第四章 数据建模

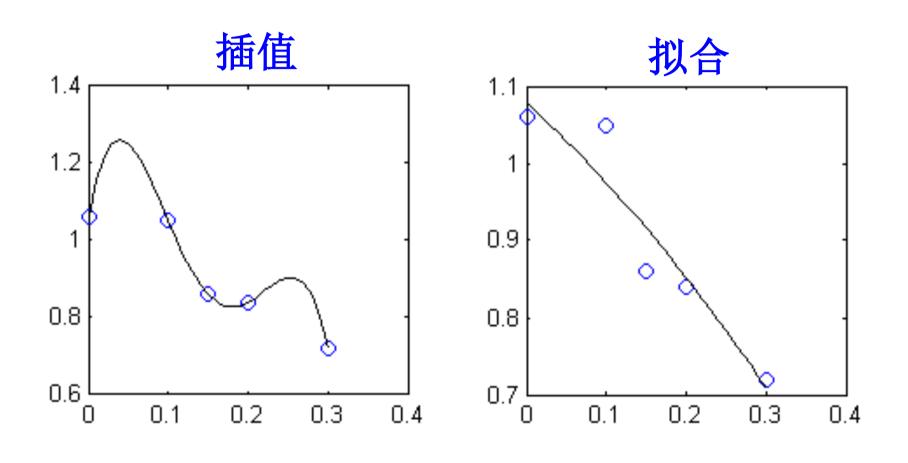
数据建模: 插值和拟合

数据建模:插值和拟合

已知函数 y=f(x)的一批数据(x₁, y₁), (x₂, y₂), …, (x_n, y_n), 若函数f(x)是已知的,但计算复杂,使用不方便或者函数是未知的,这时需要从某函数类(如多项式函数、样条函数等)中求一个函数 Φ (x)作为f(x)的近似,进而研究其规律,这类数值计算问题称为数据建模。

数据建模的两类方法:插值方法,拟合方法。

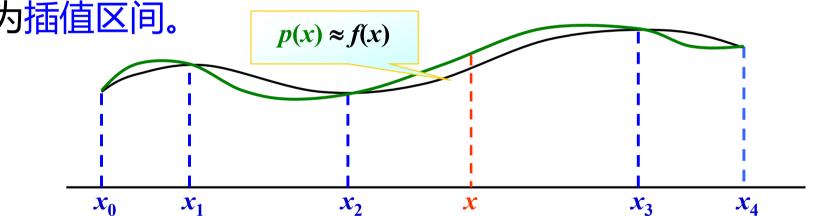
插值与拟合



(一致拟合、最小二乘拟合)

4.1 多项式插值

已知函数y=f(x)在区间[a,b]上有定义,且区间内存在若干点 x_i 上的函数值 $y_i=f(x_i)$ (i=0,1,...,n),一个插值问题就是求一个"简单"的函数p(x)满足 $p(x_i)=y_i$ 这时称p(x)为f(x)的插值函数,而f(x) 称为插值原函数, $x_0, x_1, ..., x_n$ 称为插值点。插值点的边界称为插值区间。



最常用的插值函数是 ...?多项式

当p(x)为不超过n次多项式时称为n次拉格朗日插值。

多项式插值(拉格朗日插值)

定义: 设 y = f(x) 在区间 [a, b] 上有定义,且已知它在n+1个 互异点 $a \le x_0 < x_1 ... < x_n \le b$ 上的函数值 $y_0, y_1, ..., y_n$,若 存在一个次数不超过 n 次的多项式 $p(x) = a_0 + a_1 x + ... + a_n x^n$

$$p(x) - a_0 + a_1x + \dots + a_nx$$

满足条件

$$p(x_i) = y_i \ (i = 0, \dots n)$$
 (4-1) 插值条件

则称p(x)为f(x)的n次插值多项式。 x_0 , x_1 , …, x_n 成为插值点,当固定点 \overline{x} 为一个插值点时,称 $f(\overline{x}) \approx p(\overline{x})$ 为 \overline{x} 点的插值。当 $\overline{x} \in [\min(x_0, x_1, ..., x_n), \max(x_0, x_1, ..., x_n)]$ 时,称为内插,否则称为外推。

插值多项式的唯一性

定理4. 1: 设节点 x_0 , x_1 , ..., x_n 互不相同,则存在唯一次数不超过n的多项式 $P_n(x)$ 满足 $P_n(x_i) = y_i (i = 0,1,...,n)$ **证明**: 设所要构造的插值多项式为:

