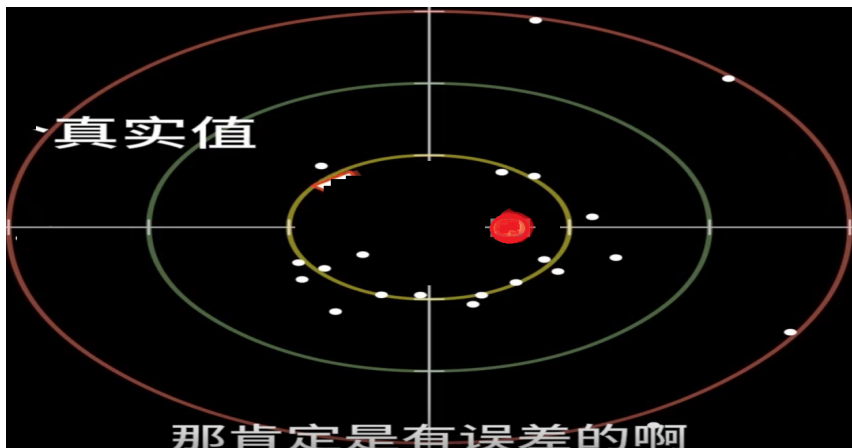


## 第二章 最小二乘拟合

中国科学技术大学 数学学院

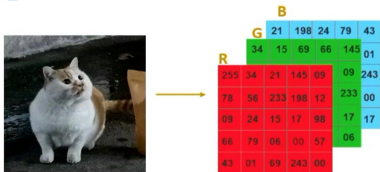
[chenxjin@ustc.edu.cn](mailto:chenxjin@ustc.edu.cn)



<https://www.toutiao.com/video/7283719078433260047/>

# 卷积神经网络

1.为什么使用卷积神经网络?

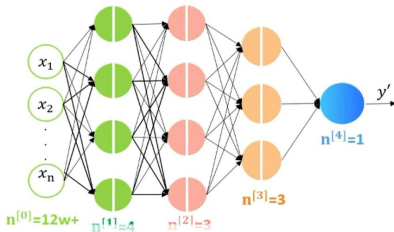


$$202 \times 206 \times 3 = 124836$$

W矩阵的维度:  $n^{[1]} * n^{[0]} = 4 * 12w$

$$x^{[1]} =$$

$$\begin{bmatrix} 255 \\ 34 \\ 21 \\ 145 \\ 09 \\ 78 \\ \vdots \\ 34 \\ 15 \\ 68 \\ \vdots \\ 21 \\ \vdots \end{bmatrix}$$

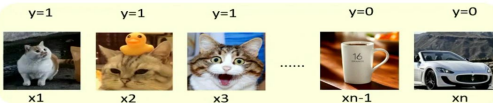


# 插值与拟合: 都是构造近似函数的方法

## 神经网络入门与实践

### 卷积神经网络

1. 为什么使用卷积神经网络?



$x_1 = \begin{bmatrix} 255 \\ 34 \\ 21 \\ 145 \\ 09 \\ 78 \\ \vdots \\ 34 \\ 15 \\ 69 \\ \vdots \\ 21 \\ \vdots \\ \vdots \end{bmatrix}$

$x_2 = \begin{bmatrix} 21 \\ 78 \\ 69 \\ 15 \\ 145 \\ 21 \\ \vdots \\ 255 \\ 09 \\ 34 \\ \vdots \\ 34 \\ \vdots \\ \vdots \end{bmatrix}$

