第二章 最小二乘拟合

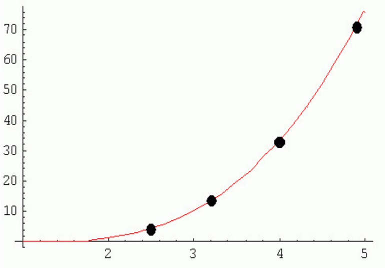
中国科学技术大学 数学学院

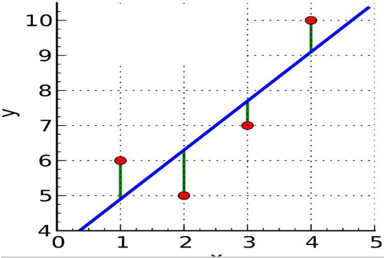
chenxjin@ustc.edu.cn

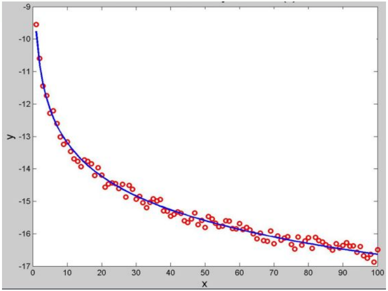
2020

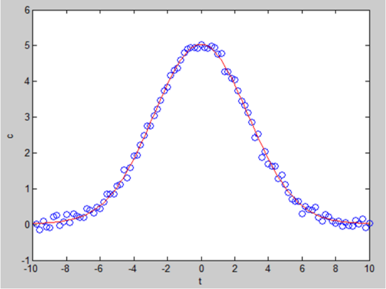
插值与拟合: 都是构造近似函数的方法

插值与拟合: 都是构造近似函数的方法









曲线拟合

对给定的一组离散数据，希望在给定的某函数类中找出一个函数来近似或逼近未知函数， 这种手段或方法

曲线拟合

对给定的一组离散数据，希望在给定的某函数类中找出一个函数来近似或逼近未知函数， 这种手段或方法 即是所谓的插值。

曲线拟合

对给定的一组离散数据，希望在给定的某函数类中找出一个函数来近似或逼近未知函数， 这种手段或方法 即是所谓的插值。 然而，在实际中，那些数据不可避免会存在着误差（或称“ 噪音”)； 这些误差会直接影响到我们所构造的插值函数， 更为严重的是，数据之间可能会包含 “矛盾”， 导致插值函数无法构造。 显然，此时采用插值，是很不好或很不科学的，甚至会失败！

曲线拟合

对给定的一组离散数据，希望在给定的某函数类中找出一个函数来近似或逼近未知函数， 这种手段或方法 即是所谓的插值。 然而，在实际中，那些数据不可避免会存在着误差（或称“ 噪音”)； 这些误差会直接影响到我们所构造的插值函数， 更为严重的是，数据之间可能会包含 “矛盾”， 导致插值函数无法构造。 显然，此时采用插值，是很不好或很不科学的，甚至会失败！

因此，我们可采用另一种近似未知函数的方法—拟合，使得我们的近似函数

1. 不要求通过所有的数据点（可以有力地消除数据误差的影响）
2. 尽可能表现离散数据的趋势，尽最大可能地靠近那些数据点

向量范数

定义：映射 $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}^+\cup\{0\}$如满足：

- 非负性　　$\|X\|\geq 0$, 且 $\|X\|=0\Leftrightarrow X=0$
- 齐次性　　$\forall a\in\mathbb{R}, \|aX\|=|a|\cdot\|X\|$
- 三角不等式　　$\|X+Y\|\leq\|X\|+\|Y\|$

则称该映射为向量的范数.

向量范数

定义：映射 $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}^+\cup\{0\}$如满足：

- 非负性 $\|X\|\geq 0$, 且 $\|X\|=0\Leftrightarrow X=0$
- 齐次性 $\forall a\in\mathbb{R}, \|aX\|=|a|\cdot\|X\|$
- 三角不等式 $\|X+Y\|\leq\|X\|+\|Y\|$

则称该映射为向量的范数.

设 $X=(x_1,x_2,\cdots,x_n)$. 几种常见的范数有：

1. $\|X\|_1=\sum\limits_{i=1}^{n}|x_i|\longrightarrow$ 1-范数
2. $\|X\|_2=\sqrt{\sum\limits_{i=1}^{n}x_i^2}\longrightarrow$ 2-范数
3. $\|X\|_\infty=\max\{|x_i|\}\longrightarrow$ ∞-范数

由范数可以定义$\mathbb{R}^n$中点 X, Y 的距离为$\|X-Y\|$.

