群組分類
Quadratic Approaches

last modified March 22, 2006

不管是 Linear Regression 或 Linear Discriminant Analysis, 所提供都是一條線性的分界線。當不同群組間的資料稍密合, 或群組內的變數存在不同程度的相依性時, 線性分隔的效果便大打折扣。如果這條分割線能彎一點, 或許能切割的更漂亮。本單元將之前的線性方式稍做修正即可達到這樣的效果。

1 背景介紹:Quadratic Discriminant Function

Regression Model

假設輸入變數爲 x_1, x_2, 其所有可能的值涵蓋二度空間。此時如果將兩個變數擴展爲五個變數 $x_1, x_2, x_{12}, x_{22}, x_1x_2$, 同樣利用迴歸模式與最小平方法建立一條分界線, 當將此分界線投映回原來的空間時, 它將呈現出一條無與倫比的漂亮曲線。這五個變數因其彼此相關的本質, 並非將空間拓展爲五度空間, 實際仍在二度空間裡, 這個所謂的 Augmented Regression Model 寫成

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 \tag{1}$$

其參數的估計與群組判別的方式與線性模式相同, 透過範例1的練習將爲更清楚, 不再贅述。

Discriminant Analysis

另外在 Linear Discriminant Analysis(LDA) 的假設裡, 所有的共變異矩陣都相同,

這無形中也侷限了分辨的能力。當群組間的分配情況不同時, 這個假設便不妥了。如果允許各群組有不同的共變異矩陣, 上個單元的 Linear Discriminant Function

$$\delta_k(\mathbf{x}) = \mathbf{x}^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log Pr(G=k) \tag{2}$$

將變成 Quadratic Discriminant Function:

$$\delta_k(\mathbf{x}) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x}-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}-\mu_k) + \log Pr(G=k) \tag{3}$$

上式是一個含變數平方項的函數, 這個方法稱爲 Quadratic Discriminant Analysis (QDA), 而介於群組 k 與群組 l 間的分界線便可以下列的集合表示:

$$\{\mathbf{x}|\delta_k(\mathbf{x}) = \delta_l(\mathbf{x})\} \tag{4}$$

QDA的群組判別方式同 LDA, 寫成

$$G(\mathbf{x}) = \arg\max_k \delta_k(\mathbf{x})$$

2 練習

以下練習的資料取自la_2.txt

範例1: 利用式(1) 的非線性複迴歸模型, 畫出群組間的分界線。

將迴歸模式從兩個變數擴展爲五個變數的模型如 (1), 在程式的編寫上只需在資料矩陣 X 上做少許的修正即可, 亦即再加入三欄分別來自 x_1x_2, x_{12}, x_{22} 的資料, 即

$$X = \begin{bmatrix} 1 & x_1(1) & x_2(1) & x_1(1)x_2(1) & x_1^2(1) & x_2^2(1) \\ 1 & x_1(2) & x_2(2) & x_1(2)x_2(2) & x_1^2(2) & x_2^2(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1(N) & x_2(N) & x_1(N)x_2(N) & x_1^2(N) & x_2^2(N) \end{bmatrix}$$

接著計算參數 $\underline{\beta}$ 的最小平方估計:$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$, 如同線性迴歸模式, 式(1) 的 Augmented Regression Model 的分界線集合表示爲

$$\left\{(X_1, X_2) | \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + \hat{\beta}_4 X_1^2 + \hat{\beta}_5 X_2^2 = 0.5\right\}$$

如圖1那條彎彎的線。這是一條利用等高線圖畫出來的線, 其函數依據分界線集合爲

$$Z = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + \hat{\beta}_4 X_1^2 + \hat{\beta}_5 X_2^2$$