$x_3 = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$

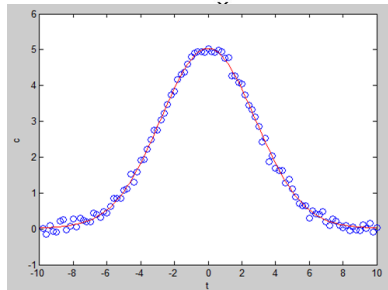
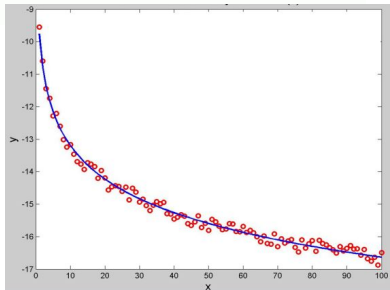
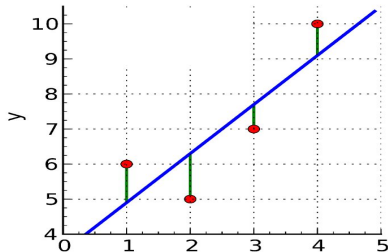
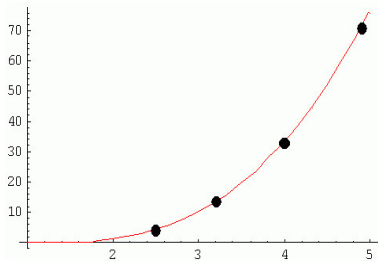
.....

$x_{n-1} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$

$x_n = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$

$X = \begin{bmatrix} 255 & 21 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 34 & 78 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 21 & 69 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 145 & 15 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 09 & 145 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 78 & 21 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 34 & 255 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 15 & 09 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 69 & 34 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 21 & 34 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$

# 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  $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+ \cup \{0\}$  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  $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}^+ \cup \{0\}$  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  $X = (x_1, x_2, \dots, x_n)$ . 几种常见的范数有：

①  $\|X\|_1 = \sum_{i=1}^n |x_i| \longrightarrow 1\text{-范数}$

②  $\|X\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \longrightarrow 2\text{-范数}$

③  $\|X\|_\infty = \max\{|x_i|\} \longrightarrow \infty\text{-范数}$

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

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

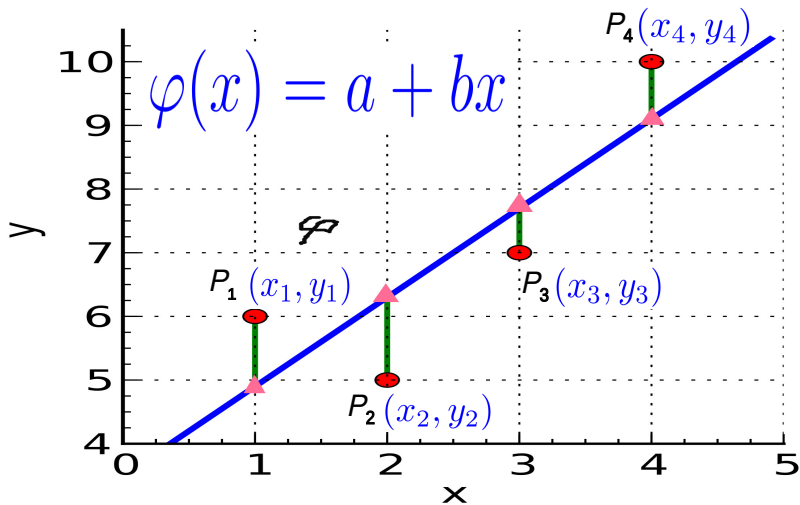
例如:

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$$

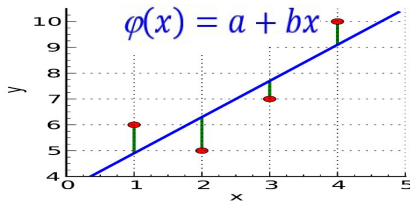
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$

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

## 拟合标准-“距离”最小/近



## 拟合标准-“距离”最近

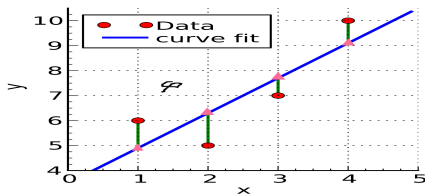


$(x_1, y_1)$

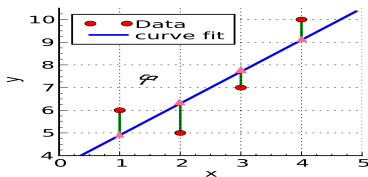
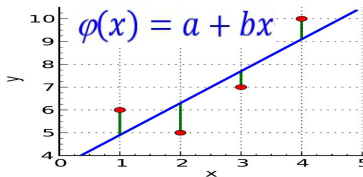
$(x_2, y_2)$