$\mathbb{R}^n$中的范数等价性

$\mathbb{R}^n$空间中的任两种范数$\|\cdot\|_p, \|\cdot\|_q$是等价的(记为$\|\cdot\|_p \sim \|\cdot\|_q$)，即存在常数$c, C > 0$满足

$$c\|x\|_p \leq \|x\|_q \leq C\|x\|_p, \ \forall x \in \mathbb{R}^n.$$

$\mathbb{R}^n$中的范数等价性

$\mathbb{R}^n$空间中的任两种范数$\|\cdot\|_p$, $\|\cdot\|_q$是等价的(记为$\|\cdot\|_p \sim \|\cdot\|_q$)，即存在常数$c, C > 0$满足

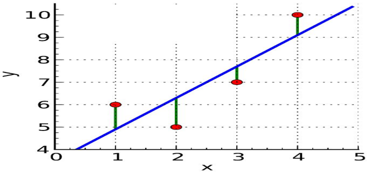
$$c\,\|x\|_p \leq \|x\|_q \leq C\,\|x\|_p, \ \forall x \in \mathbb{R}^n.$$

例如：

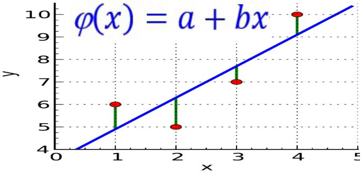
$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\,\|x\|_\infty$$
$$\|x\|_\infty \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_\infty$$

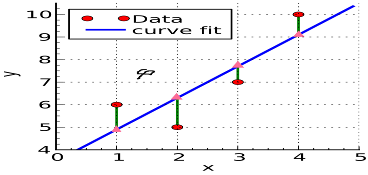
拟合标准–“距离”最小

拟合标准–“距离”最小



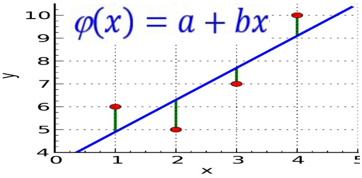
拟合标准–“距离”最小

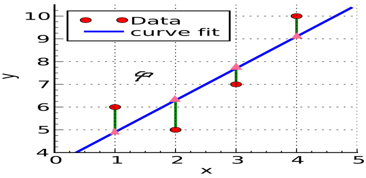




$$(x_1, y_1)$$
$$(x_2, y_2)$$
$$\vdots$$
$$(x_m, y_m)$$

拟合标准–“距离”最小





$$\begin{matrix} (x_1, y_1) \\ (x_2, y_2) \\ \vdots \\ (x_m, y_m) \end{matrix} \qquad \left\| \begin{pmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \vdots \\ \varphi(x_m) \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \right\|$$

拟合标准—“距离”最小

样本

(x_1, y_1)

(x_2, y_2)

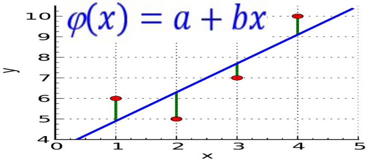
$\vdots$

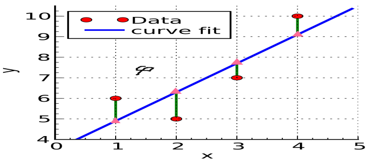
(x_m, y_m)

$\varphi(x) = ?$

$$\min \left\| \begin{pmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \vdots \\ \varphi(x_m) \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \right\|$$

拟合标准–“距离”最小





(x_1, y_1)

(x_2, y_2)

$\vdots$

(x_m, y_m)

图 3-2 一条直线公路与多个景点

那么，只能要求所做逼近函数 $\Phi(\mathbf{x})$ 最优的靠近样点，即向量 $\Phi(\mathbf{x}) = (\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_m))^T$ 和 $\mathbf{y} = (y_1, y_2, \cdots, y_m)^T$ 的误差或距离最小。按照 $\Phi(\mathbf{x})$ 与 $\mathbf{y}$ 之间误差最小的标准作为“最优”构造的逼近函数，称为拟合函数。函数 $\Phi(\mathbf{x})$ 最优也就是 $\Phi(\mathbf{x})$ 与 $\mathbf{y}$ 的距离最小 $(min\|\Phi(\mathbf{x}) - \mathbf{y}\|)$。

插值和拟合是构造逼近函数的两种方法。 $\varphi(x) = ?$

⋆ **拟合标准**

向量 Φ 与 $\mathbf{y}$ 之间的误差最小：$\min\|\Phi(\mathbf{x}) - \mathbf{y}\|$

向量 Φ 与 $\mathbf{y}$ 之间的误差或距离有多种定义方式。

用各点误差绝对值（1 范数）的和表示：

拟合标准

向量 $\Phi(\mathbf{x})$ 与 $\boldsymbol{y}$ 之间的误差或距离有各种不同的定义方式. 例如: 各种向量范数

$$\boldsymbol{R}_1 \triangleq \sum_{i=1}^{m} |\varphi(x_i) - y_i|,$$

$$\boldsymbol{R}_2 \triangleq \sqrt{\sum_{i=1}^{m} (\varphi(x_i) - y_i)^2},$$

$$\boldsymbol{R}_\infty \triangleq \max_{1 \leqslant i \leqslant n} |\varphi(x_i) - y_i|$$

特别地,

$$Q = \sum_{i=1}^{m} (\varphi(x_i) - y_i)^2$$

称为误差平方和, 由于计算最小值的方法容易实现而被广泛采用. 使误差平方和最小的原则称为最小二乘法. 本章主要讲述用最小二乘法构造拟合函数 $\varphi(x)$.

最小二乘问题

已知

问题：$f(x)$为定义在区间$[a,b]$上的函数，$\{x_i\}_{i=1}^m$ 为区间上m个互不相同 的点，Φ为给定的某一函数类。求Φ上的函数$\phi(x)$满足

$$\min \left\| \Big(f(x_1), f(x_2), \cdots, f(x_m) \Big) - \Big(\phi(x_1), \phi(x_2), \cdots, \phi(x_m) \Big) \right\|$$

即距离达到最小.