繪製 $Z = 0.5$ 的等高線即爲分界線。指令如

```
contour(X1,X2,Z,[0.5 0.5])
```

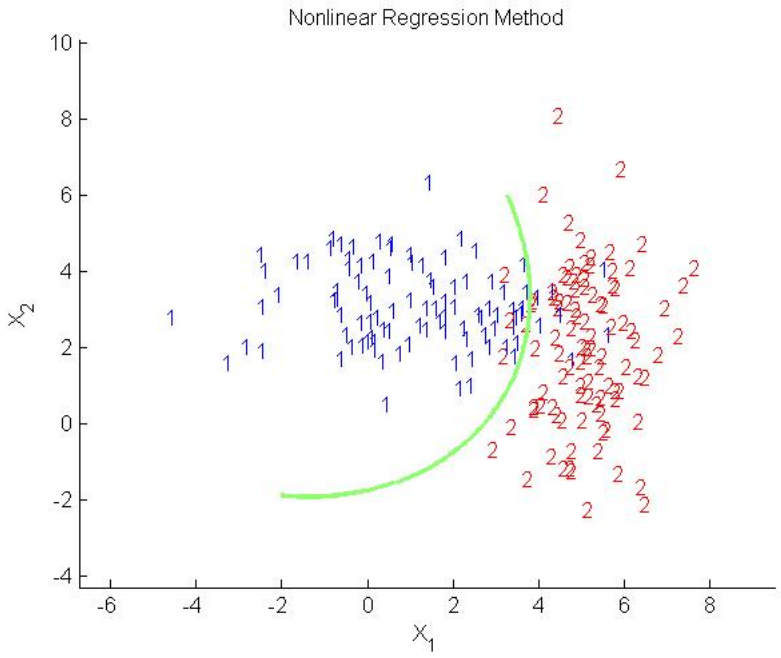

圖 1: 非線性複迴歸群組分界線

範例2: 利用式(3) 的 Quadratic Discriminant Function 與式 (4) 的 QDA 群組

分界線集合, 畫出群組間的分界線。

式 (3) 牽涉到三個統計量 $\mu_k, \Sigma_k, Pr(G = k)$ 的計算, 分別以各組的樣本平均數、樣本共變異矩陣及群組比例取代。作圖前最好將式 (3) 作一番整理並令 $\delta_1(\mathbf{x}) = \delta_2(\mathbf{x})$, 將 $\mathbf{x}$ 的平方項、一次項與常數項分開, 寫成($\mathbf{x} = [X_1\ X_2]$)

$$c = c_1X_1 + c_2X_2 + c_3X_1X_2 + c_4X_1^2 + c_5X_2^2$$

的形式, 畫等高線 $Z = c$, 其中 Z 爲上式等號右邊的函數。結果如圖2所示。

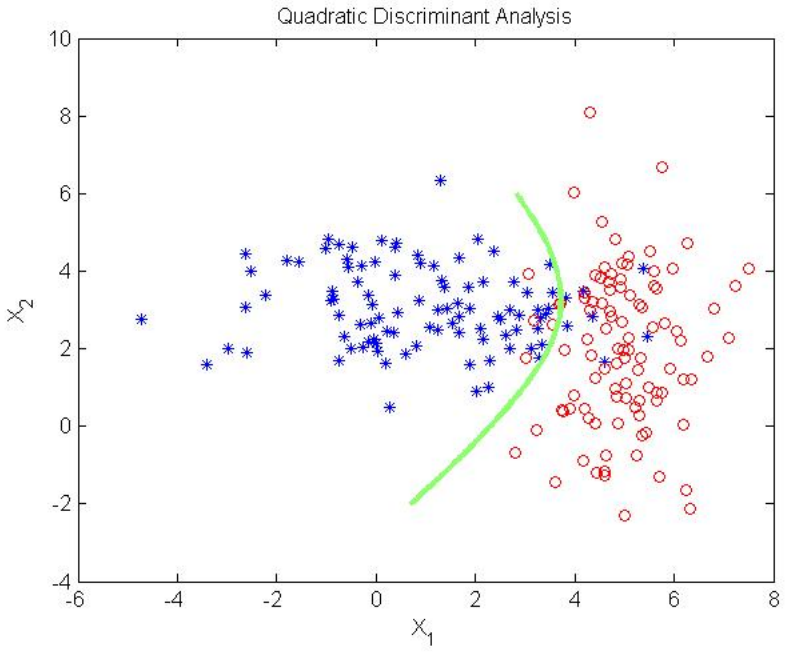

圖 2: QDA 群組分界線

3 觀察

1. 不管是 LDA 還是 QDA 其基本假設都是群組資料的常態分配, 兩者的不同僅是對於共變異矩陣的的假設。當資料呈現常態分配時, 其周邊的範圍不難想像是呈現平滑曲線的, 所以 QDA 往往能展現很好的切割效果, 成功的將群組本身

的變異情況涵蓋進來。試試看自己產生不同的常態群組資料, 改變共變異矩陣, 看看 QDA 與 LDA 的差別。

2. 圖1與2的兩張曲線圖來自 (1)(當 y=0.5) 與 (3)(當$\delta_1(\mathbf{x}) = \delta_2(\mathbf{x})$), 看起來有點像, 又有點不像。如果要從外觀上去判斷的話, 座標軸的範圍必須一致, 若要進一步確定可以從程式裡面去觀察相對應的係數是否相同?

3. 每個方法的優劣並非只是看分界線切的漂不漂亮, 那只能代表對已知資料或一般稱爲 training data 的認知, 並不代表對未知資料 (一般稱爲 testing data) 判別的能力。因此在評斷一個方法時必須同時對已知與未知做出量化的數據。就群組的分類而言可以計算「Misclassification Rate,」即分類錯誤的比例。

4 作業

1. 推導出式 (3) 的結果。

2. QDA 中兩群組的分界線方程式由 (4) 裡面的$\delta_k(\mathbf{x}) = \delta_l(\mathbf{x})$構成, 請將這個式子化解開到可以進行程式寫作的階段, 即寫出範例2中的 $c, c_1, \cdots, c_5$。

3. 自行產生兩個不同群組的常態資料, 選擇不同的 Mean 與 Covariance 組合, 試試本單元所提出的方法, 看看會畫出什麼樣的曲線。

4. 將資料分爲兩部分:training data 與 testing data。利用 training data 計算出分界線及其 Misclassification Rate, 再利用 testing data 計算 Misclassification Rate, 藉此比較這兩種 quadratic approaches。

參考文獻

[1] T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning:Data Mining, Inference, and Prediction," Springer.