$\vdots$

$(x_m, y_m)$



## 拟合标准-“距离”最近



$$(x_1, y_1)$$

$$(x_2, y_2)$$

$$\vdots$$

$$(x_m, y_m)$$

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

拟合标准-“距离”最小

样本

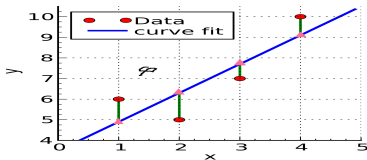
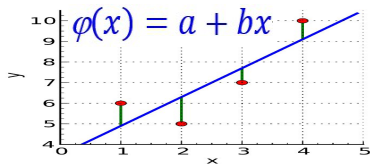
$(x_1, y_1)$   
 $(x_2, y_2)$   
 $\vdots$   
 $(x_m, y_m)$

$$\varphi(x) = ?$$

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



$(x_1, y_1)$  要求所做逼近函数  $\varphi(x)$  尽可能地靠近样本点,  
 $(x_2, y_2)$  即向量  $\Phi(\mathbf{x}) = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_m))^T$   
 $\vdots$  和  $\mathbf{Y} = (y_1, y_2, \dots, y_m)^T$  的误差或距离最小,  
$$\min \|\Phi(\mathbf{x}) - \mathbf{Y}\|.$$

$(x_m, y_m)$  ♣ 拟合标准  
向量  $\Phi$  与  $\mathbf{Y}$  之间的误差最小:  $\min \|\Phi(\mathbf{x}) - \mathbf{y}\|$   
向量  $\Phi$  与  $\mathbf{Y}$  之间的误差或距离有多种定义方式。

# 拟合标准

向量  $\Phi(\mathbf{x})$  与  $\mathbf{Y}$  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

$$\|\Phi(\mathbf{x}) - \mathbf{Y}\|_1 = \sum_{i=1}^m |\varphi(x_i) - y_i|,$$

$$\|\Phi(\mathbf{x}) - \mathbf{Y}\|_2 = \sqrt{\sum_{i=1}^m (\varphi(x_i) - y_i)^2},$$

$$\|\Phi(\mathbf{x}) - \mathbf{Y}\|_\infty = \max_{1 \leq i \leq n} |\varphi(x_i) - y_i|$$

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拟合方法称为**最小二乘法**。  
本章主要讲述用最小二乘法构造拟合函数  $\varphi(x)$ 。

已知

## 最小二乘问题

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,  $\Phi$  为给定的某一函数类。求  $\Phi$  上的函数  $\phi(x)$  满足

$$\min \left\| \begin{pmatrix} f(x_1), f(x_2), \dots, f(x_m) \end{pmatrix} - \begin{pmatrix} \phi(x_1), \phi(x_2), \dots, \phi(x_m) \end{pmatrix} \right\|$$

即距离达到最小.

## 最小二乘问题

已知

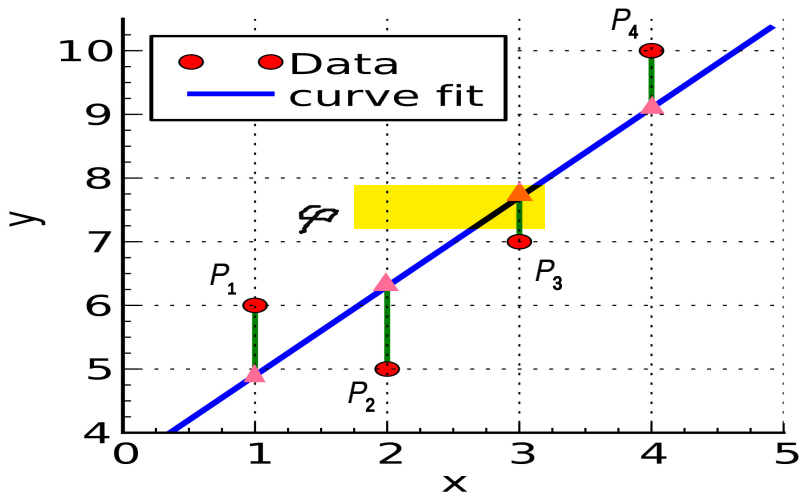
问题:  $f(x)$  为定义在区间  $[a, b]$  上的函数,  $\{x_i\}_{i=1}^m$  为区间上  $m$  个互不相同的点,  $\Phi$  为给定的某一函数类。求  $\Phi$  上的函数  $\phi(x)$  满足

$$\min \left\| \begin{pmatrix} f(x_1), f(x_2), \dots, f(x_m) \end{pmatrix} - \begin{pmatrix} \phi(x_1), \phi(x_2), \dots, \phi(x_m) \end{pmatrix} \right\|$$