最小二乘问题

已知

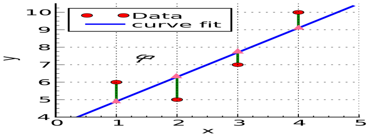
问题：$f(x)$为定义在区间$[a, b]$上的函数，$\{x_i\}_{i=1}^{m}$ 为区间上m个互不相同 的点，Φ为给定的某一函数类。求Φ上的函数$\phi(x)$满足

$$\min \left\| \left(f(x_1), f(x_2), \cdots, f(x_m) \right) - \left(\phi(x_1), \phi(x_2), \cdots, \phi(x_m) \right) \right\|$$

即距离达到最小.

在二范数意义下使得上述距离达到最小，构造拟合曲线的问题称为最小二乘问题; 构造的方法称为最小二乘法.

线性拟合



在图中，设公路为直线 $p(x)=a+bx$，景点的坐标 $(x_i,y_i), i=1,2,\cdots,m$；均方误差：

$$Q(a,b)=\sum_{i=1}^{m}(p(x_i)-y_i)^2=\sum_{i=1}^{m}(a+bx_i-y_i)^2$$

确定直线方程 $\leftrightarrow$ 计算 $\min Q(a,b)$ 的最小值。

由多元函数的极值原理，$\min Q(a,b)$ 的极小值要满足.

$$\begin{cases} \dfrac{\partial Q(a,b)}{\partial a}=0 \\ \dfrac{\partial Q(a,b)}{\partial b}=0 \end{cases}$$

线性拟合

得到满足最小均方误差的法方程组（正规方程）：

$$\begin{cases} \dfrac{\partial Q(a,b)}{\partial a} = 2\sum_{i=1}^{m}(a+bx_i-y_i)\cdot 1 = 0 \\ \dfrac{\partial Q(a,b)}{\partial b} = 2\sum_{i=1}^{m}(a+bx_i-y_i)\cdot x_i = 0 \end{cases}$$

整理得到满足最小均方误差的法方程组（正规方程）：

$$\begin{cases} ma + \left(\sum_{i=1}^{m} x_i\right) b = \sum_{i=1}^{m} y_i \\ \left(\sum_{i=1}^{m} x_i\right) a + \left(\sum_{i=1}^{m} x_i^2\right) b = \sum_{i=1}^{m} x_i y_i \end{cases}$$

写成矩阵形式：

$$\begin{pmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{pmatrix} \tag{3.1}$$

用消元法或克来姆方法解出方程组。

线性拟合的法方程组（正规方程）

$$\begin{array}{cc} & \begin{array}{cc} 1 & x \end{array} \\ \begin{array}{c} 1 \\ x \end{array} & \begin{pmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{pmatrix} \end{array} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{array}{c} y \\ \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{pmatrix} \end{array}$$

线性拟合的法方程组（正规方程）

$$\begin{array}{c|cc} & 1 & x \\ \hline 1 & & \\ x & & \end{array} \quad \begin{pmatrix} a \\ b \end{pmatrix} = \begin{array}{c} y \\ \hline \\ \\ \end{array}$$

线性拟合

$$a = \frac{\left(\sum_{i=1}^{m} x_i^2\right)\left(\sum_{i=1}^{m} y_i\right) - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} x_i y_i\right)}{m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$b = \frac{m\sum_{i=1}^{m} x_i y_i - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} y_i\right)}{m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

例 2.1 下表是 1950 至 1959 年我国的人口数据资料 (单位: 亿人).

年份 x_i	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
人口 y_i	5.52	5.63	5.75	5.88	6.03	6.15	6.28	6.46	6.60	6.72

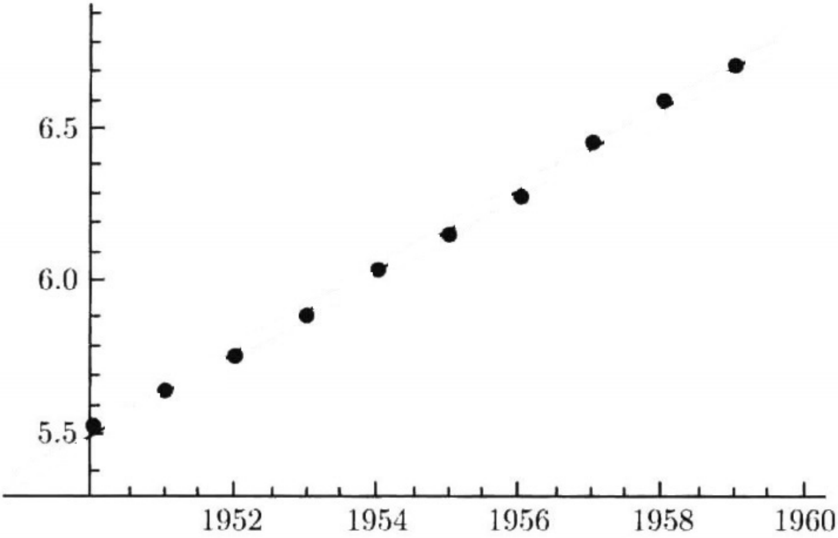


图 2.3　人口增长的线性模型

例 2.1 下表是 1950 至 1959 年我国的人口数据资料 (单位: 亿人).

