

Practice#1：線性預測：關於時間序列

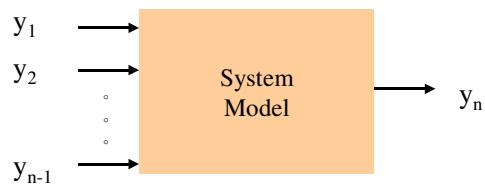
目的：

迴歸分析經常被用來作為預測的工具。它主要的作用是找出自變數與因變數間的關係，譬如

$$y=f(x_1,x_2,\dots,x_n)$$

當 $f(.)$ 被確定之後，代入一組自變數 (x_1,x_2,\dots,x_n) 的值，便可以輕易算出 y 值，作為預測值。這個模式牽涉到利用自變數與因變數的資料，來確立兩者之間的關係。迴歸分析模式所探討的資料，通常沒有時間上先後順序的關係。資料間的排序並不影響參數的估計。但另一種是時間序列式的資料模式，沒有自變數與因變數之別，只關心一個變數。資料依時間的順序產生，目的是希望藉由已產生的資料，預測下一個時間點上最有可能出現的值。

一般的假設，任何資料的產生都與過去的資料值相關。譬如說，假設今天的加權指數與之前幾天的指數相關。如果這個假設成立，對於未來資料的預測，便可以根據過去的歷史資料推算出來。資料序列間的關係如圖所示



以數學的方式來表示，可以寫成

$$y_n=f(y_1,y_2,\dots,y_{n-1})$$

其中 $f(.)$ 代表第 n 個 y 值與前面 $n-1$ 個 y 值的關係。

如果這個假設可行，這比起迴歸分析要方便多了，只要一組資料就可以做預測了。以加權指數來說，可以不去理會加權指數與其他的經濟因素或政治因素，或類似許多專家提出的充滿許多變數的數學模型。

這種僅以單一變數的時間資料做預測的模式中，以線性關係最為簡單，如

$$y(n)=a_1y(n-1) + a_2y(n-2) + \dots + a_p y(n-p)$$

其中 $y(n-1), y(n-2), \dots, y(n-p)$ 代表最近的 p 個資料值。而其相關係數 a_1, a_2, \dots, a_p 代表其的相關程度。這個模式在某些領域被稱為『線性預測』(Linear Prediction)。

至於這個模式合不合理？在實際的應用上效果好不好？並不是本練習的主要目的，我們反倒要關心面對不同的資料，這個模式的 p 如何決定？相關係數 a_1, a_2, \dots, a_p 如何估計？

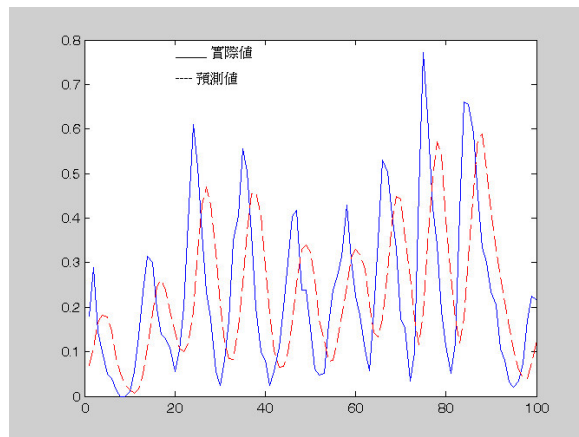
練習：

1. 上述的模式來自一個很直覺的想法：“未來的資料應該與歷史資料有某種程度上的相關”。有一個簡單的模式，如下

$$y(n) = \frac{y(n-1) + y(n-2) + \dots + y(n-p)}{p}$$

換句話說，未來值預測為過去歷史資料值的平均。試用太陽黑子 sunspot.mat 這組資料來做練習，看看預測的結果效果如何？這牽涉到使用適當的 p 值，及你如何展現預測的『效果』？

- ◇ 預測的好或不好如何判斷呢？先想一想。
- ◇ 將預測結果拿來與實際值比較是方法之一。問題是「怎麼比？」「拿多少個來比？」
- ◇ 當然也可以試圖以目視的方式來比較，譬如將實際值與預測值畫在一起，看看差多少？如下圖所示。
- ◇ 前兩個方法可以一起做，不過都牽涉到要先計算出若干個預測值。這時在 Matlab 程式的安排上要能一次解決，不需要用迴圈。
- ◇ 將「比較」的方式訂出來後，才能決定不同 p 值的優劣。



2. 同上，另外一個直覺的想法是，愈近期的資料是否給於較高的權重，來彰顯它的重要性。譬如，

$$y(n) = \frac{3}{6}y(n-1) + \frac{2}{6}y(n-2) + \frac{1}{6}y(n-3) \text{ 或}$$

$$y(n) = \alpha y(n-1) + \alpha(1-\alpha)y(n-2) + \dots + \alpha(1-\alpha)^p y(n-p)$$

其中 $0 < \alpha < 1$

選擇適當的 p 與 α 值，並與上題做一比較。

3. 第三個做法，直接採用最小平方法的方式來估計最佳的係數： a_1, a_2, \dots, a_p 。這個做法與迴歸分析的做法相同。這部分是上課分析的重點；如何將問題以矩陣的方式表示？如何運用最小平方法（其實與迴歸分析單元的方式相同）？如何準備寫程式時需要的「資料矩陣」？這些都要一一的寫出來，植入程式中。

觀察：

1. 上面的練習牽涉到對於預測結果的評估及不同方法間的比較。老師可以在課堂上與學生互相討論出『評估』與『比較』的方式。由老師先提出一些一般研究上可能的做法，再由學生自己去深入思考出比較恰當的方式。
2. 對於 p 的選擇，開始可以是隨意的，譬如從1到6分別去試試看。觀察每一個 p 值所產生的預測結果，對不同的資料是否有不同的 p 值？
3. 是否對不同的資料都有一個最佳的 p 值？如果決定的 p 值過大或過小，是否影響預測的品質呢？

4. 觀察資料矩陣 Y 的 RANK. 其中 $Y = \begin{bmatrix} y(p) & y(p-1) & \dots & y(1) \\ y(p+1) & y(p) & \dots & y(2) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \dots & y(N-p) \end{bmatrix}$

5. 觀察矩陣 $Y^T Y$ 的 eigenvalues.
6. 上面觀察的 RANK 值或 eigenvalues 是否與可以幫忙決定 p 值？

作業

採用最小平方法的方式做線性預測，資料為 lp_data1.txt, lp_data2.txt, lp_data3.txt。紀錄自己的觀察或就上述觀察的項目中，寫下值得記載的東西，特別是對 p 的選擇，及預測的誤差。