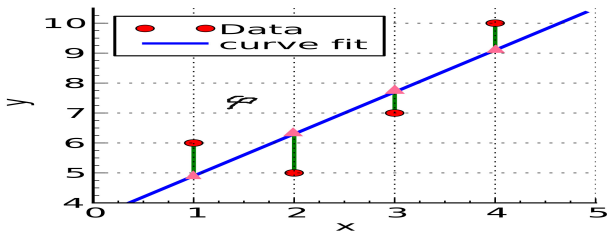
即距离达到最小.

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



## 线性拟合



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

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

# 线性拟合

得到满足最小均方误差的法方程组（正规方程）：

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

整理得到满足最小均方误差的法方程组（正规方程）：

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

写成矩阵形式：

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

用消元法或克莱姆方法解出方程组。

## 线性拟合的法方程组（正规方程）

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

## 线性拟合的法方程组（正规方程）

$$\begin{matrix} & 1 & x & & y \\ \begin{matrix} 1 \\ x \end{matrix} & \begin{pmatrix} \\ \end{pmatrix} & \begin{pmatrix} a \\ b \end{pmatrix} & = & \begin{pmatrix} \\ \end{pmatrix} \end{matrix}$$

# 线性拟合

例 2.1 下表是1950至1959年我国人口数据资料(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 |

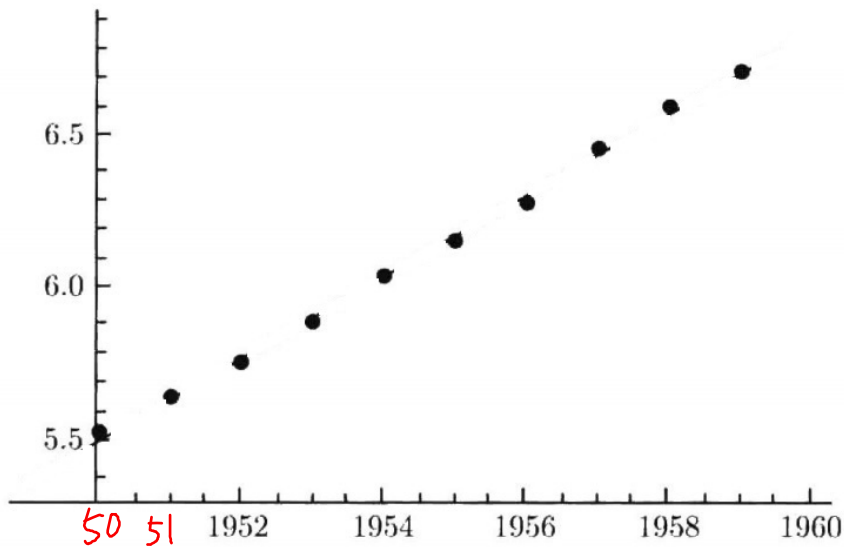


图 2.3 人口增长的线性模型

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.