年份 x_i	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
人口 y_i	5.52	5.63	5.75	5.88	6.03	6.15	6.28	6.46	6.60	6.72

解 观察数据的散点图, 发现数据点大约在一条直线上, 故可进行线性拟合. 设

$$\varphi(x) = a + b(x - 1950)$$

$$Q(a,b) = \sum_{i=1}^{10} (a + b(x_i - 1950) - y_i)^2$$

解得当 $a = 5.48945, b = 0.136121$ 时, $Q(a,b)$ 达最小值 0.00331879.

$$\varphi(x) = 5.48945 + 0.136121(x - 1950)$$

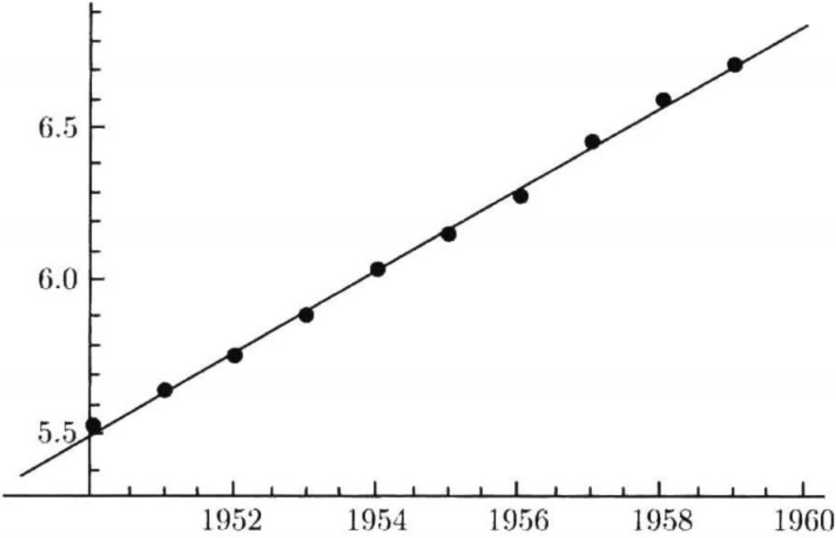


图 2.3　人口增长的线性模型

取函数类为多项式函数空间 $\mathbb{P}_n$，此时最小二乘问题为：对给定的 $\{(x_i, y_i)\}_{i=1}^m$，
求 $\phi(x) \in \mathbb{P}_n$ $(m \geq n+1)$ 使得均方误差

$$R = \sum_{i=1}^{m} \left(\phi(x_i) - y_i\right)^2$$

达到最小. 此时ϕ 称为多项式拟合函数.

最小二乘问题之多项式拟合$\Longleftrightarrow$多元函数的极值问题

取函数类为多项式函数空间 $\mathbb{P}_n$，此时最小二乘问题为：对给定的 $\{(x_i, y_i)\}_{i=1}^m$，
求 $\phi(x) \in \mathbb{P}_n$ $(m \geq n+1)$ 使得均方误差

$$R = \sum_{i=1}^{m} \left(\phi(x_i) - y_i\right)^2$$

达到最小. 此时ϕ 称为多项式拟合函数.

设 $\phi(x) = \sum_{j=0}^{n} a_j x^j$，则 $R = \sum_{i=1}^m \left(\phi(x_i) - y_i\right)^2$ 为系数 $\{a_j\}_{j=0}^n$ 的多元函数 (微积分角度).

由多元函数的极值条件，得：

$$\frac{\partial R}{\partial a_j} = 2x_i^j \sum_{i=1}^{m} \left(\phi(x_i) - y_i\right) = 0 \Longrightarrow x_i^j \sum_{i=1}^{m} \phi(x_i) = \sum_{i=1}^{m} x_i^j y_i, \quad \text{i.e.}$$

$$\begin{cases} a_0 m + a_1 \sum_{i=1}^{m} x_i + \cdots + a_n \sum_{i=1}^{m} x_i^n = \sum_{i=1}^{m} y_i, & j = 0; \\ a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^{n+1} = \sum_{i=1}^{m} x_i y_i, & j = 1; \\ \vdots & \vdots \\ a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + \cdots + a_n \sum_{i=1}^{m} x_i^{2n} = \sum_{i=1}^{m} x_i^n y_i, & j = n. \end{cases}$$

写成矩阵形式:

$$
\begin{pmatrix}
m & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \cdots & \sum_{i=1}^{m} x_i^n \\
\sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \cdots & \sum_{i=1}^{m} x_i^{n+1} \\
\sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \sum_{i=1}^{m} x_i^4 & \cdots & \sum_{i=1}^{m} x_i^{n+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{m} x_i^n & \sum_{i=1}^{m} x_i^{n+1} & \sum_{i=1}^{m} x_i^{n+2} & \cdots & \sum_{i=1}^{m} x_i^{2n}
\end{pmatrix}
\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}
=
\begin{pmatrix}
\sum_{i=1}^{m} y_i \\
\sum_{i=1}^{m} x_i y_i \\
\sum_{i=1}^{m} x_i^2 y_i \\
\vdots \\
\sum_{i=1}^{m} x_i^n y_i
\end{pmatrix}
$$

称之为多项式拟合函数 $\phi(x) = \sum_{j=0}^{n} a_j x^j$ 的法方程.

$$\begin{pmatrix} m & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \cdots & \sum_{i=1}^{m} x_i^n \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \cdots & \sum_{i=1}^{m} x_i^{n+1} \\ \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \sum_{i=1}^{m} x_i^4 & \cdots & \sum_{i=1}^{m} x_i^{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{m} x_i^n & \sum_{i=1}^{m} x_i^{n+1} & \sum_{i=1}^{m} x_i^{n+2} & \cdots & \sum_{i=1}^{m} x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{m} x_i^n y_i \end{pmatrix}$$

称之为多项式拟合函数 $\phi(x) = \sum_{j=0}^{n} a_j x^j$ 的法方程.