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

由插值条件 $P_n(x_i) = y_i$, $i = 0, 1, \dots, n$
得到如下线性代数方程组:

$$\begin{cases} 1 \cdot a_0 + x_0 a_1 + \dots + x_0^n a_n = y_0 \\ 1 \cdot a_0 + x_1 a_1 + \dots + x_1^n a_n = y_1 \\ \dots \\ 1 \cdot a_0 + x_n a_1 + \dots + x_n^n a_n = y_n \end{cases}$$

此方程组的系数行列式为

$$D = \begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} = \prod_{0 \le j < i \le n} (x_i - x_j)$$

范德蒙行列式!

因 $x_i \neq x_j$, $i = 1,2,\cdots n$; $j = 1,2,\cdots n$ 时,

 $D \neq 0$, 因此, $P_n(x)$ 由 a_0 , a_1 , …, a_n 唯一确定。

注:该定理的证明过程实质上给出了一种求插值多项式的一个方法,但此方法不适合计算机求解(计算量大、病态方程组(n比较大时)),我们要寻找用计算机的求解方法。

插值基函数

令 $P_n(x) = \{ 次数不超过 n 的多项式的全体 \} , 则 <math>P_n(x)$ 构成一个 n+1 维线性空间,设其一组基为

$$\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$$

则插值多项式 $p_n(x)$ 可以被这组基线性表出,即:

$$p_{n}(x) = a_{0}\varphi_{0}(x) + a_{1}\varphi_{1}(x) + \dots + a_{n}\varphi_{n}(x)$$

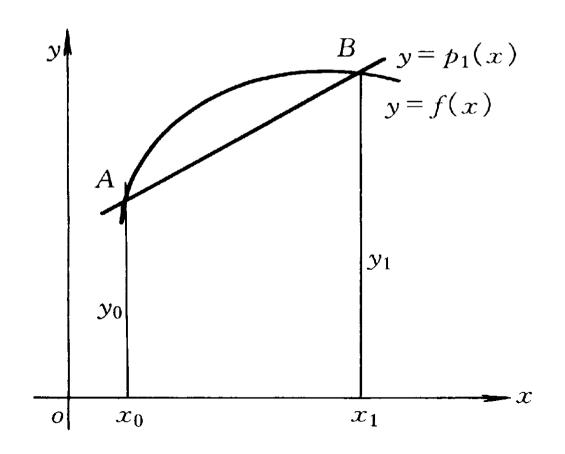
这样就可以通过不同的基来构造插值多项式 $p_n(x)$ 项,这样的方法称为基函数法。

基函数法基本步骤:

- 1) 寻找特殊的基函数组(插值基函数)
- 2) 确定插值多项式在这组基下的表示系数。

一次Lagrange插值多项式

已知函数y = f(x)在点 x_0, x_1 上的值为 y_0, y_1 ,要求多项式 $y = p_1(x)$ 使 $p_1(x_0) = y_0$, $p_1(x_1) = y_1$ 。 其几何意义就是通过两点 $A(x_0, y_0), B(x_1, y_1)$ 的一条直线,如图所示。



一次Lagrange插值多项式

由直线两点式可知,通过A,B 的直线方程为

$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = L_1(x)$$

它也可变形为

$$L_1(x) = \mathbf{p}_1(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1$$

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}, l_1(x) = \frac{x - x_0}{x_1 - x_0}$$

L. (x) = 6 (x) 10 + 6 (x) 4.

显然有:
$$l_0(x_0) = l_1(x_1) = 1$$
, $l_0(x_1) = l_1(x_0) = 0$, $L_1(x_0) = y_0$, $L_1(x_1) = y_1$,

一次Lagrange插值也称为线性插值。

2次Lagrange插值多项式

已知函数
$$y = f(x)$$
在点 x_0 , x_1 , x_2 上的值为 y_0 , y_1 , y_2 , 要求多项式 $y = L_2(x)$ 使 $L_2(x_0) = y_0$, $L_2(x_1) = y_1$, $L_2(x_2) = y_2$ 。

