3　数据拟合

代数插值是根据给定的数据表(也称为函数表),按某些条件构造一个代数多项式 $p_n(x)$ 近似代替函数 $f(x)$,所要求的条件中有一个条件是 $p_n(x_i)=y_i(i=0,1,\cdots,n)$,即要求函数 $p_n(x)$ 经过点 (x_i,y_i). 但是,由于数据表中给定的数据 x_i 和 y_i 是从实验或测量中得到的,难免有些误差,而且有个别点的误差可能比较大. 用代数插值所得到的插值多项式 $p_n(x)$ 因为要经过点 (x_i,y_i),所以必然保留原来数据的一切误差. 这是我们所不希望的. 为了避免这种情况的发生,可以使用数据拟合法.

3.1　单变量数据拟合及最小二乘法

若给定的数据表表示的是一个量与另一个量的关系,则可以使用单变量数据拟合法寻找一个近似函数代替函数 $f(x)$.

单变量数据拟合法的一般过程是:先根据给定函数 $y=f(x)$ 的数据表(如表 3-1 所示),用几何描点法或凭经验选择一个近似函数 $F(x)$,以反映数据表中数据的一般趋势,然后使用最小二乘法确定 $F(x)$ 中的未知参数,从而得到 $f(x)$ 的近似函数 $F(x)$.

表 3-1　$y=f(x)$ 的数据表

x	x_1	x_2	$\cdots$	x_n
$y=f(x)$	y_1	y_2	$\cdots$	y_n

通常 $F(x)$ 称为拟合函数,$f(x)$ 称为被拟合函数.

从单变量数据拟合法的一般过程可以看到,在使用数据拟合法求拟合函数 $F(x)$ 时,要用到最小二乘法. 那么,什么是最小二乘法呢?

与插值法的目的一样,单变量数据拟合法也是要寻找一个近似函数 $F(x)$ 来近似代替 $f(x)$. 但它与插值法又有些不同,它不要求近似函数 $F(x)$ 一定经过点 (x_i,y_i). 很自然,人们总是希望能找到一个最好的函数来近似代替 $f(x)$. 现在的问题是,什么是"最好"的函数?"最好"的函数以什么标准来衡量?为了讨论这个问题,需要引入偏差的定义.

定义 3.1　若记 $\delta_i=f(x_i)-F(x_i)$,$i=1,2,\cdots,n$,则称 δ_i 为 $f(x)$ 与 $F(x)$ 在 x_i 处的偏差.

一般情况下,使用单变量数据拟合法能找到一个近似函数 $F(x)$,使它与 $f(x)$ 的偏差 δ_i 的平方和最小,即使

$$\sum_{i=1}^{n}\delta_i^2 = \sum_{i=1}^{n}[f(x_i) - F(x_i)]^2$$

最小. 能使偏差 δ_i 的平方和最小的函数就是“最好”的函数. 因此,可以以偏差 δ_i 的平方和最小作为原则选择近似函数 $F(x)$.

定义 3.2　以“偏差的平方和最小”为原则选择近似函数的方法称为最小二乘法.

下面介绍一个呈线性关系的数据拟合的例子.

例 3－1　已知一组实验数据如表 3－2 所示,试用单变量数据拟合法求其拟合函数.

表 3－2　例 3－1 数据表

x	-1	0	1	2	3	4	5	6
$y=f(x)$	10	9	7	5	4	3	0	-1

解　按照单变量数据拟合法的一般过程,根据给定的数据表(表 3－2),用几何描点法或凭经验选择近似函数,以反映数据表中数据的一般趋势。这里,用几何描点法选择近似函数,而用几何描点法选择近似函数要画出数据表中数据的散点图,散点图如图 3－1 所示.

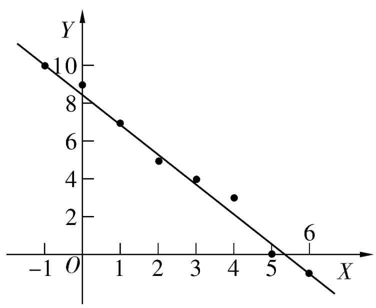

图 3－1　例 3－1 数据散点图

从图 3－1 可以看到,点 (x_i, y_i) $(i=1,2,\cdots,8)$ 在一条直线附近,这些点基本满足直线方程. 因此,可以选择线性函数来拟合这些数据,即可以选取

$$F(x) = a + bx$$

作为 $f(x)$ 的近似函数. 其中 a 和 b 为待定参数.

拟合函数选定之后,还要确定拟合函数中的待定参数. 确定拟合函数中的待定参数最常用的方法是最小二乘法。

先求出被拟合函数 $f(x)$ 与拟合函数 $F(x)$ 的偏差:

$$\delta_i = f(x_i) - F(x_i) = y_i - a - bx_i \quad (i=1,2,\cdots,8)$$

按照最小二乘法,要使偏差的平方和最小,即需选择 a 和 b,使

$$\sum_{i=1}^{8} \delta_i^2 = \sum_{i=1}^{8} (y_i - a - bx_i)^2 \tag{3.1}$$

最小.