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

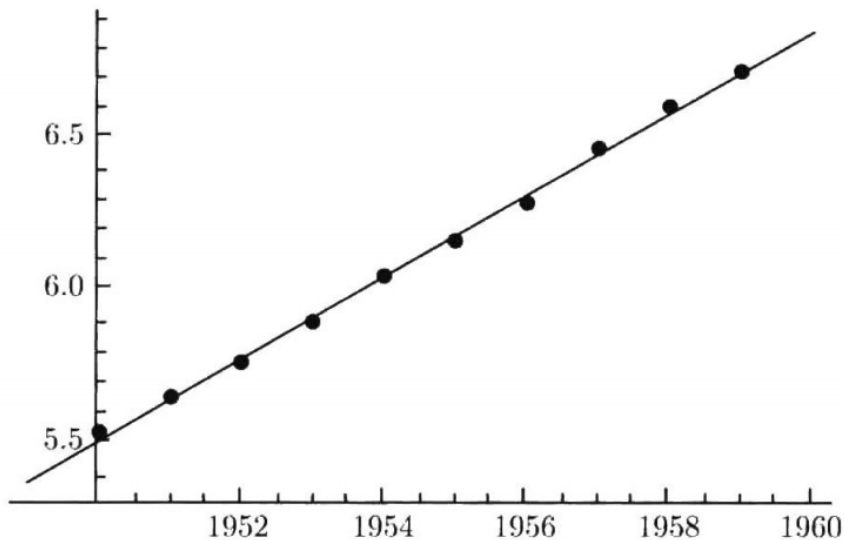


图 2.3 人口增长的线性模型

## 最小二乘问题之多项式拟合 $\iff$ 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

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

达到最小. 此时  $\phi$  称为多项式拟合函数.

# 最小二乘问题之多项式拟合 $\iff$ 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

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

达到最小. 此时  $\phi$  称为多项式拟合函数.

设  $\phi(x) = \sum_{j=0}^n a_j x^j$ , 则  $R = \sum_{i=1}^m (\phi(x_i) - y_i)^2$  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：

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

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

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  $\sum_{i=1}^m x_i^k$  ( $0 \leq k \leq 2n$ ),  $\sum_{i=1}^m x_i^k y_i$  ( $0 \leq k \leq n$ ), 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 $\Leftarrow$ 矛盾方程组(线代)

线性方程组  $Ax = b$  ( $A \in \mathbb{R}_{m \times n}$ ), 当  $\text{rank}(A) \neq \text{rank}(A, b)$  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 $\Leftarrow$ 矛盾方程组(线代)

线性方程组  $Ax = b$  ( $A \in \mathbb{R}_{m \times n}$ ), 当  $\text{rank}(A) \neq \text{rank}(A, b)$  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 $\Leftarrow$ 矛盾方程组(线代)

线性方程组  $Ax = b$  ( $A \in \mathbb{R}_{m \times n}$ ), 当  $\text{rank}(A) \neq \text{rank}(A, b)$  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  $m > n$ ,  $A \in \mathbb{R}_{m \times n}$ ,  $b \in \mathbb{R}_{m \times 1}$ ,  $\text{rank}(A) = n$ .  $A^T Ax = A^T b$  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  $\iff A^T Ax = A^T b$ , 即  $x$  为法方程的解.

1. 法方程 $A^T Ax = A^T b$ 恒有唯一解.

1. 法方程  $A^T A x = A^T b$  恒有唯一解.

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

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

1. 法方程  $A^T A x = A^T b$  恒有唯一解.

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

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

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

1. 法方程  $A^T A x = A^T b$  恒有唯一解.

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

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

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

由  $\bar{A}_n$  可逆, 故

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

即  $A^T A$  可逆.  $\implies A^T A x = A^T b$  有唯一解.

2.  $x$ 为 $\min \|Ax - b\|_2$ 的解  $\iff x$ 为法方程 $A^T Ax = A^T b$ 的解.

2.  $x$ 为 $\min \|Ax - b\|_2$ 的解  $\iff x$ 为法方程 $A^T Ax = A^T b$ 的解.

Pf: “ $\Leftarrow$ ”:  $x$ 满足法方程 $A^T Ax = A^T b$ .  $\forall y \in \mathbb{R}_{n \times 1}$ , 设 $e = y - x$ . 则

$$\begin{aligned}\|Ay - b\|_2^2 &= (Ay - b)^T (Ay - b) = (Ax + Ae - b)^T (Ax + Ae - b) \\&= (Ax - b)^T (Ax - b) + (Ax - b)^T Ae + (Ae)^T (Ax - b) + (Ae)^T (Ae) \\&= \|Ax - b\|_2^2 + \|Ae\|_2^2 + x^T A^T Ae - b^T Ae + e^T A^T Ax - e^T A^T b \\&= \|Ax - b\|_2^2 + \|Ae\|_2^2 + 0 \\&\geq \|Ax - b\|_2^2\end{aligned}$$

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

即 $x$ 为 $\min \|Ax - b\|_2$ 的解.

