PHÉP NỘI SUY VÀ HỒI QUY

PHƯƠNG PHÁP SỐ

Hoàng Văn Đông

Bộ môn Khoa học máy tính, Khoa Công nghệ thông tin, Đại học Thuỷ Lợi

30/6/2019

Mục lục

- Bài toán nội suy
- Nội suy đa thức
- Bài toán hồi quy
- Phương pháp bình phương cực tiểu

Bài toán nội suy

Thông thường trong một số lĩnh vực như kinh tế chẳng hạn, các đại lượng khảo sát thường không được cho dưới dạng hàm liên tục, mà là bảng các giá trị rời rạc.
Ví dụ, ta biết đại lượng y được biểu diễn bởi f(x) nhưng không biết công thức cụ thể mà chỉ biết bộ giá trị $(y_0, y_1, ..., y_n)$ ứng với các điểm $(x_0, x_1, ..., x_n)$.
Bài toán nội suy là bài toán yêu cầu chúng ta tìm một hàm g(x) dựa vào các điểm rời rạc $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$ biết trước để xấp xỉ hàm f(x).
Hàm g(x) chúng ta cần tim phải thoả mãn các điều kiện sau

- $g(x_i)$ i =0,1,2,...,n gần các điểm y_i nhất theo một nghĩa nào đó
- g(x) là duy nhất theo một số điều kiện nào đó
- Hàm g(x) liên tục và ít thay đổi trong từng đoạn $[x_i, x_{i+1}]$

Bài toán nội suy

Định lý Weierstrass 1 về xấp xỉ hàm

Cho f(x) là một hàm thực liên tục xác định trên khoảng [a, b]. Khi đó với mọi $\varepsilon > 0$ tồn tại một đa thức p(x) bậc m với các hệ số thực sao cho với mọi giá trị $x \in [a, b]$ ta có $|f(x) - p(x)| < \varepsilon$

Định lý Weierstrass 2 về xấp xỉ hàm

Cho f(x) là một hàm thực liên tục xác định trên khoảng $[-\pi, \pi]$ và $f(-\pi) = f(\pi)$. Khi đó với mọi $\varepsilon > 0$ tồn tại một đa thức lượng giác

$$q_m(x) = \frac{a_0}{2} + \sum_{j=1}^{m} [a_j cos(jx) + b_j sin(jx)]$$

với các hệ số thực sao cho với mọi giá trị $x \in [-\pi, \pi]$ ta có $|f(x) - q(x)| < \varepsilon$.

Nội suy đa thức

Nếu ta biết rằng các cặp giá trị $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$ là thể hiện của một hàm f(x) nào đó. ta có bài toán nội suy đa thức:

Bài toán nội suy đa thức

Cho một mẫu quan sát gồm n+1 cặp các giá trị đã biết của x và y : $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$. Hãy xây dựng một đa thức bậc m $\leq$ n

$$p_m(x) = a_0 + a_1 x^1 + ... a_{m-1} x^{m-1} + a_m x^m \tag{1}$$

sao cho $p_m(x_i) = y_i, i = 0, 1, ..., n$
Người ta gọi bài toán trên đây là bài toán nội suy đa thức, và đa thức $p_m(x)$ được gọi là đa thức nội suy.

Định lý

Có duy nhất một đa thức với bậc không quá n đi qua n + 1 điểm cho trước $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$

Nội suy đa thức

Phương pháp Horner

Trong bài toán nội suy đa thức, ta thường xuyên phải tính giá trị của đa thức $p_m(x) = a_0 + a_1x^1 + ...a_{m-1}x^{m-1} + a_mx^m$ tại điểm x. Ta thường thực hiện phép tính trên bằng luật Horner như sau:

B0: Đặt $P = a_m$

B1: Tính $P = Px + a_{m-1}$

...

Bm: Tính $P = Px + a_0$

Giá trị P cuối cùng là giá trị của $p_m(x)$.

Nội suy đa thức

Định lý Rolle

Cho f(x) là hàm số thực liên tục trên khoảng đóng [a,b] và khả vi trên khoảng mở (a,b) và f(a) = f(b). Khi đó tồn tại điểm $\xi \in (a, b)$ sao cho $f'(\xi) = 0$.

