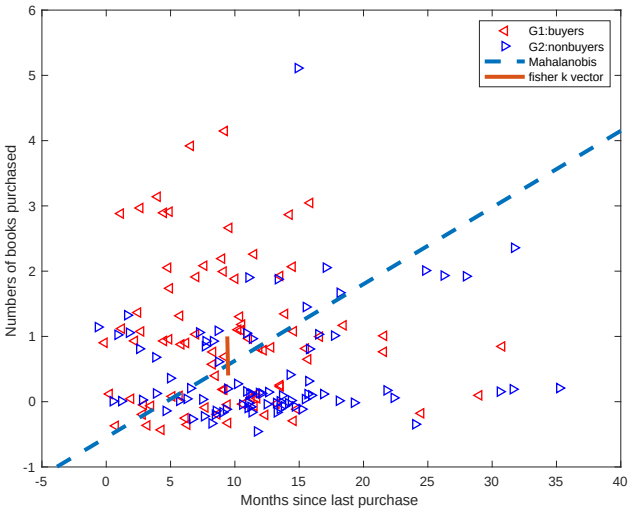

圖 4: Fisher 提出的最佳觀察角度與及 Mahalanobis 等距分界線

```
D=load('BOOKS_1.txt'); % 第 2, 3 欄爲自變數資料
X1=D(D(:,4)==0, 2:3); % Group 1
X2_all=D(D(:,4)==1 ,2:3); % Group 2
n1=size(X1,1); n2=size(X2_all,1); % Group size
b2=unidrnd(n2,n1,1); % 從大群組隨機挑出與小群組等量的資料
X2=X2_all(b2, :);
% 微調後的散佈圖
plot(X1(:,1)+normrnd(0,1,n1,1), X1(:,2)+normrnd(0,1,n1,1)*0.2, '<r')
hold on
plot(X2(:,1)+normrnd(0,1,n1,1), X2(:,2)+normrnd(0,1,n1,1)*0.2, '>b')
mu1=mean(x1); mu2=mean(x2); d=mu2-mu1;
Cw=(cov(X1)*(n1-1)+cov(X2)*(n2-1))/(n1+n2-2);
k=Cw\d'; % Fisher's method
tc=(mu1+mu2)*k/2;
f=@(x1,x2) k(1)*x1+k(2)*x2 - tc;  % Mahalanobis 的等距線
fimplicit(f, 'LineWidth', 3, 'LineStyle', '-')
```

2　配適性與預測（Goodness of Fit and Prediction）

Fisher 提出兩群組間的鑑別函數

$$f(\mathbf{x}) = \mathbf{x^T k} \tag{16}$$

其中最佳組合係數 $\mathbf{k^o}$ 與 $C_W^{-1}\mathbf{d}$ 成正比。這個組合根據群組的樣本資料，提供了分辨兩個群組的最佳「視野」。但當兩群組交錯而存在模糊的界線時，這個最佳的「角度」到底有多好？配適已知的樣本的準確率有多高？或說「誤判率」多低？且式 (16) 如何配置樣本資料的群組呢？有一個方式是提出函數的臨界值 (cutoff score) t_c，也就是當個別資料代入函數 (16)，若其值大於 t_c，便認定爲某一群組，否則爲另一群組。

Mahalanobis 提出兩群組的等距點當作臨界值 t_c，即

$$\mathbf{x}^T\mathbf{k} = t_c = \frac{t_{(1)} + t_{(2)}}{2} \tag{17}$$

其中 $t_{(1)} = \bar{\mathbf{x}}_{(1)}^T\mathbf{k}$，$t_{(2)} = \bar{\mathbf{x}}_{(2)}^T\mathbf{k}$ 爲兩群組中心點的鑑別函數值。不過當群組的大小不等時，式 (17) 需要做些修正以矯正因樣本數不均所造成的誤差；譬如:

$$t_c = \frac{n_1\bar{t}_{(1)} + n_2\bar{t}_{(2)}}{n_1 + n_2} \tag{18}$$

其中 n_1, n_2 分別代表群組 1 與群組 2 的樣本數。

當然鑑別函數對資料的配適性高低並不能保證期「預測」能力，也就是對未知資料的鑑別能力。特別當我們考慮其他因素，如判別錯誤的代價（Cost of Misclassification）。誤判的代價或許因群組而異，譬如將群組一誤判爲群組二，其損失是將群組二誤判爲群組一的 10 倍。將誤判列入考慮是比較符合實際情況的作法。或者說，將更多有力判斷的因素加入，以提高預測的能力或降低預測錯誤的損失。