显然,(3.1)右边是关于未知参数 a 和 b 的函数,所以可设

$$\sum_{i=1}^{8} \delta_i^2 = \sum_{i=1}^{8} (y_i - a - bx_i)^2 = \varphi(a,b)$$

这样一来,选择 a 和 b 使偏差的平方和最小的问题就转化成选择 a 和 b 使二元函数 $\varphi(a,b)$ 最小的问题. 而选择 a 和 b 使二元函数 $\varphi(a,b)$ 最小的问题实际上就是求二元函数 $\varphi(a,b)$ 的极小值. 根据求多元函数极小值的方法,先对 $\varphi(a,b)$ 分别求关于 a 和 b 的偏导数,得

$$\frac{\partial\varphi(a,b)}{\partial a} = -2\sum_{i=1}^{8} (y_i - a - bx_i) \tag{3.2}$$

$$\frac{\partial\varphi(a,b)}{\partial b} = -2\sum_{i=1}^{8} (y_i - a - bx_i)x_i \tag{3.3}$$

令(3.2)式和(3.3)式等于 0,整理后得

$$8a + \left(\sum_{i=1}^{8} x_i\right) b = \sum_{i=1}^{8} y_i \tag{3.4}$$

$$\left(\sum_{i=1}^{8} x_i\right) a + \left(\sum_{i=1}^{8} x_i^2\right) b = \sum_{i=1}^{8} x_i y_i \tag{3.5}$$

把(3.4)式和(3.5)式联立起来,就得到含有两个未知参数 a 和 b 的有两个方程的线性代数方程组. 它通常称为正规方程组.

把 x_i 和 y_i 代入正规方程组得

$$\begin{cases} 8a + 20b = 37 \\ 20a + 92b = 25 \end{cases} \tag{3.6}$$

解方程组(3.6)得 $a = 8.6429$, $b = -1.6071$.

于是,拟合函数

$$y = 8.6429 - 1.6071x$$

通过这个具体的例子,可以把单变量数据拟合法的一般步骤归纳如下:

①按给定数据表画出散点图;

②分析散点图,确定近似函数 $F(x)$ 的类型,以反映给定数据的一般趋势;

③用最小二乘法确定近似函数 $F(x)$ 的未知参数,从而得到最小二乘拟合函数 $F(x)$.

仿照例 3-1 的求解过程,不难得到如下定理.

定理 3.1 给定 $y=f(x)$ 的数据表(表 3-1),若点 (x_i, y_i) $(i=1,2,\cdots,n)$ 大体上满足线性函数,即最小二乘拟合函数为

$$F(x) = a + bx$$

则待定参数 a 和 b 是正规方程组

$$\begin{cases} na + \left(\sum_{i=1}^{n} x_i\right) b = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{3.7}$$

的解。

单变量线性拟合法算法

①读入数据 x_i 和 $y_i(i=1,2,\cdots,n)$.

②计算

$$s_1 = \sum_{i=1}^{n} x_i \qquad s_2 = \sum_{i=1}^{n} y_i \qquad s_3 = \sum_{i=1}^{n} x_i^2 \qquad s_4 = \sum_{i=1}^{n} x_i y_i$$

③解正规方程组

$$\begin{cases} na + s_1 b = s_2 \\ s_1 a + s_3 b = s_4 \end{cases}$$

即按下列公式求 a 和 b:

$$a = \frac{s_2 s_3 - s_1 s_4}{n s_3 - s_1^2} \qquad b = \frac{n s_4 - s_1 s_2}{n s_3 - s_1^2}$$

④输出 a 和 b.

3.2　多变量数据拟合

在实际问题中,很多问题反映的不是一个量与一个量的关系,而是一个量与若干个量的关系. 具体来说,是一个量由若干个量确定. 这就是所谓的多元函数问题. 在数学上,这若干个量通常称为自变量,而由这些自变量确定的量通常称为因变量. 若假设这些自变量为 $x_1,x_2,\cdots,x_k$,因变量为 y,则每经过一次实验或测量就会得到一组数据 $x_1,x_2,\cdots,x_k,y$,而经过 n 次实验或测量就会得到 n 组数据,由这 n 组数据构成的数据表如表 3－3 所示.

表 3－3　多变量拟合数据表

实验或测量次数	x_1	x_2	$\cdots$	x_k	$y=f(x_1,x_2,\cdots,x_k)$
1	x_{11}	x_{12}	$\cdots$	x_{1k}	y_1
2	x_{21}	x_{22}	$\cdots$	x_{2k}	y_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{nk}	y_n

根据表 3－3，希望能找到一个函数来近似表达这些量的关系. 要做到这一点，可以采用多变量数据拟合法.