Định lý

Giả sử hàm f(x) có đạo hàm liên tục đến cấp n+1 trên đoạn [a,b] và $p_m(x)$ là đa thức nội suy, tức là: $p_m(x_i) = f(x_i) = y_i$, i = 0, 1,..., n. Với các mốc nội suy là $a = x_0 < x_1 < ... < x_n = b$.
Đặt $\omega_{n+1}(x) = \prod_{i=0}^{n}(x - x_i)$ và $R(x) = f(x) - p_m(x)$.
Khi đó, với mọi x thuộc [a, b], tồn tại $\eta \in [a, b]$ (phụ thuộc vào x) sao cho

$$R(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!}\omega_{n+1}(x)$$

Nội suy đa thức

Hệ quả

Gọi $M = \sup_{a \le x \le b} |f^{(n+1)}(x)|$, khi đó ta có:

$$|R(x)| \le |f(x) - p_m(x)| \le \frac{M}{(n+1)!} |\omega_{n+1}(x)| \tag{2}$$

Đây là công thức đánh giá sai số của đa thức nội suy

Nội suy Lagrange

Giả sử ta có các điểm quan sát $x_0, x_1, ...x_n$ với khoảng chia đều hoặc không đều và một dãy các giá trị quan sát $y_0, y_1, ...y_n$.
Ý tưởng đơn giản đầu tiên là tìm một đa thức nội suy có bậc không quá n là tổng của các đa thức như sau:

$$p_n(x) = H_0(x) + H_1(x) + ... + H_n(x)$$

Ở đây các hàm $H_i(x)$ đều có bậc không quá n và $H_i(x_i) = y_i, H_i(x_j) = 0$ khi $j \neq i$.

Nội suy Lagrange

Để $H_i(x_j) = 0$ khi $j \neq i$ thì $H_i(x)$ phải có dạng:

$$H_i(x) = K(x)(x - x_0)(x - x_1)...(x - x_{i-1})(x - x_{i+1})...(x - x_n)$$

Vì $H_i(x)$ có bậc không quá n nên K(x) là hằng số. Thay x = x_i vào công thức $H_i(x)$ ta được:

$$H_i(x_i) = K(x)(x_i - x_0)(x_i - x_1)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n) = y_i$$

$$\Leftrightarrow K(x) = \frac{y_i}{(x_i - x_0)(x_i - x_1)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)}$$

$$\Leftrightarrow H_i(x) = y_i \frac{(x - x_0)(x - x_1)...(x - x_{i-1})(x - x_{i+1})...(x - x_n)}{(x_i - x_0)(x_i - x_1)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)}$$

Nội suy Lagrange

Ký hiệu $L_i(x) = \dfrac{(x-x_0)(x-x_1)...(x-x_{i-1})(x-x_{i+1})...(x-x_n)}{(x_i-x_0)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_n)}$

Như vậy $L_i(x)$ có tính chất:

$$L_i(x_j) = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

Và đa thức $p_n(x)$ có dạng

$$p_n(x) = y_0L_0(x) + y_1L_1(x) + ... + y_nL_n(x) \tag{4}$$

Nội suy Lagrange

Ví dụ 1: Cho các nút giá trị sau:

I	x_i	y_i
0	0	0.000
1	1.5	0.682
2	2	0.841

Hãy xác định đa thức nội suy Lagrange đi qua các điểm trên? Hãy tính giá trị gần đúng của hàm số tại điểm x=1? Hãy đánh giá sai số lý thuyết tại x=1.

Sai phân

Cho f(x) là hàm của x và $h = \Delta x$ là một hằng số không đổi biểu thị cho khoảng thay đổi trên biến x và được gọi là số gia của x. Khi đó số gia tương ứng trên f(x):

$$\Delta f(x) = f(x + \Delta x) - f(x)$$

được gọi là sai phân tiến cấp 1 tại x của f(x) ứng với h.
Ngoài ra số gia tính bởi:

$$\nabla f(x) = f(x) - f(x - \Delta x)$$

được gọi là sai phân lùi cấp 1 của x tại f(x) ứng với h.
Sai phân của sai phân cấp 1 của f(x) là sai phân cấp 2 của f(x), tương tự, theo quy nạp ta có thể xây dựng các sai phân cấp k bất kì của f(x).