2.  $x$ 为 $\min \|Ax - b\|_2$ 的解  $\iff x$ 为法方程 $A^T Ax = A^T b$ 的解.

Pf: “ $\Leftarrow$ ”:  $x$ 满足法方程 $A^T Ax = A^T b$ .  $\forall y \in \mathbb{R}_{n \times 1}$ , 设 $e = y - x$ . 则

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

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

即 $x$ 为 $\min \|Ax - b\|_2$ 的解.

“ $\Rightarrow$ ”: 设 $x = (x_1, x_2, \dots, x_n)$ , 由极小值条件 $\frac{\partial(\|Ax - b\|_2^2)}{\partial x_i} = 0 \implies A^T Ax = A^T b$ .

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

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

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

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

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

# 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 $\phi$ ；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 $\Phi$ 不是 $\mathbb{P}_n$ ，但 $\Phi = \text{Span}\{\phi_0, \dots, \phi_n\}$ 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.,  
 $\Phi = \text{Span}\{1, x^3\}$ .
- 4 若给定的函数空间 $\Phi$ 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

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

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

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

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

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

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

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

其法方程为

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

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

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

其法方程为

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

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

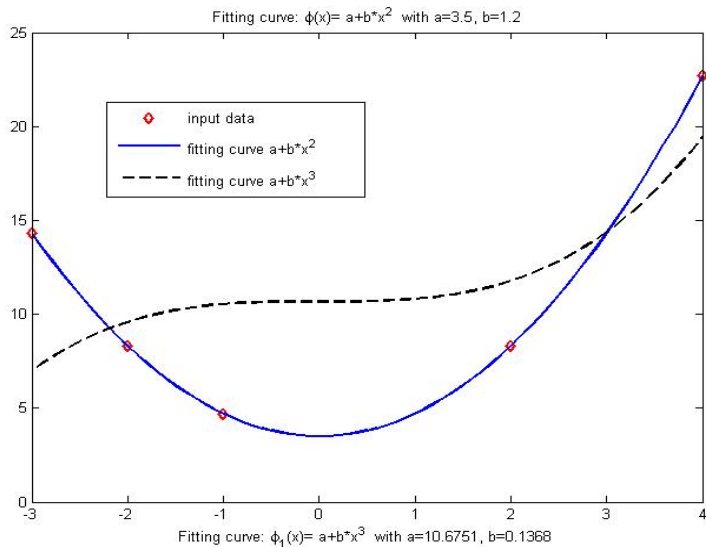
$$\begin{cases} a + b(-3)^2 = 14.3 \\ a + b(-2)^2 = 8.3 \\ a + b(-1)^2 = 4.7 \\ a + b \times 2^2 = 8.3 \\ a + b \times 4^2 = 22.7 \end{cases}, \text{ 即 } \begin{pmatrix} 1 & (-3)^2 \\ 1 & (-2)^2 \\ 1 & (-1)^2 \\ 1 & 2^2 \\ 1 & 4^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 14.3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

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



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.

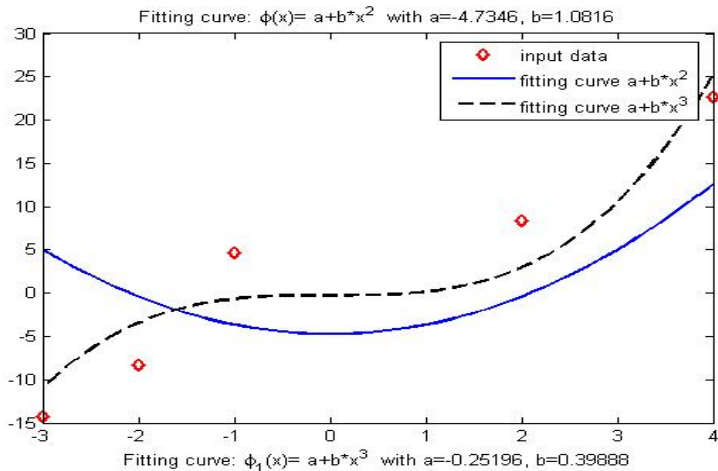
|       |       |      |     |     |      |
|-------|-------|------|-----|-----|------|
| $x_i$ | -3    | -2   | -1  | 2   | 4    |
| $y_i$ | -14.3 | -8.3 | 4.7 | 8.3 | 22.7 |