多变量数据拟合法的一般过程是：先根据表 3－3 选择变量 y 与变量 $x_1,x_2,\cdots,x_k$ 的一个近似函数 $F(x_1,x_2,\cdots,x_k)$ 以反映 y 与变量 $x_1,x_2,\cdots,x_k$ 的函数关系，然后使用最小二乘法确定近似函数 $F(x_1,x_2,\cdots,x_k)$ 中的未知参数，从而得到 $F(x_1,x_2,\cdots,x_k)$.

通常 $F(x_1,x_2,\cdots,x_k)$ 称为拟合函数，$y=f(x_1,x_2,\cdots,x_k)$ 称为被拟合函数.

从多变量数据拟合法的一般过程可以看出，若使用这种方法求拟合函数，需根据表 3－3 选择一个近似函数 $F(x_1,x_2,\cdots,x_k)$. 这是多变量数据拟合的关键. 这与两个变量的情况不同，两个变量的情况可以使用几何描点法画出散点图辅助选择近似函数. 但对多变量的情况，一般来说，作图是困难的，通常是凭经验或根据实际问题的物理背景和一些专业知识来找.

为了说明多变量数据拟合的一般过程，现在讨论一种特殊情况. 假定表 3－3 中的数据呈线性关系，这时选择线性函数

$$F(x_1,x_2,\cdots,x_k)=a_0+a_1x_1+a_2x_2+\cdots+a_kx_k \tag{3.8}$$

来近似表达 y 与变量 $x_1,x_2,\cdots,x_k$ 的函数关系，其中 $a_0,a_1,\cdots,a_k$ 为待定参数.

按数据拟合的方法，要确定这些参数，需使用最小二乘法. 现在使用最小二乘法来确定待定参数 $a_0,a_1,\cdots,a_k$.

先把 $x_{m1},x_{m2},\cdots,x_{mk}(m=1,2,\cdots,n)$ 代入(3.8)式得

$$F(x_{m1},x_{m2},\cdots,x_{mk})=a_0+a_1x_{m1}+a_2x_{m2}+\cdots+a_kx_{mk}$$

则 $F(x_1,x_2,\cdots,x_k)$ 与 $y=f(x_1,x_2,\cdots,x_k)$ 在 $x_{m1},x_{m2},\cdots,x_{mk}$ 处的偏差为

$$\begin{aligned}\delta_m&=y_m-F(x_{m1},x_{m2},\cdots,x_{mk})\\&=y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk}\end{aligned}$$

偏差的平方和为

$$\sum_{m=1}^{n}\delta_m^2=\sum_{m=1}^{n}(y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk})^2\overset{\text{记为}}{=\!=}\varphi(a_0,a_1,\cdots,a_k)$$

根据最小二乘法，要选择近似函数使 $\sum\limits_{m=1}^{n}\delta_m^2$ 最小，而 $\sum\limits_{m=1}^{n}\delta_m^2$ 是关于 $a_0,a_1,\cdots,a_k$ 的函数，所以可以转化为选择 $a_0,a_1,\cdots,a_k$ 使 $\varphi(a_0,a_1,\cdots,a_k)$ 最小. 实际上就是求 $\varphi(a_0,a_1,\cdots,a_k)$ 的极小值.

根据多元函数求极小值的方法，对 $\varphi(a_0,a_1,\cdots,a_k)$ 分别求关于 $a_0,a_1,\cdots,a_k$ 的偏导数并令其等于0，得

$$\frac{\partial\varphi}{\partial a_0}=-2\sum_{m=1}^{n}(y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk})=0$$

$$\frac{\partial\varphi}{\partial a_1}=-2\sum_{m=1}^{n}(y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk})x_{m1}=0$$

$$\frac{\partial\varphi}{\partial a_2}=-2\sum_{m=1}^{n}(y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk})x_{m2}=0$$

$$\vdots$$

$$\frac{\partial\varphi}{\partial a_k}=-2\sum_{m=1}^{n}(y_m-a_0-a_1x_{m1}-a_2x_{m2}-\cdots-a_kx_{mk})x_{mk}=0$$

整理化简后联立起来得方程组(3.9)：

$$\begin{cases} na_0+a_1\sum_{m=1}^{n}x_{m1}+a_2\sum_{m=1}^{n}x_{m2}+\cdots+a_k\sum_{m=1}^{n}x_{mk}=\sum_{m=1}^{n}y_m \\ a_0\sum_{m=1}^{n}x_{m1}x_{m1}+a_1\sum_{m=1}^{n}x_{m1}+a_2\sum_{m=1}^{n}x_{m2}x_{m1}+\cdots+a_k\sum_{m=1}^{n}x_{mk}x_{m1}=\sum_{m=1}^{n}y_mx_{m1} \\ a_0\sum_{m=1}^{n}x_{m2}+a_1\sum_{m=1}^{n}x_{m1}x_{m2}+a_2\sum_{m=1}^{n}x_{m2}x_{m2}+\cdots+a_k\sum_{m=1}^{n}x_{mk}x_{m2}=\sum_{m=1}^{n}y_mx_{m2} \\ a_0\sum_{m=1}^{n}x_{mk}+a_1\sum_{m=1}^{n}x_{m1}x_{mk}+a_2\sum_{m=1}^{n}x_{m2}x_{mk}+\cdots+a_k\sum_{m=1}^{n}x_{mk}x_{mk}=\sum_{m=1}^{n}y_mx_{mk} \end{cases}$$

方程组(3.9)是一个含有 $k+1$ 个方程和 $k+1$ 个未知数的线性代数方程组,也称为正规方程组.解这个方程组就能得到 $a_0,a_1,\cdots,a_k$.