Sai phân

Tổng quát, ta có công thức sai phân tiến cấp k:

$$\Delta^k y_i = \sum_{j=0}^{k} (-1)^j C_k^j y_{i+k-j}$$

Và công thức sai phân lùi cấp k:

$$\nabla^k y_i = \sum_{j=0}^{k} (-1)^j C_k^j y_{i-j}$$

Sai phân]

Bảng sai phân tiến

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\Delta^4 y_2$
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\Delta^3 y_3$	$\Delta^4 y_3$
x_4	y_4	Δy_4	$\Delta^2 y_4$	$\Delta^3 y_4$	$\Delta^4 y_4$
.	.	.	.	.	.
					$\Delta^4 y_{n-5}$
				$\Delta^3 y_{n-4}$	$\Delta^4 y_{n-4}$
			$\Delta^2 y_{n-3}$	$\Delta^3 y_{n-3}$	
.	.	Δy_{n-2}	$\Delta^2 y_{n-2}$		
x_{n-1}	y_{n-1}	Δy_{n-1}			
x_n	y_n				

Sai phân

Bảng sai phân lùi

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
x_0	y_0				
x_1	y_1	∇y_1			
.	.	∇y_2	$\nabla^2 y_2$		
			$\nabla^2 y_3$	$\nabla^3 y_3$	
				$\nabla^3 y_4$	$\nabla^4 y_4$
					$\nabla^4 y_5$
.	.	.	.	.	.
x_{n-4}	y_{n-4}	∇y_{n-4}	$\nabla^2 y_{n-4}$	$\nabla^3 y_{n-4}$	$\nabla^4 y_{n-4}$
x_{n-3}	y_{n-3}	∇y_{n-3}	$\nabla^2 y_{n-3}$	$\nabla^3 y_{n-3}$	$\nabla^4 y_{n-3}$
x_{n-2}	y_{n-2}	∇y_{n-2}	$\nabla^2 y_{n-2}$	$\nabla^3 y_{n-2}$	$\nabla^4 y_{n-2}$
x_{n-1}	y_{n-1}	∇y_{n-1}	$\nabla^2 y_{n-1}$	$\nabla^3 y_{n-1}$	$\nabla^4 y_{n-1}$
x_n	y_n	∇y_n	$\nabla^2 y_n$	$\nabla^3 y_n$	$\nabla^4 y_n$

Nội suy Newton

Ta sẽ tìm một đa thức nội suy có bậc không quá n sao cho trong đó các cặp (x_i, y_i) i = 0,1, ..., n, sao cho mỗi lần bổ sung thêm số liệu thì ta vẫn tận dụng được đa thức nội suy đã tính trước đó.
Đa thức đó có dạng:

$$p_n(x) = D_0(x) + D_1(x) + ... + D_n(x)$$

sao cho đa thức $p_i(x) = D_0(x) + D_1(x) + ... + D_i(x)$ là nội suy của mẫu $(x_0, y_0), (x_1, y_1), ..., (x_i, y_i)$ với mọi i $\leq$ n.
Từ những điều kiện trên, ta thấy $D_i(x)$ là đa thức bậc cao nhất là i, và các hệ số của nó chỉ phụ thuộc vào mẫu con $(x_0, y_0), (x_1, y_1), ..., (x_i, y_i)$.

Nội suy Newton

$D_0(x)$ là đa thức nội suy đi qua một điểm duy nhất (x_0, y_0), do đó đây là đa thức bậc 0, $D_0(x) = a_0$, trong đó a_0 là hằng số và ta có thể suy ra là $a_0 = y_0$.
$p_i(x)$ là đa thức nội suy của mẫu $(x_0, y_0), (x_1, y_1), ..., (x_i, y_i)$, ta có:

$$p_i(x_j) = D_0(x_j) + D_1(x_j) + ... + D_i(x_j) = y_j \ j = 0, 1, ..., i$$

Ta còn có

$$p_n(x_j) = D_0(x_j) + D_1(x_j) + ... + D_n(x_j) = y_j \ j = 0, 1, ..., i$$

Kết hợp hai đẳng thức trên ta suy ra:

