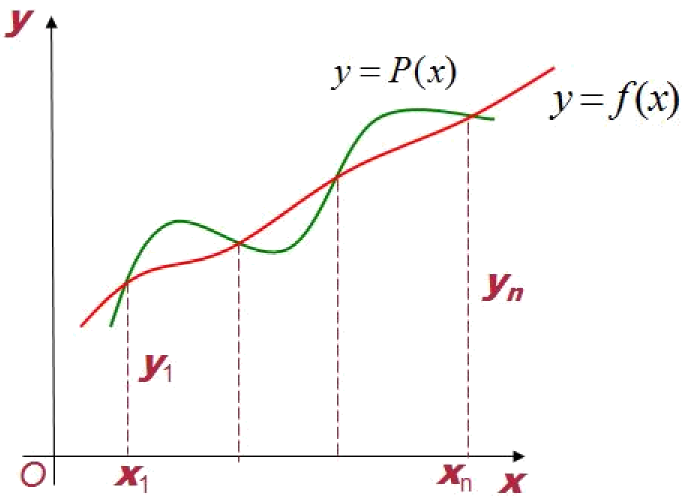
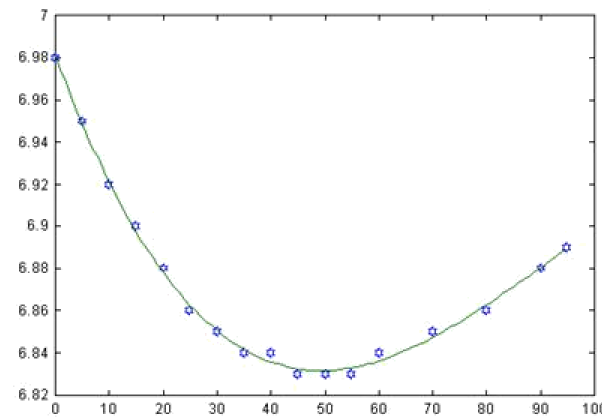
数据拟合