求多项式拟合函数的一般步骤:

1. 画出所给数据点的粗略图形，确定拟合次数 n;

2. 列表计算 $\sum_{i=1}^{m} x_i^k \ (0 \le k \le 2n)$, $\sum_{i=1}^{m} x_i^k y_i \ (0 \le k \le n)$, 给出法方程.

3. 解法方程，求出 a_i，最后给出拟合函数 $\phi(x)$.

最小二乘问题$\Longleftarrow$矛盾方程组(线代)

线性方程组$Ax=b(A\in\mathbb{R}_{m\times n})$，当rank($A$)$\neq$ rank(A,b)时无解，称为矛盾方程组. 通常，当方程个数多于未知数个数时，为矛盾方程组.

最小二乘问题$\Longleftarrow$矛盾方程组(线代)

线性方程组$Ax = b (A \in \mathbb{R}_{m\times n})$，当rank($A$)$\neq$ rank(A, b)时无解，称为矛盾方程组. 通常，当方程个数多于未知数个数时，为矛盾方程组.

问题：求解这样的$x \in \mathbb{R}^n$，使得 $\|Ax - b\|_2$最小.

最小二乘问题$\Longleftarrow$矛盾方程组(线代)

线性方程组$Ax=b(A\in\mathbb{R}_{m\times n})$，当rank($A$)$\neq$ rank(A,b)时无解，称为矛盾方程组. 通常，当方程个数多于未知数个数时，为矛盾方程组.

问题：求解这样的$x\in\mathbb{R}^n$，使得 $\|Ax-b\|_2$最小.

定理：

1. 设$m>n$，$A\in\mathbb{R}_{m\times n}$，$b\in\mathbb{R}_{m\times 1}$，rank($A$)=$n$. $A^TAx=A^Tb$ 称为矛盾方程组$Ax=b$的法方程， 法方程恒有唯一解.

2. x为$\min\|Ax-b\|_2$的解 $\Longleftrightarrow$ $A^TAx=A^Tb$，即x为法方程的解.

1. 法方程$A^TAx = A^Tb$恒有唯一解.

1. 法方程$A^TAx = A^Tb$恒有唯一解.

Pf：由rank(A)=$n$$\Longrightarrow$$\exists P_{m\times m}$可逆,

$s.t.$ $PA = \begin{pmatrix} \bar{A}_n \\ O \end{pmatrix}$ $\&|\bar{A}_n| \neq 0.$

1. 法方程$A^TAx=A^Tb$恒有唯一解.

Pf：由rank(A)=$n$$\Longrightarrow\exists P_{m\times m}$可逆,

$s.t.$ $PA=\begin{pmatrix}\bar{A}_n\\ O\end{pmatrix}$ & $|\bar{A}_n|\neq 0$.

$\Longrightarrow PA(PA)^T=PAA^TP^T=\begin{pmatrix}\bar{A}_n\bar{A}_n^T & O\\ O & O\end{pmatrix}$.

1. 法方程$A^TAx=A^Tb$恒有唯一解.

Pf：由rank(A)=$n\Longrightarrow\exists P_{m\times m}$可逆,

$s.t.\ PA=\begin{pmatrix}\bar{A}_n\\ O\end{pmatrix}$ & $|\bar{A}_n|\neq 0.$

$$\Longrightarrow PA(PA)^T=PAA^TP^T=\begin{pmatrix}\bar{A}_n\bar{A}_n^T & O\\ O & O\end{pmatrix}.$$

由$\bar{A}_n$可逆，故

$$\text{rank}(\bar{A}_n\bar{A}_n^T)=n=\text{rank}(AA^T)=\text{rank}(A^TA)$$

即A^TA可逆. $\Longrightarrow A^TAx=A^Tb$有唯一解.

2. x为$\min \|Ax-b\|_2$的解 $\Longleftrightarrow$ x为法方程$A^TAx=A^Tb$的解.

2. x为$\min \|Ax-b\|_2$的解 $\Longleftrightarrow$ x为法方程$A^TAx=A^Tb$的解.

Pf: "$\Longleftarrow$"：x满足法方程$A^TAx=A^Tb$. $\forall y\in\mathbb{R}_{n\times 1}$，设$e=y-x$. 则

$$\begin{aligned}
&\|Ay-b\|_2^2=(Ay-b)^T\cdot(Ay-b)=(Ax+Ae-b)^T\cdot(Ax+Ae-b)\\
=\;&(Ax-b)^T\cdot(Ax-b)+(Ax-b)^TAe+(Ae)^T(Ax-b)+(Ae)^T\cdot(Ae)\\
=\;&\|Ax-b\|_2^2+\|Ae\|_2^2+x^TA^TAe-b^TAe+e^TA^TAx-e^TA^Tb\\
=\;&\|Ax-b\|_2^2+\|Ae\|_2^2+0\\
\geq\;&\|Ax-b\|_2^2
\end{aligned}$$

$$\Longleftrightarrow \|Ax-b\|_2\leq\|Ay-b\|_2,\quad\forall y$$

即x为$\min\|Ax-b\|_2$的解.