综上所述,得到如下定理.

定理 3.2 给定 $y=f(x_1,x_2,\cdots,x_k)$的数据表(表3-3),若数据表中的数据呈线性关系,这时选取线性函数

$$F(x_1,x_2,\cdots,x_k)=a_0+a_1x_1+a_2x_2+\cdots+a_kx_k$$

作为 y 与变量 $x_1,x_2,\cdots,x_k$ 的最小二乘拟合函数,则待定参数 $a_0,a_1,\cdots,a_k$ 是正规方程组(3.9)的解.

可以证明:当 $n>k$ 时,正规方程组(3.9)有惟一解.

例 3-2 已知一组测量数据如表 3-4 所示,求其线性拟合函数.

表 3-4 例 3-2 数据表

测量次数	x_1	x_2	$y=f(x_1,x_2)$
1	1	1	7
2	1	2	9
3	2	1	10
4	2	2	11
5	2	3	12

解 据题意,选择线性函数

$$F(x_1,x_2)=a_0+a_1x_1+a_2x_2\ (a_0,a_1,a_2\text{为待定参数})$$

拟合给定数据表中的数据.

由定理 3.2 得到正规方程组

$$\begin{cases} na_0 + a_1\sum_{m=1}^{5}x_{m1} + a_2\sum_{m=1}^{5}x_{m2} = \sum_{m=1}^{5}y_m \\ a_0\sum_{m=1}^{5}x_{m1} + a_1\sum_{m=1}^{5}x_{m1}x_{m1} + a_2\sum_{m=1}^{5}x_{m2}x_{m1} = \sum_{m=1}^{5}y_m x_{m1} \\ a_0\sum_{m=1}^{5}x_{m2} + a_1\sum_{m=1}^{5}x_{m1}x_{m2} + a_2\sum_{m=1}^{5}x_{m2}x_{m2} = \sum_{m=1}^{5}y_m x_{m2} \end{cases} \tag{3.10}$$

而 $\sum_{m=1}^{5}x_{m1}=8$, $\sum_{m=1}^{5}x_{m2}=9$, $\sum_{m=1}^{5}x_{m1}x_{m1}=14$, $\sum_{m=1}^{5}x_{m1}x_{m2}=\sum_{m=1}^{5}x_{m2}x_{m1}=15$, $\sum_{m=1}^{5}x_{m2}x_{m2}=19$, $\sum_{m=1}^{5}y_m=49$, $\sum_{m=1}^{5}y_m x_{m1}=82$, $\sum_{m=1}^{5}y_m x_{m2}=93$,把它们代入正规方程组(3.10)得

$$\begin{cases} 5a_0 + 8a_1 + 9a_2 = 49 \\ 8a_0 + 14a_1 + 15a_2 = 82 \\ 9a_0 + 15a_1 + 19a_2 = 93 \end{cases}$$

解方程组得 $a_0=3.8$, $a_1=2.4$, $a_2=1.2$.

于是,所求的拟合函数为

$$y=F(x_1,x_2)=3.8+2.4x_1+1.2x_2$$

多变量线性拟合法算法

①输入数据 x_{mi} 和 $y_m(m=1,2,\cdots,n;i=1,2,\cdots,k)$.

②计算正规方程组的系数:

$$l_{00}=n \qquad l_{0i}=l_{i0}=\sum_{m=1}^{n}x_{mi}\ (i=1,2,\cdots,k) \qquad g_0=\sum_{m=1}^{n}y_m$$

$$l_{ij}=\sum_{m=1}^{n}x_{mi}x_{mj}\ (i,j=1,2,\cdots,k) \qquad g_i=\sum_{m=1}^{n}y_m x_{mi}\ (i=1,2,\cdots,k)$$

③解正规方程组

$$\begin{cases} l_{00}a_0+l_{01}a_1+\cdots+l_{0k}a_k=g_0 \\ l_{10}a_0+l_{11}a_1+\cdots+l_{1k}a_k=g_1 \\ \qquad\vdots \\ l_{k0}a_0+l_{k1}a_1+\cdots+l_{kk}a_k=g_k \end{cases}$$

求出 $a_0,a_1,\cdots,a_k$.

④输出 $a_0,a_1,\cdots,a_k$.