求解过程与例1-2完全类似;

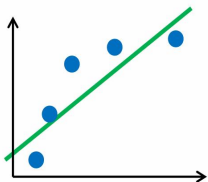
例 2': 分别对以下数据作形如 $a + bx^2$ 以及 $a + bx^3$ 的曲线拟合.

|       |       |      |     |     |      |
|-------|-------|------|-----|-----|------|
| $x_i$ | -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：

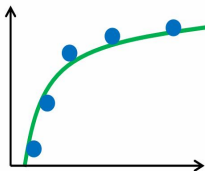


# 欠拟合或过拟合



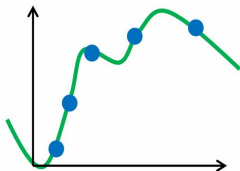
$$w_0 + w_1x$$

High bias  
(underfitting)



$$w_0 + w_1x + w_2x^2$$

"Just right"



$$w_0 + w_1x + w_2x^2 + w_3x^3 +$$

High variance  
(overfitting)

- 需要根据不同的应用与需求，不断尝试  
(不断“调参”)

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:  $y = ae^{bx} \iff \ln y = \ln a + bx$ . 令 $z_i = \ln y_i$ , 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:  $y = ae^{bx} \iff \ln y = \ln a + bx$ . 令  $z_i = \ln y_i$ , 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:  $y = ae^{bx} \iff \ln y = \ln a + bx$ . 令 $z_i = \ln y_i$ , 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.

新形式  $\rightarrow$  齐次  $\rightarrow$

|       |      |      |      |      |      |      |      |       |
|-------|------|------|------|------|------|------|------|-------|
| $x_i$ | 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 |

法方程

解:  $y = ae^{bx} \iff \ln y = \ln a + bx$ . 令  $z_i = \ln y_i$ , 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:

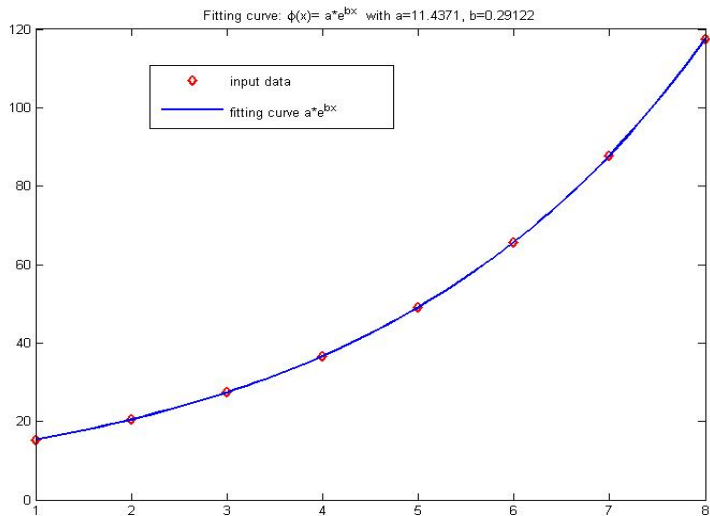
$$\begin{pmatrix} 8 & 36 \\ 36 & 204 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 29.9787 \\ 147.1350 \end{pmatrix}$$

$\implies A \approx 2.4369, B \approx 0.29122$ . 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



# 多元线性回归问题

# 多元线性回归问题(见9章)

前面主要讨论了函数  $y$  依赖于**单个变量**的情形. 如果  $y$  是多个自变量  $x_1, x_2, \dots, x_m$  的函数, 那么可以把这些自变量本身视为一组基, 建立如下模型:

$$y(x_1, x_2, \dots, x_m) = c_1 x_1 + c_2 x_2 + \dots + c_m x_m,$$

来预测  $y$ , 其中系数  $\{c_j\}_{j=1}^m$  可利用**最小二乘法**求解, 即

$$\min_{c_1, c_2, \dots, c_m \in \mathbb{R}} \sum_{i=1}^n \left[ \sum_{j=1}^m c_j x_{ij} - y_i \right]^2.$$

在数理统计中, 上述方法又称为**多元线性回归**.