2. x为$\min \|Ax-b\|_2$的解 $\Longleftrightarrow$ x为法方程$A^TAx=A^Tb$的解.

Pf: "$\Longleftarrow$"：x满足法方程$A^TAx=A^Tb$. $\forall y \in \mathbb{R}_{n\times 1}$，设$e=y-x$. 则

$$\begin{aligned}
& \|Ay-b\|_2^2 = (Ay-b)^T\cdot(Ay-b) = (Ax+Ae-b)^T\cdot(Ax+Ae-b) \\
= \; & (Ax-b)^T\cdot(Ax-b)+(Ax-b)^TAe+(Ae)^T(Ax-b)+(Ae)^T\cdot(Ae) \\
= \; & \|Ax-b\|_2^2+\|Ae\|_2^2+x^TA^TAe-b^TAe+e^TA^TAx-e^TA^Tb \\
= \; & \|Ax-b\|_2^2+\|Ae\|_2^2+0 \\
\geq \; & \|Ax-b\|_2^2
\end{aligned}$$

$$\Longleftrightarrow \|Ax-b\|_2 \leq \|Ay-b\|_2, \quad \forall y$$

即x为$\min \|Ax-b\|_2$的解.

"$\Longrightarrow$"：设$x=(x_1,x_2,\cdots,x_n)$，由极小值条件$\frac{\partial(\|Ax-b\|_2^2)}{\partial x_i}=0 \Longrightarrow$ $A^TAx=A^Tb$.

在多项式拟合问题中, 设待拟合函数

为 $\phi(x) = \sum_{i=0}^{n} a_i x^i$, 对给定的一组数据 $\{(x_i, y_i)\}_{i=1}^{m}$,

即要找一组系数$\{a_i\}_{i=0}^{n}$使得 $\|Aa - y\|_2$ 最小;

在多项式拟合问题中，设待拟合函数

为 $\phi(x) = \sum_{i=0}^{n} a_i x^i$，对给定的一组数据 $\{(x_i, y_i)\}_{i=1}^{m}$，

即要找一组系数$\{a_i\}_{i=0}^{n}$使得 $\|Aa - y\|_2$ 最小；其中

$$A = \begin{pmatrix} 1 & x_1 & \cdots & x_1^n \\ 1 & x_2 & \cdots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \cdots & x_m^n \end{pmatrix}$$

$$\begin{cases} a = (a_0, a_1, \cdots, a_n)^T \\ y = (y_1, y_2, \cdots, y_m) \end{cases} \quad (m > n+1)$$

由前面定理，即是要求方程组 $A^TAa = A^Ty$ (i.e., 法方程)的解

$$
\begin{pmatrix}
m & \sum_{i=1}^{m} x_i & \cdots & \sum_{i=1}^{m} x_i^n & \\
\sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \cdots & \sum_{i=1}^{m} x_i^{n+1} & \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{m} x_i^n & \sum_{i=1}^{m} x_i^{n+1} & \cdots & \sum_{i=1}^{m} x_i^{2n} &
\end{pmatrix}
\begin{pmatrix}
a_0 \\ a_1 \\ \vdots \\ a_n
\end{pmatrix}
=
\begin{pmatrix}
\sum_{i=1}^{m} y_i \\
\sum_{i=1}^{m} x_i y_i \\
\vdots \\
\sum_{i=1}^{m} x_i^n y_i
\end{pmatrix}
$$

Remark

1. 求最小二乘问题的多项式拟合，可这样理解：先假定拟合（多项式）函数$\phi(x)$通过所有的数据点$\{(x_i, y_i)\}$，即得到关于函数$\phi(x)$的多项式系数的矛盾方程组. 再解矛盾方程组（两边左乘系数矩阵的转置再解线性方程组）即得ϕ；这与直接由法方程求解的结果是一样的.
2. 法方程组一般是病态的，阶数越高，$\{x_i\}$相差量级越大或偏离原点越远，病态越严重. 通常要用高精度来求解.
3. 若给定的函数空间Φ 不是 $\mathbb{P}_n$，但 Φ=Span$\{\phi_0, \cdots, \phi_n\}$是某个有限维线性空间(可能是 $\mathbb{P}_n$的某线性子空间)，以上过程仍然成立; e.g., Φ=Span$\{1, x^3\}$.
4. 若给定的函数空间Φ不是线性空间，通常采用先作变元代换，将其转化为多项式拟合的问题；此时结果虽已不是最小二乘意义下最小， 但也有直观的拟合意义.

例 1：对以下数据作形如$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

例 1：对以下数据作形如$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^3$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^3=14.3 \\ a+b(-2)^3=8.3 \\ a+b(-1)^3=4.7 \\ a+b\times 2^3=8.3 \\ a+b\times 4^3=22.7 \end{cases}，即 \begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ 1 & (-1)^3 \\ 1 & 2^3 \\ 1 & 4^3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

例 1：对以下数据作形如$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^3$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^3=14.3 \\ a+b(-2)^3=8.3 \\ a+b(-1)^3=4.7 \\ a+b\times 2^3=8.3 \\ a+b\times 4^3=22.7 \end{cases}，即 \begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ 1 & (-1)^3 \\ 1 & 2^3 \\ 1 & 4^3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