鑑別函數的好壞如何評斷呢？看一看這個條件式機率密度函數

$$P(Group\ k|\mathbf{x}) \qquad\qquad k = 1, 2, \cdots, C$$

這說明當資料 **x** 出現時，它來自群組 k 的機率。機率最高的那個群組，代表資料 **x** 來自該群組的可能性最高。這個函數稱爲最佳的鑑別函數，也可以拿來作爲評斷其他鑑別函數的依據。但問題是這個條件式機率密度函數，一般也稱爲後驗機率（或事後機率，Posterior Probability），通常是不可知的。還好有個貝式定理可以緩和這個限制

$$P(Group\ k|\mathbf{x}) = \frac{P(\mathbf{x}|Group\ k)P(Group\ k)}{P(\mathbf{x})}$$

其中 $P(\mathbf{x}|Group\ k)$ 一般稱爲群組條件式機率密度函數（或概似函數，Class Conditional Density Function），$P(Group\ k)$ 稱爲群組的先驗機率（或稱事前機率，Prior Probability）。這兩個密度函數相對比較容易「取得（透過估計或假設），」因此最佳的鑑別函數可以改寫爲[3]

$$P(\mathbf{x}|Group\ k)P(Group\ k)$$

對於只有兩個群組而言，最佳（貝氏）的鑑別函數可以寫成（作業 6）

$$f(\mathbf{x}) = \frac{P(\mathbf{x}|Group\ 1)}{P(\mathbf{x}|Group\ 2)} = t_c \tag{19}$$

其中 t_c 爲其臨界值

$$t_c = \frac{P(Group\ 2)}{P(Group\ 1)}$$

當進一步假設 (a) 所有群組資料遵循常態分配，(b) 每個群組的共變異矩陣相同，[4]允許我們寫出群組 1 的條件機率密度函數:

$$P(\mathbf{x}|Group\ 1) = \frac{1}{det(C_W)\sqrt{2\pi}} exp(-\frac{1}{2}(\mathbf{x} - \overline{\mathbf{x}}_{(1)})^T C_W^{-1}(\mathbf{x} - \overline{\mathbf{x}}_{(1)})) \tag{20}$$

其中 $\overline{\mathbf{x}}_{(1)}$ 及 C_W 爲群組 1 的平均數與共變異矩陣的估計。觀察上式指數的部分恰是 Mahalanobis 對於「距離」的定義，於是可以改寫爲

$$P(\mathbf{x}|Group\ 1) = \frac{1}{det(C_W)\sqrt{2\pi}} exp(-\frac{1}{2}D_1^2) \tag{21}$$

同理，群組 2 的條件機率密度函數爲

$$P(\mathbf{x}|Group\ 2) = \frac{1}{det(C_W)\sqrt{2\pi}} exp(-\frac{1}{2}D_2^2) \tag{22}$$

根據式 (21)、(22) 並假設群組的先驗機率分別爲 q_1 及 q_2，式 (19) 取對數後的群界分界線可以進一步寫成

$$\ln\frac{q_1}{q_2} - \frac{D_1^2 - D_2^2}{2} = 0 \tag{23}$$

[3]分母 $P(\mathbf{x})$ 爲共同項，可以刪去不計入。

[4]即 Linear Discriminant Analysis(LDA) 的假設

其中（作業 2）

$$\frac{D_1^2 - D_2^2}{2} = \mathbf{x}^T\mathbf{k} - \frac{\bar{t}_{(1)} + \bar{t}_{(2)}}{2} = t - t_c \tag{24}$$