3.3 非线性数据线性化

本章介绍的例 3－1 和例 3－2 都是线性拟合的例子. 在这两个例子中,数据与数据之间的关系呈线性关系. 如果数据与数据之间的关系呈线性关系,则可以按这两个例子所介绍的方法求拟合函数;否则,就不能直接使用这两个例子所介绍的方法求拟合函数. 但是,在这类问题中有一些问题,经过数据与数据之间的变换之后能得到线性关系,这时就可以直接使用前面介绍过的方法.

例 3－3 某炼钢厂出钢时用的钢包(用来装钢水的容器)是用特殊耐火材料制成的,在使用过程中,由于钢水及炉渣对包衬耐火材料的侵蚀,使其容量随着使用次数的增多而增大. 为了找出使用次数 x 与容量 y 之间的函数关系,工程技术人员做了 15 次测试,测试数据如表 3－5 所示. 试用数据拟合法找出使用次数 x 与容量 y 之间的函数关系.

表 3－5 例 3－3 数据表

i	使用次数 x_i	容量 y_i	i	使用次数 x_i	容量 y_i	i	使用次数 x_i	容量 y_i
1	2	6.42	6	7	10.00	11	12	10.60
2	3	8.20	7	8	9.93	12	13	10.80
3	4	9.58	8	9	9.99	13	14	10.60
4	5	9.50	9	10	10.49	14	15	10.90
5	6	9.70	10	11	10.59	15	16	10.76

解 为了找出使用次数 x 和容量 y 之间的函数关系,先画出散点图如图 3－2 所示.

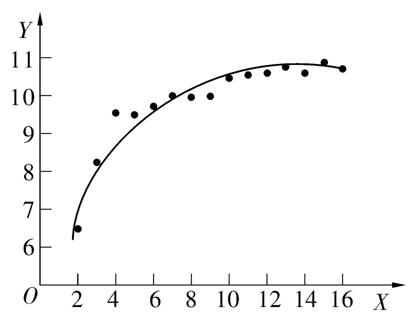

图 3－2 例 3－3 数据散点图

按照图 3－2 中散点的趋势,凭直观可以画出一条近似曲线,使这些点或者落在这条曲线上,或者落在曲线的两侧. 而这条近似曲线大致上像一条双曲线,因而可以把使用次数 x 与容量 y 之间的关系近似表示为

$$\frac{1}{y}=a+b\,\frac{1}{x}$$

这是双曲线方程的一种形式,其中 a 和 b 为待定参数.

为了确定待定参数 a 和 b,可以直接采用最小二乘法.但这样得出的方程组为非线性方程组,而非线性方程组一般是比较难求解的.

令 $X=\frac{1}{x}$和 $Y=\frac{1}{y}$,得到 $Y=a+bX$.这是一个线性函数,现取它作为拟合函数.

由关系式 $X=\frac{1}{x}$和 $Y=\frac{1}{y}$得到 $X_i=\frac{1}{x_i}(i=1,2,\cdots,15)$和 $Y_i=\frac{1}{y_i}(i=1,2,\cdots,15)$,从而得到一个新的数据表,如表 3-6 所示.

表 3-6 例 3-3 新数据表

i	x_i	y_i	$X_i=\frac{1}{x_i}$	$Y_i=\frac{1}{y_i}$	i	x_i	y_i	$X_i=\frac{1}{x_i}$	$Y_i=\frac{1}{y_i}$
1	2	6.42	0.5000	0.1558	9	10	10.49	0.1000	0.0953
2	3	8.20	0.3333	0.1220	10	11	10.59	0.0909	0.0944
3	4	9.58	0.2500	0.1044	11	12	10.60	0.0833	0.0943
4	5	9.50	0.2000	0.1053	12	13	10.80	0.0769	0.0926
5	6	9.70	0.1667	0.1031	13	14	10.60	0.0714	0.0943
6	7	10.00	0.1429	0.1000	14	15	10.90	0.0667	0.0917
7	8	9.93	0.1250	0.1007	15	16	10.76	0.0625	0.0929
8	9	9.99	0.1111	0.1001					

有了新数据表和选定的拟合函数,就可以按最小二乘法确定待定参数 a 和 b.由定理 3.1 得到正规方程组

$$\begin{cases} 15a+\left(\sum_{i=1}^{15}X_i\right)b=\sum_{i=1}^{15}Y_i \\ \left(\sum_{i=1}^{15}X_i\right)a+\left(\sum_{i=1}^{15}X_i^2\right)b=\sum_{i=1}^{15}X_iY_i \end{cases} \tag{3.11}$$

而 $\sum_{i=1}^{15}X_i\approx2.3807$, $\sum_{i=1}^{15}Y_i\approx1.5469$, $\sum_{i=1}^{15}X_i^2\approx0.5842$, $\sum_{i=1}^{15}X_iY_i\approx0.2727$.把它们代入正规方程组(3.11)得

$$\begin{cases} 15a+2.3807b=1.5469 \\ 2.3807a+0.5842b=0.2727 \end{cases}$$

解方程组得 $a\approx0.0824$,$b\approx0.1318$.故

$$Y=0.0824+0.1318X$$

即有

$$\frac{1}{y}=0.0824+0.1318\frac{1}{x}$$

于是,拟合函数

$$y=\frac{x}{0.0824x+0.1318}$$

例 3－4　已知一组数据如表 3－7 所示,求一个经验函数,形如 $y=ae^{bx}$（a,b 为常数）,使它与表 3－7 中的数据相拟合.

