以 least square method 搜尋非線性方程式的係數值

作者: 何岳峰
website: http://www.hoamon.info/
email: hoamon@hoamon.info

version: 4
export time: 2011-03-08 10:26:19 (UTC+8)

應用問題說明

當我們有一堆觀測數據，想要找出一條方程式可以將自變數、應變數的關係明確表示時，可以使用 least square method 求解，例如我們有 10 個觀察值 (1, 4), (4, 5), (1, 2), (4, 5), (6, 9), (1, 8), (4, 5), (3, 4), (5, 5), (6, 6) 等，而我們先猜測它們的關係是二次的，應該要符合 $Y = a \times X^2 + b \times X + c$ 方程式，有了觀測值及預估方程式形式後，接下來就可以利用 least square method 計算出 a, b, c 三個係數。

評估非線性方程式的係數值

假定一含有 n 個未知係數的方程式 $E(x_1, x_2, ..., x_k, p_1, p_2, ..., p_n)$,在 m 組觀測值 $((y_1, x_{11}, ..., x_{1k}),$ $(y_2, x_{21}, ..., x_{2k}), ..., (y_m, x_{m1}, ...x_{mk}))$ 下，欲求出此 n 個係數的值，期使 $E(x_1, x_2, ..., x_k, p_1, p_2, ..., p_n)$ 與觀測值的誤差最小。

首先定義理論值與觀測值的誤差如式 1 的 f_i 。

$$f_i(p_1, p_2, ..., p_n) = y_i - E(x_{i1}, x_{i2}, ..., x_{ik}, p_1, p_2, ..., p_n), i \in 1, ..., m \tag{1}$$

$$F(p_1, p_2, ..., p_n) = \sum_{i=0}^{m} (f_i(p_1, p_2, ..., p_n))^2, m \geq n \tag{2}$$

將每一觀測值誤差平方並加總起來，可得到一目標函式 F 其定義如式 2 。當 F 函式的值為全域最小值或是區域最小值，其特定 $[p_1^*, p_2^*, ..., p_n^*]$ 組合，也就是 P^* ，即為該 E 函式的係數全域或區域最佳解。

所有的非線性最佳化都是用迭代方法計算的：從一個起點 P_0 開始搜尋，產生一系列的向量 $P_1, P_2, ...$ ，希望可以收斂至一個 P^* ，使得函式成為全域或區域最佳解 [2] 。

$$F(P_{k+1}) = F(P_k) + h^\top F'(P_k) + \frac{1}{2} h^\top F''(P_k) h + O(||h||^3) \tag{3}$$

$$h = P_{k+1} - P_k \tag{4}$$

如式 3 ，首先以 Taylor expansion 改寫 F 函式，式 3 中的 $O(||h||^3)$ 代表後面無窮盡的項目， k 代表的是第幾次 iteration 。在每次 iteration 中，我們要移動 h 的距離使得 $F(P_k+h) < F(P_k)$ 。目前常用來計算 h 的迭代法有 The Steepest Descent method, Newton's method, Guass-Newton method, The Levenberg-Marquardt method。

The Steepest Descent method

Steepest Descent method 將式 3 改寫為式 5 ，忽略二階以後項目。

$$F(P_k+\alpha h) \simeq F(P_k) + \alpha h^\top F'(P_k), \alpha > 0 \tag{5}$$

而原來的 h 改寫成一個純量 α 再乘以向量，這樣在 P_{k+1} 接近 P^* 時， α 必定近似 0 ，所以我們又可以得到式 6 ，

$$\lim_{\alpha\to 0} \frac{F(P_k) - F(P_k+\alpha h)}{\alpha \|h\|} = -\frac{1}{\alpha \|h\|} h^\top F'(P_k) = -\frac{h^\top}{\alpha \|h\|} \frac{F'(P_k)}{\|F'(P_k)\|} \|F'(P_k)\| = -\frac{cos\theta}{\alpha} \|F'(P_k)\| \tag{6}$$