t 與 t_c 分別是 Fisher 的鑑別函數值及 Mahalanobis 提出的臨界值。式 (24) 做群組判別的預測時，表示爲：將觀察資料判斷爲群組 1，當

$$t < t_c + \ln\frac{q_1}{q_2} \tag{25}$$

式 (25) 看得出，當群組大小一致時，或說當 $q_1 = q_2$ 時，這個群組的判斷與 Mahalanobis 提出的等距軌跡相同（式 (17)）。但當群組大小不一時，式 (25) 的臨界值有別於前述的式 (18)。當進一步考慮判斷錯誤的代價時，譬如 $C(1|2)$ 代表將屬於群組 2 的資料誤判爲群組 1 的代價，$C(2|1)$ 剛好相反。判斷式 (25) 可以修正爲

$$t < t_c + \ln\frac{q_1 C(2|1)}{q_2 C(1|2)} \tag{26}$$

範例 3 利用 Fisher 的鑑別函數 (16) 及 Mahalanobis 提出的臨界值 t_c，計算前一個單元使用的的資料 Book_1.txt，計算其配適性，並製作一張所謂的 hits-and-misses table（或稱 confusion matrix）。在這組資料裡，兩個群組的大小相差很多，因此可以朝兩方面去做：(a) 兩群組相同大小，採式 (17) 的 臨界值 (b) 兩群組不同大小，但採式 (18) 的臨界值，並仔細觀察這兩個結果。

這裡所謂的「配適性」可以簡單說是「命中率」。即將一組已知群組的資料以某種群組分界線（譬如式 (25)）做群組判斷，將判別結果與資料已知的群組做比較，計算判別正確與錯誤的個數與比例，製成一張所謂的 hits-and-misses table。以 Book_1.txt 的資料爲例，共有 1000 筆，其中群組 1（購買者）有 83 筆，群組 2（非購買者）有 917 筆。我們可以依群組大小的相同與否來做配適性計算，測試式 (18) 的必要性。因爲群組 1 筆數較少，爲使兩組大小相等，我們自群組 2 隨意抽取 83 筆資料來做測試。其結果如圖 5 所示（以兩群組大小相同爲例）。

```
Hits Misses Hits/misses
51.00 32.00  0.61
61.00 22.00  0.73
```

圖 5: Hits and Misses Table

程式碼如下（僅列出計算 Hits and Misses 部分）:

```
G1_hits=sum(t1<tc);
G1_miss=sum(t1>tc); % 或 n1 - G1_hits
G2_hits=sum(t2>tc);
G2_miss=sum(t2<tc); % 或 n2 - G2_hits
fprintf('%5.2f %5.2f %5.2f\n',HM')
```

3 觀察與延伸

1. 式 (5) 右邊第一項，正比於轉換變數後屬第一群組的變異數。這個變異數愈小愈有利於分辨。這也是爲什麼圖 1 中間的常態圖變異數比較小的原因
2. Fisher 提出最佳的組合係數 $\mathbf{k}^o$，將自變數組合成一個單一變數，如式 (4)。當代入樣本值時，式 (4) 又稱爲鑑別分數（函數）（Discriminant score (function)）。本單元並未提及如何應用這個 score 來做群組區隔的判別。似乎需要定義（或找出）一個 score 來做爲群組判斷（預測）的關鍵值（cut-off value），Mahalanobis 提出的觀點補足了這個關鍵值。

4 習題

1. 推導式 (6) 與 (8)。
2. 推導式 (11)。
3. 推導式 (15)。
4. 如範例 2，但使用全部的樣本資料。
5. 試著畫出圖 1 。這牽涉到下列的技巧:
   - 畫出雙變量常態（Binormal）的密度函數圖的等高線圖（contour plot），當兩變數間具相關性，看起來像是個傾斜的橢圓。
   - 圖形轉向及位移（可以利用指令 pol2cart 及 cart2pol 做角度的轉移）。
   - 估計常態分配從不同角度觀察（變數轉換）的變異數。