表 3－7　例 3－4 数据表

x	0	1	2	3	4
$y=f(x)$	1.5	2.5	3.5	5	7.5

解　据题意,表 3－7 中的数据大体上满足函数关系 $y=ae^{bx}$. 所以,表 3－7 的拟合函数为 $y=ae^{bx}$. 为了便于确定 a 和 b,可以通过变量变换把 $y=ae^{bx}$ 转换为线性函数,然后用最小二乘法来确定 a 和 b.

先对 $y=ae^{bx}$ 两边取自然对数得

$$\ln y=\ln a+bx$$

令 $Y=\ln y, A=\ln a, B=b$ 和 $X=x$,则有

$$Y=A+BX$$

这是一个线性函数,取它作为拟合函数.

由 $Y=\ln y$ 和 $X=x$ 得到 $Y_i=\ln y_i$ 和 $X_i=x_i$,从而得到新的数据表,如表 3－8 所示.

表 3－8　例 3－4 新数据表

i	x_i	y_i	X_i	$Y_i=\ln y_i$
1	0	1.5	0	0.405465
2	1	2.5	1	0.916291
3	2	3.5	2	1.252763
4	3	5.0	3	1.609438
5	4	7.5	4	2.014903

有了拟合函数,又有了新的数据表,就可以使用最小二乘法确定 a 和 b. 由定理 3.1 得到正规方程组

$$\begin{cases} 5A+\left(\sum_{i=1}^{5}X_i\right)B=\sum_{i=1}^{5}Y_i \\ \left(\sum_{i=1}^{5}X_i\right)A+\left(\sum_{i=1}^{5}X_i^2\right)B=\sum_{i=1}^{5}X_iY_i \end{cases} \tag{3.12}$$

而 $\sum_{i=1}^{5}X_i=10$, $\sum_{i=1}^{5}Y_i\approx 6.19886$, $\sum_{i=1}^{5}X_i^2=30$, $\sum_{i=1}^{5}X_iY_i\approx 16.309743$,把它们代入

正规方程组(3.12)得

$$\begin{cases} 5A+10B=6.19886 \\ 10A+30B=16.309743 \end{cases}$$

解上述方程组得 $A\approx0.457367$, $B\approx0.3912023$. 又由 $A=\ln a$ 得 $a\approx1.579910$, $b=B\approx0.3912023$.

于是,所求的拟合函数为

$$y=1.579910\mathrm{e}^{0.3912023x}$$

在实践中,有时要进行多项式拟合. 多项式拟合也可以采用非线性数据线性化的方法.

设有两个量 z 和 y 基本满足 m 次多项式函数,经过实验或测量得到数据表(表 3-9),从而可以把

$$y=a_0+a_1z+a_2z^2+\cdots+a_mz^m \tag{3.13}$$

作为给定数据的拟合函数,其中 $a_0,a_1,\cdots,a_m$ 为待定参数.

表 3-9 多项式拟合数据表

z	z_1	z_2	$\cdots$	z_n
$y=f(z)$	y_1	y_2	$\cdots$	y_n

令 $x_1=z, x_2=z^2,\cdots,x_m=z^m$,则由(3.13)式有

$$y=a_0+a_1x_1+a_2x_2+\cdots+a_mx_m \xlongequal{\text{记为}} F(x_1,x_2,\cdots,x_m)$$

使它变成多变量线性拟合.

在数据拟合法中,当选定拟合函数后,非线性拟合函数能否线性化是首先要考虑的问题. 为了方便读者对非线性拟合函数进行线性化,表 3-10 中给出了常用非线性拟合函数线性化的方法.

表 3-10 常用非线性拟合函数线性化的方法

非线性拟合函数的形式	对非线性拟合函数线性化为 $Y=A+BX$	需做的变换
$y=a+b\ln x$	$y=a+b\ln x$	$X=\ln x, Y=y$ $A=a, B=b$
$y=a\mathrm{e}^{bx}$	$\ln y=\ln a+bx$	$X=x, Y=\ln y$ $A=\ln a, B=b$
$y=a+\dfrac{b}{x}$	$y=a+b\dfrac{1}{x}$	$X=\dfrac{1}{x}, Y=y$ $A=a, B=b$