$$L_2(x) = y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x)$$

$$l_0(x_0) = l_1(x_1) = l_2(x_2) = 1;$$

 $l_0(x_1) = l_0(x_2) = 0; l_1(x_0) = l_1(x_2) = 0; l_2(x_0) = l_2(x_1) = 0$

$$l_{0}(x) = \frac{(x - x_{1})(x - x_{2})}{(x_{0} - x_{1})(x_{0} - x_{2})}$$

$$l_{1}(x) = \frac{(x - x_{0})(x - x_{2})}{(x_{1} - x_{0})(x_{1} - x_{2})}$$

$$l_{2}(x) = \frac{(x - x_{0})(x - x_{1})}{(x_{2} - x_{0})(x_{2} - x_{1})}$$

$$l_{2}(x) = \frac{(x - x_{0})(x - x_{1})}{(x_{2} - x_{0})(x_{2} - x_{1})}$$

抛物(线)插值 (二次拉格朗日插值)

n次拉格朗日(Lagrange)插值

已知函数y = f(x)在点 x_0 , x_1 , …, x_n 上的值为 y_0 , y_1 , …, y_n , 要求多项式 $y = L_n(x)$ 使 $L_n(x_0) = y_0$, $L_n(x_1) = y_1$, …, $L_n(x_n) = y_n$ 。

设
$$L_n(x) = y_0 l_0(x) + y_1 l_1(x) + \dots + y_n l_n(x)$$

其中, $l_k(x_i) = \begin{cases} 1, i = k \\ 0, i \neq k \end{cases}$, $k = 0, 1, \dots, n$

则称 $l_k(x)$ 为节点上的拉格朗日插值基函数。

由构造法可得

$$l_k(x) = rac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

可以证明 $l_0(x)$, $l_1(x)$, ..., $l_n(x)$ 线性无关,即它们构成线性空间 $L_n(x)$ 的一组基。

 $L_{\rm n}(x)$ 称为n次拉格朗日插值多项式。 程序4.1 P75

与节点有关, 但与*f(x)*无关.

插值余项

问题:如何估计用 $L_n(x)$ 近似f(x)的误差?

$$R_n(x) = f(x) - L_n(x)$$

$$R_n(x) = \int_{(n+1)}^{(n+1)} \frac{f^{(n+1)}(x)}{f^{(n+1)}(x)} \int_{(n+1)}^{n} (x-x) dx$$

定理 4.2 假设 $x_0,...,x_n \in [a,b], f(x)$ 在[a,b]上连续的n+1阶导数, $L_n(x)$ 为f(x) 关于节点 $x_0,...,x_n$ 的n次拉格朗日插值多项式,则对任意 $x \in [a,b]$,插值余项公式为:

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega(x)$$
$$\omega(x) = \prod_{i=0}^n (x - x_i)$$

插值举例

例:已知函数 $y = \ln x$ 的函数值如下

X	10	11	12	13	14
lnx	2.3026	2.3979	2.4849	2.5649	2.6391

试分别用线性插值和抛物插值计算 1n11.75的近似值。

解: 在插值计算中,为了减小截断误差,通常选取与插值点x 邻接的插值节点。

线性插值: 取 $x_0=11$, $x_1=12$ 得

$$L_1(x) = \frac{(x - x_1)}{(x_0 - x_1)} y_0 + \frac{(x - x_0)}{(x_1 - x_0)} y_1 = 0.087x + 1.4409 (1)$$

插值举例

将 x=11.75 代入 $L_1(x)$ 可得: $\ln 11.75 \approx L_1(11.75) \approx 2.4632$

$$L_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_1 - x_0)(x_1 - x_1)} y_2$$

将 x=11.75 代入可得:

可以计算出 In11.75 的近似值为:

 $\ln 11.75 \approx L_2(11.75) \approx 2.4638$

精确解

 $\ln 11.75 \approx 2.4638532405902$

可见,抛物插值的精度比线性插值要高。

Lagrange插值多项式简单方便, 只要取定节点就可写出基函数, 进而得到插值多项式。易于计算机实现。

插值误差举例

例:已知函数 $y = \ln x$ 的函数值如下

X	10	11	12	13	14
lnx	2.3026	2.3979	2.4849	2.5649	2.6391

试给出线性插值和抛物插值计算 ln11.75的误差。

解: 1. 线性插值误差
$$R_1(x) = \frac{f''(\xi)}{2!}(x - x_0)(x - x_1), \xi \in (x_0, x_1)$$

$$\[\forall f''(x) = -1/x^2, \frac{1}{2}x_0 = 11, x_1 = 12, \xi \in (11, 12) \]
$$|R_1(11.75)| = \left| \frac{f''(\xi)}{2}(11.75 - x_0)(11.75 - x_1) \right|$$