其法方程为

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ (-3)^3 & (-2)^3 & (-1)^3 & 2^3 & 4^3 \end{pmatrix}\begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ \vdots & \vdots \\ 1 & 4^3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 58.3 \\ 1062 \end{pmatrix}$$

例 1：对以下数据作形如$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^3$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^3=14.3 \\ a+b(-2)^3=8.3 \\ a+b(-1)^3=4.7 \\ a+b\times 2^3=8.3 \\ a+b\times 4^3=22.7 \end{cases}，即 \begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ 1 & (-1)^3 \\ 1 & 2^3 \\ 1 & 4^3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

其法方程为

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ (-3)^3 & (-2)^3 & (-1)^3 & 2^3 & 4^3 \end{pmatrix} \begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ \vdots & \vdots \\ 1 & 4^3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 58.3 \\ 1062 \end{pmatrix}$$

$\Rightarrow a=10.6751,\ b=0.1368.$

例 1：对以下数据作形如$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^3$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^3=14.3 \\ a+b(-2)^3=8.3 \\ a+b(-1)^3=4.7 \\ a+b\times 2^3=8.3 \\ a+b\times 4^3=22.7 \end{cases}，即 \begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ 1 & (-1)^3 \\ 1 & 2^3 \\ 1 & 4^3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

其法方程为

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ (-3)^3 & (-2)^3 & (-1)^3 & 2^3 & 4^3 \end{pmatrix}\begin{pmatrix} 1 & (-3)^3 \\ 1 & (-2)^3 \\ \vdots & \vdots \\ 1 & 4^3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 58.3 \\ 1062 \end{pmatrix}$$

$\Rightarrow a=10.6751,\ b=0.1368.\ \Longrightarrow$ 拟合函数为$\phi(x)=10.6751+0.1368x^3$.

例 2：对以下数据作形如$a+bx^2$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

例 2：对以下数据作形如$a+bx^2$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^2$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^2=14.3 \\ a+b(-2)^2=8.3 \\ a+b(-1)^2=4.7 \\ a+b\times 2^2=8.3 \\ a+b\times 4^2=22.7 \end{cases}，\text{即} \begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ 1 & (-1)^2 \\ 1 & 2^2 \\ 1 & 4^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

例 2：对以下数据作形如$a+bx^2$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^2$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^2=14.3 \\ a+b(-2)^2=8.3 \\ a+b(-1)^2=4.7 \\ a+b\times 2^2=8.3 \\ a+b\times 4^2=22.7 \end{cases}，即 \begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ 1 & (-1)^2 \\ 1 & 2^2 \\ 1 & 4^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

其法方程为

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ (-3)^2 & (-2)^2 & (-1)^2 & 2^2 & 4^2 \end{pmatrix}\begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ \vdots & \vdots \\ 1 & 4^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 58.3 \\ 563 \end{pmatrix}$$

例 2：对以下数据作形如$a+bx^2$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	14.3	8.3	4.7	8.3	22.7

解：设$\phi(x)=a+bx^2$通过所有给定的数据点，则得矛盾方程组

$$\begin{cases} a+b(-3)^2=14.3 \\ a+b(-2)^2=8.3 \\ a+b(-1)^2=4.7 \\ a+b\times 2^2=8.3 \\ a+b\times 4^2=22.7 \end{cases}\text{，即}\begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ 1 & (-1)^2 \\ 1 & 2^2 \\ 1 & 4^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 14.3 \\ 8.3 \\ 4.7 \\ 8.3 \\ 22.7 \end{pmatrix}.$$

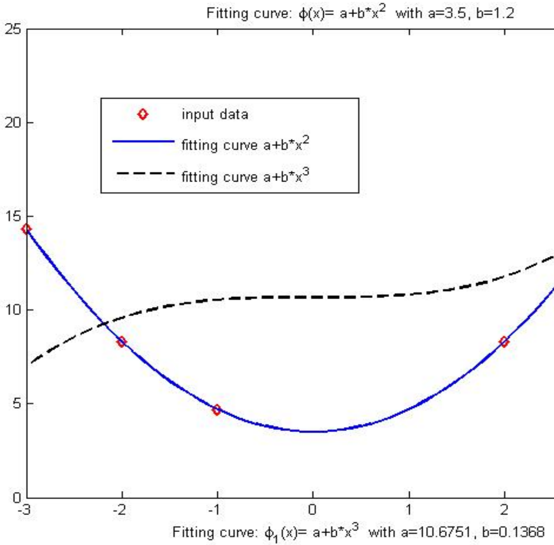
其法方程为

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ (-3)^2 & (-2)^2 & (-1)^2 & 2^2 & 4^2 \end{pmatrix}\begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ \vdots & \vdots \\ 1 & 4^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 58.3 \\ 563 \end{pmatrix}$$

$\Rightarrow a=3.5,\ b=1.2$. 故拟合函数$\phi(x)=3.5+1.2x^2$.

拟合结果

拟合结果



Fitting curve: ϕ(x)= a+b*x^2 with a=3.5, b=1.2
input data
fitting curve a+b*x^2
fitting curve a+b*x^3
25
20
15
10
5
0
-3
-2
-1
0
1
2
3
4
Fitting curve: ϕ_1(x)= a+b*x^3 with a=10.6751, b=0.1368

例 2'：分别对以下数据作形如$a+bx^2$以及$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	-14.3	-8.3	4.7	8.3	22.7

例 2'：分别对以下数据作形如$a+bx^2$以及$a+bx^3$的曲线拟合.