续上表

非线性拟合函数的形式	对非线性拟合函数线性化为 $Y = A + BX$	需做的变换
$y = \dfrac{x}{a + bx}$	$\dfrac{1}{y} = b + a\dfrac{1}{x}$	$X = \dfrac{1}{x}, Y = \dfrac{1}{y}$ $A = b, B = a$
$y = (a + bx)^{-2}$	$y^{-\frac{1}{2}} = a + bx$	$X = x, Y = y^{-\frac{1}{2}}$ $A = a, B = b$
$y = axe^{-bx}$	$\ln\dfrac{y}{x} = \ln a - bx$	$X = x, Y = \ln\dfrac{y}{x}$ $A = \ln a, B = -b$
$y = \dfrac{1}{1 + ae^{bx}}$	$\ln\left(\dfrac{1}{y} - 1\right) = \ln a + bx$	$X = x, Y = \ln\left(\dfrac{1}{y} - 1\right)$ $A = \ln a, B = b$

3.4 正交多项式拟合

多项式拟合的做法是:先把多项式拟合函数变成多变量拟合函数,然后按照多变量拟合法求出多项式的系数 $a_0, a_1, \cdots, a_m$. 在多变量拟合法中,要解一个正规方程组. 理论上可以证明该正规方程组有惟一解,但当多项式的次数比较高时,正规方程组就会变成病态方程组. 所谓病态方程组,就是如果方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 中的系数矩阵 $\boldsymbol{A}$ 和常数项 $\boldsymbol{b}$ 有微小变化,就会引起方程组的解很大变化,则方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 就是一个病态方程组. 为了克服这个缺点,把多项式拟合函数取为

$$\begin{aligned} y^* &= \Psi_m(x) = a_0 p_0(x) + a_1 p_1(x) + \cdots + a_m p_m(x) \\ &= \sum_{k=0}^{m} a_k p_k(x) \end{aligned} \tag{3.14}$$

其中 $a_0, a_1, \cdots, a_m$ 为待定参数,$p_k(x)(k = 0, 1, \cdots, m)$是 k 次多项式.

很显然,$y^* = \Psi_m(x)$是一个 m 次多项式函数.

把 x_i 代入(3.14)式得到 y_i 的近似值 y_i^*,则 y_i^* 与 y_i 就可能产生偏差. 记 y_i^* 与 y_i 的偏差为 δ_i,有

$$\delta_i = y_i - y_i^* = y_i - \sum_{k=0}^{m} a_k p_k(x_i) \quad (i = 1, 2, \cdots, n)$$

为了求得 $a_0, a_1, \cdots, a_m$,按最小二乘法选择 $a_0, a_1, \cdots, a_m$ 使误差 δ_i 的平方和最小,即选择 $a_0, a_1, \cdots, a_m$,使 $\sum\limits_{i=1}^{n} \delta_i^2$ 最小. 由于从实验或测量中得到的数据的精度

不同,通常在每一个 δ_i 的前面乘上一个表示数据精度的权数 α_i. 即要使

$$\sum_{i=1}^{n}(\alpha_i\delta_i)^2=\sum_{i=1}^{n}\alpha_i^2\delta_i^2=\sum_{i=1}^{n}\omega_i\delta_i^2$$

最小. 其中 $\omega_i=\alpha_i^2$ 称为权因子. 而

$$\sum_{i=1}^{n}\omega_i\delta_i^2=\sum_{i=1}^{n}\omega_i\left[y_i-\sum_{k=0}^{m}a_kp_k(x_i)\right]^2\overset{\text{记为}}{=\!=}\varphi(a_0,a_1,\cdots,a_m)$$

因此,可以把选择 $a_0,a_1,\cdots,a_m$ 使 $\sum\limits_{i=1}^{n}\omega_i\delta_i^2$ 最小的问题转化为求函数 $\varphi(a_0,a_1,\cdots,a_m)$ 极小值的问题. 按多元函数求极小值的方法,对函数 $\varphi(a_0,a_1,\cdots,a_m)$ 分别求关于 $a_0,a_1,\cdots,a_m$ 的导数,令其为 0 并联立起来得到方程组

$$\frac{\partial\varphi}{\partial a_j}=-2\sum_{i=1}^{n}\omega_i\left[y_i-\sum_{k=0}^{m}a_kp_k(x_i)\right]p_j(x_i)=0\quad(j=0,1,\cdots,m)\tag{3.15}$$

将方程组(3.15)整理后得

$$\sum_{i=1}^{n}\omega_i\left[\sum_{k=0}^{m}a_kp_k(x_i)p_j(x_i)\right]-\sum_{i=1}^{n}\omega_i\,y_ip_j(x_i)=0$$

即

$$\sum_{k=0}^{m}a_k\left[\sum_{i=1}^{n}\omega_ip_k(x_i)p_j(x_i)\right]=\sum_{i=1}^{n}\omega_i\,y_ip_j(x_i)\tag{3.16}$$

令 $c_{jk}=\sum\limits_{i=1}^{n}\omega_ip_k(x_i)p_j(x_i)$, $b_j=\sum\limits_{i=1}^{n}\omega_i\,y_ip_j(x_i)$,则方程组(3.16)可以写成

$$\sum_{k=0}^{m}c_{jk}a_k=b_j\tag{3.17}$$

方程组(3.17)也称为正规方程组.