$$< \left| -\frac{1}{11^2 \times 2} \right| |(11.75 - 11)(11.75 - 12)|$$

$$\approx 0.77479 \times 10^{-3} < 0.5 \times 10^{-2},$$
有1+2=3位有效数字.$$

插值误差举例

2. 抛物线插值误差

$$R_2(x) = \frac{f^{(3)}(\xi)}{3!} (x - x_0)(x - x_1)(x - x_2), \xi \in (x_0, x_2)$$

$$\nabla f^{(3)}(x) = 2/x^3, \xi \in (11,13)$$
 $|f^{(3)}(\xi)| = 2/\xi^3 < 2/11^3$

$$|R_2(11.75)| < \left| -\frac{2}{11^3 \times 6} \right| |(11.75 - 11)(11.75 - 12)(11.75 - 13)|$$
≈ $0.58696 \times 10^{-5} < 0.5 \times 10^{-4}$, $有1+4=5$ 位有效数字.

$$|R_1(11.75)| \approx 0.77479 \times 10^{-3} < 0.5 \times 10^{-2}$$
, 有 $1+2=3$ 位有效数字.

高次插值通常优于低次插值

但绝对不是次数越高就越好...

4.埃尔米特(Hermite)插值

许多实际问题不但要求插值函数P(x)在插值节点处与被插函数f(x)有相同的函数值 $P(x_i) = f(x_i)$ (i = 0, 1, 2, ..., n),而且要求在有些节点或全部节点上与f(x)的导数值也相等,甚至要求高阶导数值也相等,即

$$P(x_i) = f(x_i), P'(x_i) = f'(x_i), \dots, P^{(m)}(x_i) = f^{(m)}(x_i)$$

$$(i=0,1,2,\dots,n),$$

能满足这种要求的插值问题就称为埃尔米特插值(Hermite)。 这是泰勒插值和拉格朗日插值的综合和推广。

埃尔米特(Hermite)插值

定义 已知 n+1个互异点上 $a=x_0 < x_1 < \cdots < x_n = b$ 的函数值 $f(x_i)$ 和导数值 $f'(x_i)$,若存在一个次数不超过2n+1的多项式H(x),满足

$$H(x_i) = f(x_i), \qquad H'(x_i) = f'(x_i) \qquad (i = 0, 1, \dots, n)$$

则称H(x)为f(x)的2n+1次埃尔米特(Hermite)插值。

$$\mathbf{H}(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{2n+1} x^{2n+1}$$



两点3次Hermite插值

具有节点的导函数值约束的插值

已知
$$y_0 = f(x_0), y_1 = f(x_1), y'_0 = f'(x_0), y'_1 = f'(x_1)$$

求不超过3次埃尔米特插值多项式H3(x),使满足

$$H_3(x_0) = y_0, H_3(x_1) = y_1, H_3'(x_0) = y_0', H_3'(x_1) = y_1'$$

- 存在唯一性
- 计算公式
- 余项公式(截断误差)

存在唯一性

设埃尔米特插值多项式为

$$H_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

$$a_0 + a_1 x_0 + a_2 x_0^2 + a_3 x_0^3 = y_0$$

$$a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 = y_1$$

$$a_1 + 2a_2 x_0 + 3a_3 x_0^2 = y_0'$$

$$a_1 + 2a_2 x_1 + 3a_3 x_1^2 = y_1'$$

$$D = \begin{vmatrix} 1 & x_0 & x_0^2 & x_0^3 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 0 & 1 & 2x_0 & 3x_0^2 \\ 0 & 1 & 2x_1 & 3x_1^2 \end{vmatrix} = -(x_1 - x_0)^4 \neq 0$$

 $D \neq 0$, 因此, $H_3(x) \oplus a_0$, a_1 , a_2 , a_3 唯一确定。

注:该定理的证明过程实质上给出了一种求埃尔米插值多项式的一个方法,但此方法不适合计算机求解(计算量大,尤其是方程组较大时,我们要寻找用计算机的求解方法。这里可采用求Lagrange插值多项式的基函数的方法。

导出计算公式 (基函数法)

设
$$H_3(x) = \alpha_0(x)y_0 + \alpha_1(x)y_1 + \beta_0(x)y_0' + \beta_1(x)y_1',$$