$\alpha \|F'(P_k)\|$ 為純量，如果我們希望 $F(P_k) - F(P_k+\alpha h)$ 的值愈大愈好，這樣 $cos\theta$ 就必須為 -1 ，而 $cos\theta$ 是 h 與 $F'(P_k)$ 的夾角，所以如式 7 ，最佳的 h_{sd} 必為 $-F'(P_k)$ 。

$$h_{sd} = -F'(P_k) \tag{7}$$

如式 7 ，陡降法中的 h_{sd} 是以斜率值負值為移動方向。而 α 的值，我們需用 line search 求得，但效率通常會是個問題，所以也可以使用 binary search 方式來求得，其概念是先取得一個 α_{min} 值讓 $F(P_k) - F(P_k+\alpha h)$ 大於 0 ，再取得一個 α_{max} 值讓 $F(P_k) - F(P_k+\alpha h)$ 小於 0 ，接下來以 $\frac{1}{2}(\alpha_{min}+\alpha_{max})$ 為新的 α_{middle} 值，去計算 $F(P_k) - F(P_k+\alpha h)$ 是大於 0 還是小於 0 。若大於 0 ，則新的 α 值應為 $\frac{1}{2}(\alpha_{middle}+\alpha_{max})$ ，若小於 0 ，則新的 α 應為 $\frac{1}{2}(\alpha_{middle}+\alpha_{min})$ ，如此迭代計算後，當滿足預設條件或達迭代次數即可決定 α 。

Newton's method

牛頓法則考慮以 F 函式的二階 Hessian 矩陣來計算 h 。它將式 5 再取其一次微分得到式 8 ，若 $(P_k+\alpha h) = P^*$ ，則 $F'(P_k+\alpha h)$ 必等於 0 。

$$F'(P_k+\alpha h) \simeq F'(P_k) + \alpha h^\top F''(P_k), \alpha > 0 \tag{8}$$

所以我們可以得到式 9 ，而 h_n 就等於式 10 定義。

$$0 = F'(P_k) + \alpha h^\top F''(P_k) \tag{9}$$

$$h_n = -\frac{F'(P)}{\alpha F''(P)} \tag{10}$$

在搜尋效率上，牛頓法為二元收斂，而陡降法為線性收斂，所以牛頓法在接近最佳解時比較快，而陡降法則是離最佳解較遠時比較快，且因 Hessian matrix 在計算上不一定為 positive definite ，所以牛頓法往往會混合陡降法來實作。

The Guass-Newton method

Least squares problems 一般能用陡降法或是牛頓法求解，但要追求效率的話，我們應該作部份調整，像是盡量使用二元收斂或是不需實作出 F 函式 [2] 。

我們將式 3 的 Taylor expansion 套用在 f 上，如式 11 。

$$f(x+h) = f(x) + f'(x)h + O(||h||^2) \tag{11}$$

式 11 可再整理為式 12 。

$$f(x+h) \simeq l(h) \equiv f(x) + f'(x)h \tag{12}$$

$$F(x+h) \simeq L(h) \equiv \frac{1}{2} l(h)^\top l(h) = \frac{1}{2} f^\top f + h^\top f'^\top f + \frac{1}{2} h^\top f'^\top f' h \tag{13}$$

再將式 13 對 h 作一次微分得到式 14 。

$$L'(h) = f'^\top f + f'^\top f' h \tag{14}$$

因為在最佳解時， $L'(h)$ 等於 0 ，所以我們可以得到 h_{gn} 如式 15 。

$$h_{gn} = -\frac{f'^\top f}{f'^\top f'} \tag{15}$$

The Levenberg—Marquardt method

The Newton's method may not be defined beacause of the singularity of $F''(P_k)$ at a given point P_k , or the search direction h_n may not be a descent direction [1] . The algorithm used the gradient methods their ability to converge from an initial guess which may be outside the region of convergence of other methods. The algorithm uses the Taylor series method the ability to close in on the converged values rapidly after the vicinity of the converged values has been reached. Thus, The method combines the best features of its predecessors while avoiding their most serious limitations. [3] .