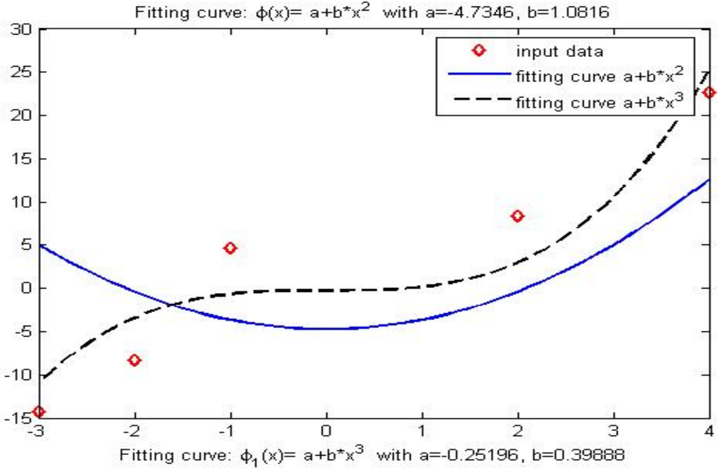
x_i	-3	-2	-1	2	4
y_i	-14.3	-8.3	4.7	8.3	22.7

求解过程与例1-2完全类似；

例 2'：分别对以下数据作形如$a+bx^2$以及$a+bx^3$的曲线拟合.

x_i	-3	-2	-1	2	4
y_i	-14.3	-8.3	4.7	8.3	22.7

求解过程与例1-2完全类似；拟合结果如下：



例 3：对以下数据作形如ae^{bx}的曲线拟合.

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

例 3：对以下数据作形如ae^{bx}的曲线拟合.

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

解：$y = ae^{bx} \Longleftrightarrow lny = lna + bx$. 令$z_i = lny_i$，可得离散数据$\{(x_i, z_i)\}_{i=1}^{8}$：

例 3：对以下数据作形如ae^{bx}的曲线拟合.

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

解：$y = ae^{bx} \Longleftrightarrow lny = lna + bx$. 令$z_i = lny_i$，可得离散数据$\{(x_i, z_i)\}_{i=1}^{8}$：

x_i	1	2	3	4	5	6	7	8
z_i	2.7279	3.0204	3.3105	3.6000	3.8939	4.1836	4.4751	4.7673

例 3：对以下数据作形如ae^{bx}的曲线拟合.

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

解：$y = ae^{bx} \Longleftrightarrow lny = lna + bx$. 令$z_i = lny_i$，可得离散数据$\{(x_i, z_i)\}_{i=1}^{8}$：

x_i	1	2	3	4	5	6	7	8
z_i	2.7279	3.0204	3.3105	3.6000	3.8939	4.1836	4.4751	4.7673

对数组$\{(x_i, z_i)\}_{i=1}^{8}$进行线性拟合$z = A + Bx$，可得法方程为：

$$\begin{pmatrix} 8 & 36 \\ 36 & 204 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 29.9787 \\ 147.1350 \end{pmatrix}$$

例 3：对以下数据作形如ae^{bx}的曲线拟合.

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

解：$y = ae^{bx} \Longleftrightarrow lny = lna + bx$. 令$z_i = lny_i$，可得离散数据$\{(x_i, z_i)\}_{i=1}^{8}$：

x_i	1	2	3	4	5	6	7	8
z_i	2.7279	3.0204	3.3105	3.6000	3.8939	4.1836	4.4751	4.7673

对数组$\{(x_i, z_i)\}_{i=1}^{8}$进行线性拟合$z = A + Bx$，可得法方程为：

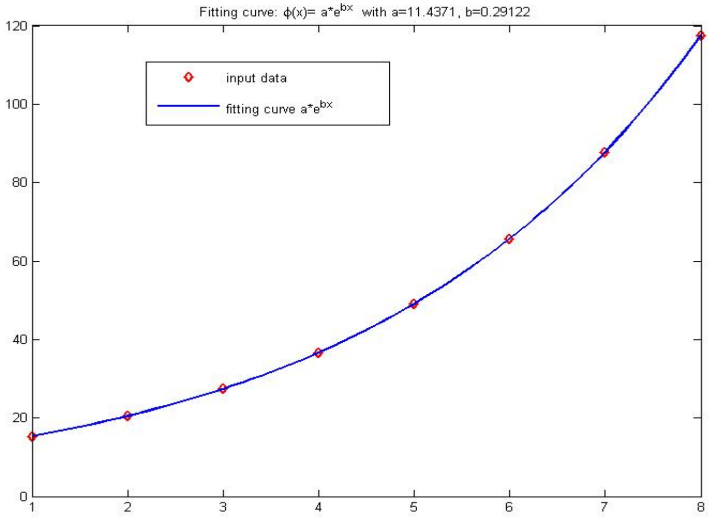
$$\begin{pmatrix} 8 & 36 \\ 36 & 204 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 29.9787 \\ 147.1350 \end{pmatrix}$$

$\Longrightarrow A \approx 2.4369,\ B \approx 0.29122$. 于是 $a = e^A \approx 11.4371,\ b = B \approx 0.29122$.
所求拟合曲线为

$$y = 11.4371 \times e^{0.29122x}$$

拟合结果

拟合结果



Fitting curve: ϕ(x)= a*e^bx with a=11.4371, b=0.29122
input data
fitting curve a*e^bx
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8