其中基函数 $\alpha_0(x), \alpha_1(x), \beta_0(x), \beta_1(x)$ 均为3次多项式,
且满足:

$$\begin{cases} \alpha_0(x_0) = 1 \\ \alpha_0(x_1) = 0 \end{cases} \qquad \begin{cases} \alpha_1(x_0) = 0 \\ \alpha_1(x_1) = 1 \end{cases} \\ \alpha'_0(x_0) = 0 \\ \alpha'_0(x_1) = 0 \end{cases} \qquad \begin{cases} \alpha_1(x_0) = 0 \\ \alpha'_1(x_0) = 0 \\ \alpha'_1(x_1) = 0 \end{cases} \\ \begin{cases} \beta_0(x_0) = 0 \\ \beta_0(x_1) = 0 \end{cases} \qquad \begin{cases} \beta_1(x_0) = 0 \\ \beta_1(x_1) = 0 \end{cases} \\ \beta'_0(x_1) = 0 \end{cases} \\ \beta'_0(x_1) = 0 \end{cases} \qquad \begin{cases} \beta'_1(x_1) = 0 \\ \beta'_1(x_1) = 0 \end{cases}$$

基函数求法:

$$\Re \alpha_0(x)$$

$$\begin{cases} \alpha_0(x_1) = 0 \\ \alpha'_0(x_1) = 0 \end{cases} \longrightarrow \alpha_0(x) = [a + b(x - x_0)](x - x_1)^2$$

$$\alpha_0(x_0) = 1$$
 $a = \frac{1}{(x_0 - x_1)^2}$

$$\alpha_0'(x_0) = 0$$
 $b = \frac{2}{(x_1 - x_0)(x_0 - x_1)^2}$

$$\alpha_0(x) = (1 + 2\frac{x - x_0}{x_1 - x_0})(\frac{x - x_1}{x_0 - x_1})^2$$

同理
$$\alpha_1(x) = (1+2\frac{x-x_1}{x_0-x_1})(\frac{x-x_0}{x_1-x_0})^2$$

或

$$\alpha_0(x) = (1 + 2l_1(x))l_0^2(x)$$

$$\alpha_1(x) = (1 + 2l_0(x))l_1^2(x)$$

其中

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}, l_1(x) = \frac{x - x_0}{x_1 - x_0}$$

 $l_0(x)$, $l_1(x)$ 为线性插值基函数。

曲
$$\begin{cases} \beta_0(x_0) = 0 \\ \beta_0(x_1) = 0 \end{cases}$$
 设 $\beta_0(x) = a(x - x_0)(x - x_1)^2$
$$\begin{cases} \beta'_0(x_0) = 1 \\ \beta'_0(x_1) = 0 \end{cases}$$

曲
$$\beta'_0(x_0)=1$$
 , 得
$$a = \frac{1}{(x_0 - x_1)^2}$$
 ,

$$\beta_0(x) = (x - x_0)(\frac{x - x_1}{x_0 - x_1})^2 = (x - x_0)l_0^2(x)$$

同理

$$\beta_1(x) = (x - x_1)(\frac{x - x_0}{x_1 - x_0})^2 = (x - x_1)l_1^2(x)$$

余项公式(截断误差)

Hermite余项

$$R_3(x) = f(x) - H_3(x) = \frac{f^{(4)}(\xi)}{4!} (x - x_0)^2 (x - x_1)^2$$
$$\xi \in \left[\min(x_0, x_1, \mathbf{x}), \max(x_0, x_1, \mathbf{x}) \right]$$

证明: 由插值条件知

$$R_3(x_0)=R_3'(x_0)=0, R_3(x_1)=R_3'(x_1)=0$$

取x异于 x_0 和 x_1 ,设

$$R_3(x) = C(x)(x - x_0)^2(x - x_1)^2$$

利用
$$f(x) - H_3(x) = C(x)(x - x_0)^2(x - x_1)^2$$

构造辅助函数

$$F(t) = f(t) - H_3(t) - C(x)(t - x_0)^2 (t - x_1)^2$$

显然,F(t)有三个零点 x_0 , x, x_1 ,由Rolle定理知, F'(t)至少有两个零点 t_0 , t_1 满足 x_0 < t_0 < t_1 < t_1 ,而 t_0 和 t_1 也是F'(t)零点,故F'(t)至少有四个相异零点.