Levenberg-Marquardt 方法乃加入一 damping parameter μ 。

$$h_{lm} = -\frac{f'^\top f(P_k)}{f'^\top f' + \mu I}, I = Identity\ Matrix \tag{16}$$

此自定參數的效益在於，For all $\mu > 0$ the coefficient matrix is positive definite, and this ensures that h_{lm} is a descent direction. For large values of μ we get

$$h_{lm} \simeq -\frac{1}{\mu} F'(P_k) \tag{17}$$

If μ is very small, then $h_{lm} \simeq h_{gn}$, which is a good step in the final stages of the iteration, when P_k is close to P^* . If $F(P^*) = 0$ (or very small), then we can get (almost) quadratic final convergence [2] .

μ 值在每個 iteration 中，皆不同，而 μ_k 的選擇，主要有兩種方法，一是看 $f'(P_k)^\top f'(P_k)$ 中，對角線元素中最大值再乘以 γ ，一般來說， γ 的值介於 $10^{-6} \sim 1$ 之間。或是也可以用 $F(P_k) - F(P_{k-1})$ 的值 s_k 來判斷，當 $s_k \geq 0$ 時， μ_k 增加 10 倍，當 $s_k \leq 0$ 時， μ_k 減少 10 倍， μ_0 的初始值則是設為 0.001 。

小結

理論上，非線性最佳化是很難求得全域最佳解的，就算你運氣真的很好，瞎貓怕上死耗子，讓你遇到全域最佳解，但你還是沒有辦法去驗證它的確是全域最佳解。只有在某些特殊題型下，你才知道所求出的解是全域最佳解，最簡單的例子就一元二次方程式，當它是凹向上時，有全域最小值，當它是凸向上時，有全域最大值。

所以，以上介紹的 4 種找尋 h 的方法，都只是一種有系統的找尋技巧，這些方法經過數學推導，可算是比較有效率罷了，不代表只有這些方法可以用在 least square problem 上。你也可以任意混合這 4 種方法，如在第 1 ~ 10 次迭代時，使用陡降法，在第 11 ~ 20 次時，使用 Levenberg-Marquardt 方法，在第 20 次以後一律使用高斯-牛頓法，這種方法是沒有對錯的，只有解題時間的快與慢而已，而這快與慢就又牽扯到起始值的挑選及非線性函式的特性。這次這樣用比較快，不表示這對其他問題的解題速度就會比較快。

而在實務上，我們也往往不需要真正計算出 F' 、 F'' 、 f' 的數學式，甚至 f 、 F 有可能是不可微分的。所以要利用數值分析的概念，直接取函式的數值來作計算，如式 18 ：

$$f'(p) = \frac{f(p+\Delta) - f(p)}{\Delta} \tag{18}$$

要知道 f 函式在 p 點的斜率為何? 不需要求取 f 的一階微分，只要將 $f(p)$ 的數值計算出來，並取一個非常小的 Δ 位移值代入計算出 $f(p+\Delta)$ 數值，並除以 Δ 即可。

也因為 Levenberg-Marquardt Method 只須取出 f 函式的一階微分，所以大部份的數值分析方法都是採用 LM 方法來實作的。

References

[1] M. S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear programming*. John Wiley and Sons, 2006.

[2] K. Madsen, H. B. Nielsen, O. Tingleff, and Mathematical Modelling. IMM METHODS FOR NON-LINEAR LEAST SQUARES PROBLEMS, 2004.

[3] Donald W. Marquardt. An algorithm for Least-Squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, June 1963. ArticleType: primary_article / Full publication date: Jun., 1963 / Copyright 1963 Society for Industrial and Applied Mathematics.