$$D_k(x_j) = 0 \ \ k = i+1, i+2, ..., n; \ \ j = 0, 1, 2, ..., i$$

Nội suy Newton

Suy ra các đa thức $D_k(x)$ có dạng

$$D_k(x) = R(x)(x - x_0)(x - x_1)...(x - x_i), \ \ k = i + 1, i + 2, ..., n$$

Với k = i+1 thì đa thức $D_k(x)$ có bậc không cao hơn i+1, do đó nó có dạng

$$D_{i+1}(x) = a_{i+1}(x - x_0)(x - x_1)...(x - x_i)$$

trong đó a_{i+1} là hằng số phụ thuộc vào mẫu $(x_0, y_0), (x_1, y_1), ..., (x_{i+1}, y_{i+1})$.

Nội suy Newton

Như vậy đa thức nội suy được xây dựng có dạng

$$\begin{aligned} p_n(x) = a_0 + & a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots \\ & + a_i(x - x_0)(x - x_1)\dots(x - x_{i-1}) + \dots \\ & + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}) \end{aligned} \tag{5}$$

Trong đó a_i là hằng số phụ thuộc vào mẫu quan sát $(x_0, y_0), (x_1, y_1), \dots, (x_i, y_i)$.
Đây là công thức nội suy Newton tiến.

Nội suy Newton khoảng chia đều

Xét bài toán nội suy với các mốc cách đều, tức

$$x_{i+1} - x_i = h \ \forall i = 0, 1, ...$$

Ta cần xác định các hệ số trong công thức của đa thức nội suy (5).

Hệ số đa thức nội suy Newton với khoảng chia đều

Với trường hợp khoảng chia đều, các hệ số a_i trong (5) được tính bởi công thức:

$$a_i = \frac{\Delta^i y_0}{i!h^i}, \ \ i \leq n$$

Nội suy Newton với khoảng chia đều

Ví dụ 2: Cho bảng sai phân

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
2	23	70	96	48
4	93	166	144	48
6	259	310	192	48
8	569	502	240	48
10	1071	742	288	
12	1813	1030		
14	2843			

xây dựng đa thức nội suy Newton tiến bằng công thức với khoảng chia đều

Nội suy Newton tổng quát

Tỷ sai phân

Với số số nguyên $p \geq 0$ bất kỳ, ta định nghĩa tỷ sai phân cấp 1 là đại lượng

$$y[x_{p+1}, x_p] = \frac{y_{p+1} - y_p}{x_{p+1} - x_p}$$

tỷ sai phân cấp 2

y[xp+2,xp+1]=(yp+2-yp+1)/(xp+2-xp+1)

$$y[x_{p+2}, x_{p+1}, x_p] = \frac{y[x_{p+2}, x_{p+1}] - y[x_{p+1}, x_p]}{x_{p+2} - x_p}$$

...

tỷ sai phân cấp k

$$y[x_{p+k}, \dots, x_{p+1}, x_p] = \frac{y[x_{p+k}, \dots, x_{p+1}] - y[x_{p+k-1}, \dots, x_p]}{x_{p+k} - x_p}$$

Nội suy Newton tổng quát

Hệ số đa thức nội suy Newton tổng quát

Với trường hợp tổng quát, các hệ số a_i trong (5) được tính bởi công thức:

$$a_i = y[x_i, ..., x_1, x_0], \ \ i \leq n$$

Ta thường sử dụng bảng tỷ sai phân để tính toán và ghi nhớ hệ số:

X	y	t.s.p bậc 1	t.s.p bậc 2	t.s.p bậc 3	t.s.p bậc 4
X_0	y_0	$y[x_1,x_0]$	$y[x_2,x_1,x_0]$	$y[x_3,x_2,x_1,x_0]$	$y[x_4,x_3,x_2,x_1,x_0]$
X_1	y_1	$y[x_2,x_1]$	$y[x_3,x_2,x_1]$	$y[x_4,x_3,x_2,x_1]$	
X_2	y_2	$y[x_3,x_2]$	$y[x_4,x_3,x_2]$		
X_3	y_3	$y[x_4,x_3]$			
X_4	y_4				