反复应用Rolle定理, 得 $F^{(4)}(t)$ 至少有一个零点设为 $\xi \in (a, b)$

$$F(t) = f(t) - H_3(t) - C(x)(t - x_0)^2(t - x_1)^2$$

$$F^{(4)}(\xi) = f^{(4)}(\xi) - C(x) \times 4! = 0$$

$$C(x) = \frac{f^{(4)}(\xi)}{4!}$$

$$R_3(x) = C(x)(x - x_0)^2(x - x_1)^2 = \frac{f^{(4)}(\xi)}{4!}(x - x_0)^2(x - x_1)^2$$

Hermite余项

$$R_3(x) = f(x) - H_3(x) = \frac{f^{(4)}(\xi)}{4!} (x - x_0)^2 (x - x_1)^2$$

$$\xi \in [\min(x_0, x_1, \mathbf{x}), \max(x_0, x_1, \mathbf{x})]$$

比较:

Lagrange余项(n=3)

$$f(x) - L_3(x) = \frac{f^{(4)}(\xi)}{4!}(x - x_0)(x - x_1)(x - x_2)(x - x_3)$$

```
3. 利用函数 fcx) = 区区 100, 141, 144的值, 分别用注偿超值和 抛构超值本1000 的近似值, 治计计算结果的设度限和有效效率.
 解: fc100) = 10 fcx = 2x = fcx = 1x = fcx = - x = 2
                                               ∫ (a) : 3 × - ±
              fc144) = 12
 0 \neq \frac{1}{4} \neq \frac{1}{4} \neq \frac{1}{4}
(a \in A) = \frac{A - A_1}{A_1 - A_2} = \frac{A - 1a_1}{100 - 1a_1} = -\frac{A - 1a_2}{a_1}
     I_{1}(X) = \frac{X - X_{2}}{X_{1} - X_{2}} = \frac{X - 100}{121 - 100} = \frac{X - 100}{21}
      LICK) = 6 CK ) 4. + LICK ) 4. = - \frac{x - \text{ii}}{-1} \cdot 10 + \frac{x - 100}{-1} \cdot 11
     L. cits) = 10,7145 8 0,107145 x 10
     读盖指计 Riens; = |fens; - Liens; |
                                                         = 1250 (118-100)(115-121)
                                                          = \frac{100}{\pi} \ \ \display \frac{1}{2} \ \ \display \frac{1}{2} \ \din \frac{1}{2} \ \display \frac{1}{2} \ \d
                                                           $ 0.0 1145 $ 0.5 × 10-1
 为证 有效 數字
 θ 抛物戏插值
   (x-x, (x-x, ) = (x-10)(x-100) = (x-10)(x-100) = (x-10)(x-100)
   L_{1}(\lambda) = \frac{(\lambda - \lambda_{1})(\lambda - \lambda_{2})}{(\lambda - \lambda_{1})(\lambda - \lambda_{2})} = \frac{(\lambda - 100)(\lambda - 144)}{(100)(0.161 - 144)} = \frac{(\lambda - 100)(\lambda - 144)}{485}
  L_{*}(A) = \frac{(A - A_{*})(A - A_{*})}{(A_{*} - A_{*})(A - A_{*})} = \frac{(A - 100)(A - 101)}{(100 + 101)} = \frac{(A - 100)(A - 101)}{101a}
   Lacis = 6000 y. + 6000 y. + 4000 y.
                         =\frac{47+}{(112-121)(12-12+3)}\times 10^{-} - \frac{48?}{(112-100)(12-12+3)}\times 11^{-} + \frac{10.17}{(12-100)(12-171)}\times 17^{-}
                        = 10. 7118 = 0. 107118 × 10"
 设备估计 R_ clis> = |fclis) - 2clis)
                                                                   = 31 (115-100) (115-121) (115-144)
                                                                  3xc100)-4 x 15 x c-6) x c-19)
20,00165 さ 0.5 A 10-4 作品 有 公宝 整本 等
3. 水不超点4次多级式HCA),接满足HCO)=H'CO)=O, HCD=H'CD=I, HCD=I,
并含fcA)性满足上生是于HCA)的均原时,有出余项fcA)-HCA)。
解: 4次多项式HCA)=OA<sup>**</sup>· bA<sup>*</sup>· cA<sup>*</sup>· da + 2
               H'CA) & 40x3+36x+1CX+d
               Hco) : & : 0
               H'co; = d = 0
               Hci) = 0+6+6= 1
              H'C1) = 40+56+46 = 1
               H (2) = 160 + 86 + 46 = 1
              可知 a=本 b=-主 c=章
                   = #x cx -6x +9)
           = *** (x-5) (t) (x-0) (x-1) (x-4)
```