显然,如果能找到 $p_k(x)$ 满足下列关系式:

$$c_{jk}=\sum_{i=1}^{n}\omega_ip_j(x_i)p_k(x_i)=0\quad(j\neq k)\tag{3.18}$$

$$c_{jj}=\sum_{i=1}^{n}\omega_ip_j(x_i)^2>0\quad(j,k=0,1,\cdots,m)\tag{3.19}$$

则正规方程组(3.17)变为

$$c_{kk}a_k=b_k\tag{3.20}$$

由(3.20)式可以求出

$$a_k=\frac{b_k}{c_{kk}}\tag{3.21}$$

定义 3.3 满足(3.18)式和(3.19)式的多项式 $p_k(x)\,(k=0,1,\cdots,m)$ 通常称为正交多项式簇,确切地说,称为对数据 x_i 和对应的权数 ω_i 的正交多项式簇.

常用的正交多项式簇有一些,在这里只介绍一个常用的等距节点正交多项式簇.

假设给定一组 $n+1$ 个等距节点 $\xi_i(i=0,1,\cdots,n)$,它们的间隔为 h,选取权数 $\omega_i=1$,引入变换

$$x=\frac{\xi-\xi_0}{h}$$

则 ξ_i 变为 $x_i=i$ 等 $n+1$ 个整数等距节点.

构造多项式

$$p_{m,n}(x)=\sum_{k=0}^{m}(-1)^k\binom{m}{k}\binom{m+k}{k}\frac{x^{(k)}}{n^{(k)}}$$

其中 $x^{(k)}=x(x-1)\cdots(x-k+1)$ 且 $x^{(0)}=1$. 可以证明,$p_{m,n}(x)$是在 $x_i=i$ 等$n+1$个整数等距节点上权数 $\omega_i=1$ 的正交多项式簇,其前 6 个多项式为

$$p_{0,n}(x)=1$$

$$p_{1,n}(x)=1-2\frac{x}{n}$$

$$p_{2,n}=1-6\frac{x}{n}+6\frac{x(x-1)}{n(n-1)}$$

$$p_{3,n}=1-12\frac{x}{n}+30\frac{x(x-1)}{n(n-1)}-20\frac{x(x-1)(x-2)}{n(n-1)(n-2)}$$

$$p_{4,n}=1-20\frac{x}{n}+90\frac{x(x-1)}{n(n-1)}-140\frac{x(x-1)(x-2)}{n(n-1)(n-2)}+70\frac{x(x-1)(x-2)(x-3)}{n(n-1)(n-2)(n-3)}$$

$$p_{5,n}=1-30\frac{x}{n}+210\frac{x(x-1)}{n(n-1)}-560\frac{x(x-1)(x-2)}{n(n-1(n-2)}+630\frac{x(x-1)(x-2)(x-3)}{n(n-1)(n-2)(n-3)}+252\frac{x(x-1)(x-2)(x-3)(x-4)}{n(n-1)(n-2)(n-3)(n-4)}$$

练习与思考

1. 求方程组

$$\begin{cases}x_1+2x_2=1\\x_1+x_2=0\\x_1-x_2=1\end{cases}$$

的最小二乘解.

2. 给定数据表如表 3-11 所示,求形如 $y=a+bx^2$ 的拟合函数.

表 3 - 11 数据表

x	19	25	31	38	44
y	19	32.3	49	73.3	97.8

3. 用最小二乘法求一个形如 $y=\dfrac{1}{a+bx}$ 的经验公式，使之与如下数据(表 3 - 12)相拟合.

表 3 - 12 数据表

x	1	1.4	1.8	2.2	2.6
y	0.931	0.473	0.297	0.224	0.168

4. 在某个低温过程中，函数 y 依赖于温度 θ(℃)的测试数据如表 3 - 13 所示，而且已知经验公式是 $y=a\theta+b\theta^2$，试用最小二乘法确定 a 和 b.

表 3 - 13 数据表

θ	1	2	3	4
y	0.8	1.5	1.8	2

5. 给定数据表如表 3 - 14 所示，试用三次多项式拟合表中的数据.

表 3 - 14 数据表

x	-2	-1	0	1	2
y	-0.1	0.1	0.4	0.9	1.6

6. 在某一化学反应里，据实验得到分解生成物的浓度与时间的数据如表 3 - 15 所示，试使用正交多项式拟合法构造一个 5 次多项式求 y 的近似值.

表 3 - 15 数据表

时间 t(min)	0	5	10	15	20	25	30	35	40	45	50
浓度 y($\times 10^{-4}$)	0	1.27	2.16	2.86	3.44	3.87	4.15	4.37	4.51	4.60	4.66