   圖形的轉向即座標位置的轉換，可以利用轉置矩陣 T 來幫忙。譬如樣將向量 $a = [2 \;\; 1]^T$ 逆時針轉 $\theta = 30^o$ ，可以這麼做

$$b = Ta, \qquad T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

如圖 6 所示。左圖將向量 **a** 逆時針轉 θ 角度（譬如，30^o）；右圖則將橢圓旋轉 90 度。程式碼如下

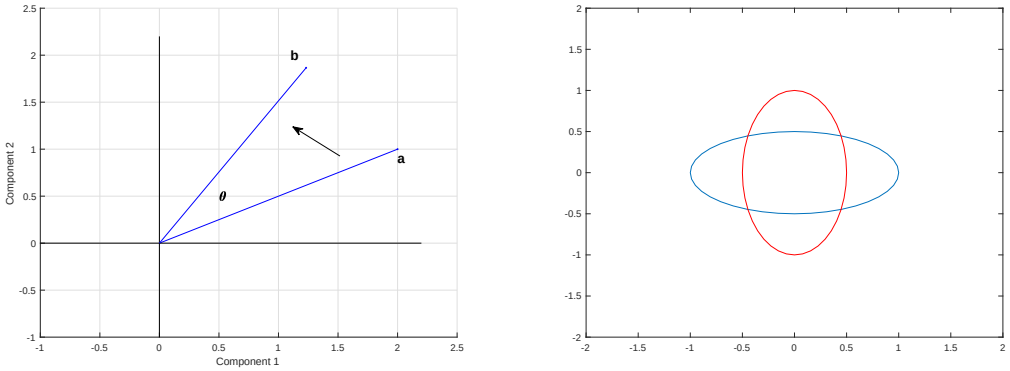

圖 6: 座標（圖形）轉換

左圖

```
a=[2 1];
biplot(a); % 繪製向量 a
t=pi/6;
T=[cos(t) -sin(t);sin(t) cos(t)];
b=T*a'; % 逆時針旋轉向量 a
hold on;
biplot(b'),axis([-1 2.5 -1 2.5]), hold off
```

右圖

```
t=0:pi/20:2*pi;
ellips_x=1*sin(t);
ellips_y=0.5*cos(t);
plot(ellips_x,ellips_y),hold  on
axis([-2 2 -2 2])
theta=pi/2; % 旋轉 90 度
T=[cos(theta)  -sin(theta);sin(theta)  cos(theta)];
R=T*[ellips_x;ellips_y]; % 針對所有的座標點做轉置
plot(R(1,:),R(2,:),'r'),hold off
```

除使用轉置矩陣外，MATLAB 提供了兩個指令 cart2pol, pol2cart 也可以達到

圖形（座標點）轉置的目的。程式碼如下，其中 cart2pol 將所有座標點的表示法從 Cartesian coordinate 改爲 polar coordinate，第二行將原來每個座標點的角度加上欲旋轉的角度，再利用 pol2cart 轉回 Cartesian coordinate 的 X-Y 座標。

```
[theta,rho]=cart2pol(ellips_x,ellips_y);
[x,y] = pol2cart(theta+pi/2,rho); % 旋轉 90 度
plot(x,y,'r')
```

6. 推導兩個群組的最佳鑑別函數 (19)。
7. 推導出式 (23) 及 (24)。
8. 利用練習 3 得到的組合係數與臨界值，對另一組資料 Book_ 2.txt 作「預測」測試。同樣製作一張 hits-and-misses table，比較看看這兩個結果的差別。
9. 同上，但應用式 (25) 的結果。
10. 同上，但應用式 (26) 的結果，其中 $C(2|1) : C(1|2) = 6 : 1$。

References

[1] J. Latin, D. Carroll, P. E. Green, ”Analyzing Multivariate Data,”2003, Duxbruy.

[2] A. C. Rencher, ”Multivariate Statistical Inference and Applications,”1998, John Wily and Sons.

[3] 黃俊英，” 多變量分析（第七版），” 中國經濟企業研究所。