Nội suy Newton

Ví dụ 3: Biết rằng đại lượng y là một tam thức bậc hai của x và tại x =0,78; 1,56; 2,34; 3,12; 3,81 các giá trị của y tuần tự là 2,5; 1,20; 1,12; 4,28. Hãy lập công thức y biểu diễn qua x.
Ví dụ 4: Cho hàm số $y = e^x$ tại x = 0,65 + 0,1i; i=0,1,...,5 tuần tự là 1,91554; 2,11700; 2,33965; 2,58571; 2,85765; 3,15819. Hãy tính ln2.

Phương pháp bình phương cực tiểu

Trong phương pháp nội suy đa thức đã xét, ta đòi hỏi đa thức phải đi qua các điểm quan sát. Điều này đôi khi là điều kiện quá chặt trong thực tế.
Sau đây ta sẽ xác định tham số của một hàm f(x) có dạng đã biết, sao cho các giá trị $f(x_i)$ xấp xỉ tốt nhất các giá trị yi theo một tiêu chuẩn gọi là bình phương cực tiểu.
Bây giờ ta xét một trường hợp hay được áp dụng nhất là hàm f(x) có dạng đa thức bậc m:

$$p_m(x) = a_0 + a_1x^1 + ...a_{m-1}x^{m-1} + a_mx^m$$

sao cho

$$y_i = p_m(x_i) + e_i, \ \forall i \in 1, 2, ..., n$$

ở đây e_i là sai số của đa thức xấp xỉ.

Phương pháp bình phương cực tiểu

Phương pháp bình phương cực tiểu là phương pháp xây dựng đa thức xấp xỉ sao cho tổng bình phương sai số là nhỏ nhất, tức:

$$S = \sum_{i=0}^{n} e_i^2 = \sum_{i=0}^{n} (y_i - \sum_{j=0}^{m} a_j x_i^j)^2$$

đạt giá trị cực tiểu.
Ta giải các hệ số bằng cách giải hệ phương trình đạo hàm riêng:

$$\frac{\delta S}{\delta a_k} = 0 \ \forall k = 0, 1, ..., m$$

Phương pháp bình phương cực tiểu

Hệ phương trình đạo hàm riêng trên có thể đưa về dạng $Ca = d$ với

$$C = \begin{bmatrix} \sum_{i=0}^{n} x_i^0 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \cdots & \sum_{i=0}^{n} x_i^m \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \cdots & \sum_{i=0}^{n} x_i^{m+1} \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^5 & \cdots & \sum_{i=0}^{n} x_i^{m+1} \\ \cdots & & & & & \\ \sum_{i=0}^{n} x_i^m & \sum_{i=0}^{n} x_i^{m+1} & \sum_{i=0}^{n} x_i^{m+2} & \sum_{i=0}^{n} x_i^{m+3} & \cdots & \sum_{i=0}^{n} x_i^{2m} \end{bmatrix}$$

và

$$d = (\sum_{i=0}^{n} y_i x_i^0, \sum_{i=0}^{n} y_i x_i^1, \ldots, \sum_{i=0}^{n} y_i x_i^m)$$

Phương pháp bình phương cực tiểu

Đặt $y = (y_0, y_1, ..., y_n)^T$, $e = (e_0, e_1, ..., e_n)^T$, $a = (a_0, a_1, ..., a_n)^T$ và

$$F = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^m \\ 1 & x_1 & x_1^2 & \dots & x_1^m \\ . & . & & \dots & . \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}$$

Hệ phương trình Ca = d có thể viết dưới dạng

$$F^T F a = F^T y$$

Chúng ta giải a bằng các phương pháp giải hệ phương trình tuyến tính đã biết.

Phương pháp bình phương cực tiểu

Phương pháp bình phương cực tiểu giải bài toán phi tuyến

$y = ae^{bx}, a > 0$

Lấy logarit 2 vế, ta có

$$lny = lna + bx$$

bài toán đưa về dạng tuyến tính!

$y = ax^{b}, a > 0$

Lấy logarit 2 vế, ta có

$$lny = lna + blnx$$

bài toán đưa về dạng tuyến tính!