一般地, 当  $n \gg m$  时, 若线性方程组  $A\mathbf{c} = \mathbf{f}$  无解, 则称它为**矛盾方程组**. 对于矛盾方程组, 称取到

$$\min_{\mathbf{c} \in \mathbb{R}^m} \|A\mathbf{c} - \mathbf{f}\|_2^2 \quad (9.11)$$

的  $\mathbf{c}$  为**矛盾方程的最小二乘解**. 容易看出, 矛盾方程的最小二乘解是内积空间  $\mathbb{R}^n$  中的最佳平方逼近问题, 即在  $A$  的列空间中选取元素最佳逼近  $\mathbb{R}^n$  中的元素  $\mathbf{f}$ . 此外, 数据拟合的最小二乘问题 (9.8) 可以视为矛盾方程组的最小二乘解. 从前面的讨论可知, 矛盾方程的最小二乘解可通**法方程组**求解, 即

$$G\mathbf{c} = \mathbf{b} \iff A^T A\mathbf{c} = A^T \mathbf{f}.$$

关于矛盾方程的最小二乘解有下面的结论.

**定理 9.11** 设  $A \in \mathbb{R}^{n \times m}$ ,  $\mathbf{f} \in \mathbb{R}^n$ , 则

- (1) 线性方程组  $A^T A \mathbf{x} = A^T \mathbf{f}$  恒有解;
- (2) 函数  $\|A \mathbf{x} - \mathbf{f}\|_2$  取到最小值  $\iff \mathbf{x}$  满足  $A^T A \mathbf{x} = A^T \mathbf{f}$ ;
- (3) 线性方程组  $A^T A \mathbf{x} = A^T \mathbf{f}$  的解是唯一的  $\iff \text{rank } A = n$ .

**例 9.8** 为了开拓市场, 某公司对其新产品作了一系列调查, 他们发现这一新产品的销量与下列事件关系密切: 其一是**温度**, 其二是**上证指数**, 其三是**广告费**, 其四是**推销员数**, 其五是**返修率**, 详细数据见下表.

| 记录 | 温度 | 上证指数 | 广告费   | 推销员数 | 返修率  | 销售量 |
|----|----|------|-------|------|------|-----|
| 1  | 39 | 567  | 10000 | 2    | 0.20 | 75  |
| 2  | 37 | 679  | 0     | 3    | 0.15 | 68  |
| 3  | 30 | 346  | 5000  | 3    | 0.10 | 105 |
| 4  | 25 | 987  | 5000  | 3    | 0.08 | 136 |
| 5  | 25 | 1101 | 0     | 4    | 0.07 | 152 |
| 6  | 10 | 1004 | 5000  | 5    | 0.07 | 191 |
| 7  | 15 | 667  | 0     | 4    | 0.05 | 148 |
| 8  | 5  | 604  | 10000 | 6    | 0.04 | 234 |

假设这些事件与销量量近似**成线性关系**, 试求这种关系的数学表达式.

解 设销售量为  $y$ , 用  $x_1, x_2, x_3, x_4, x_5$  分别表示温度、上证指数、广告费、推销员数、返修率, 则需确定函数

$$y = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5.$$

利用最小二乘法, 系数  $c_1, c_2, \dots, c_5$  需要满足**法方程组**:

$$\begin{pmatrix} 5390 & 132881 & 765000 & 594 & 21.75 \\ 132881 & 4906337 & 23395000 & 22886 & 533.67 \\ 765000 & 23395000 & 275000000 & 135000 & 3650 \\ 594 & 22886 & 135000 & 124 & 2.46 \\ 21.75 & 533.67 & 3650 & 2.46 & 0.0928 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 21091 \\ 858427 \\ 5250000 \\ 4636 \\ 87.35 \end{pmatrix}$$

解得:  $y \approx 0.68772x_1 + 0.043703x_2 + 0.0048631x_3 + 29.607x_4 - 447.36x_5$ .