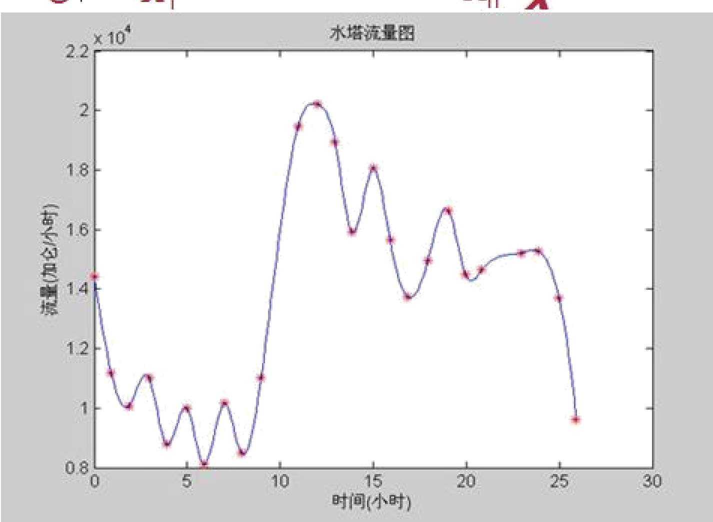
最小二乘

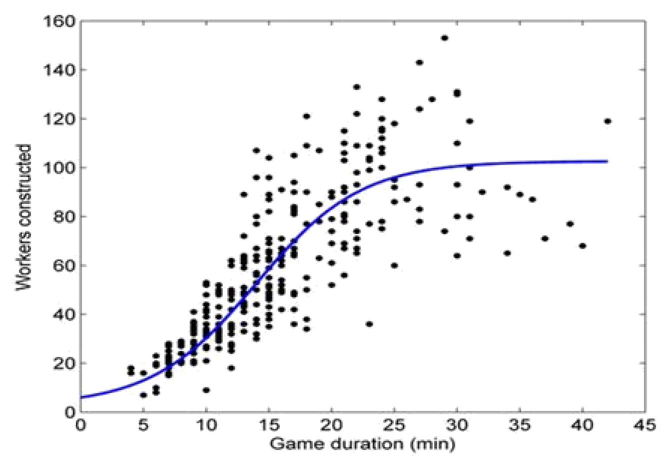
拟合

- 拟合 vs 插值







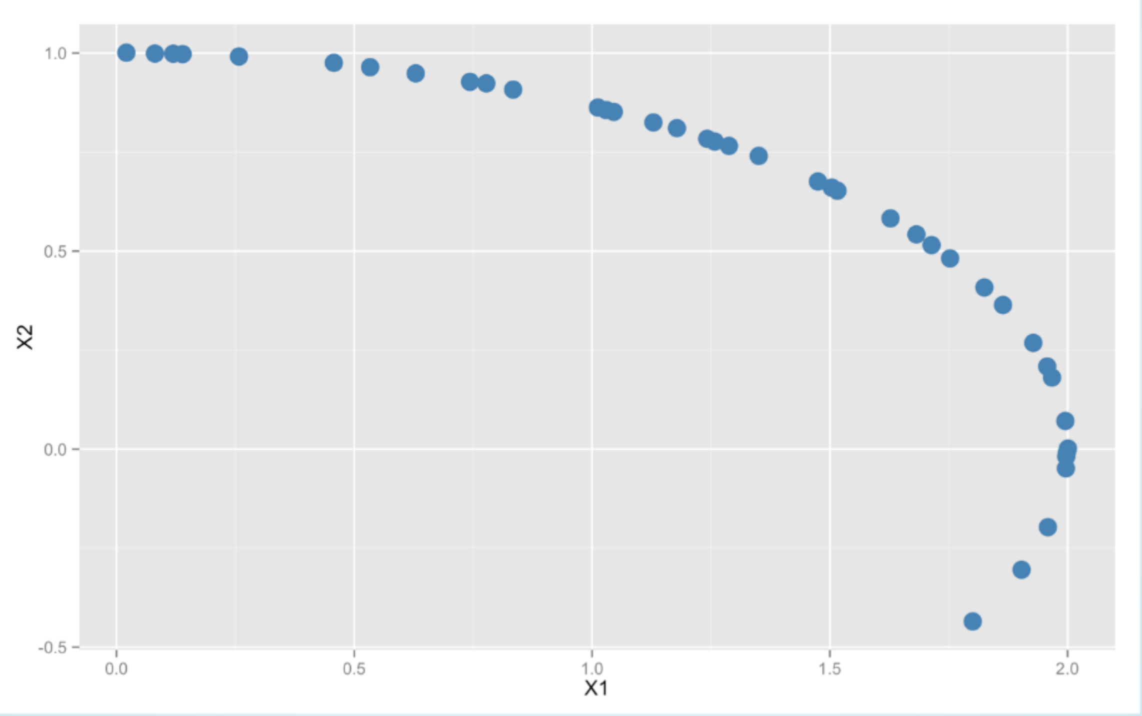


从离散观测点中找数据规律

拟合

拟合

例 1. 已知行星的运行轨道是一个椭圆。几百个观测数据，这些数据会带有误差，因此它们几乎不可能满足同一个椭圆方程。如何计算行星的运行轨道?



拟合

拟合的定义

定义 1.
$f(x)$是定义在区间$[a,b]$上的函数，x_i， $i=0,\cdots,n$是$[a,b]$区间上互不相同的点。Φ是已知的函数空间。 在Φ上找函数$g(x)$，使得$g(x)-f(x)$在某种范数意义下最小。函数$g(x)$称为$f(x)$的*拟合函数*。

注. 拟合同插值类似，都是构造近似函数的手段。

注. 范数可以用来度量线性空间中两个向量之间的距离，当然也可以度量函数空间中两个函数是否靠近。

拟合

最小二乘

定义 2.
在拟合问题中，取$f(x)-g(x)$的范数为2-范数的话，得到的问题就叫作*最小二乘问题*(Least Square)。即，找$g(x)$，使得

$$\|f(x)-g(x)\|_2=\sqrt{\sum_{i=0}^{n}(f(x_i)-g(x_i))^2}$$

达到最小。

拟合

最小二乘

$g(x)$在空间Φ中，则令

$$g(x) = a_0\phi_0(x) + a_1\phi_1(x) + \cdots + a_m\phi_m(x)$$

其中$\Phi = span\{\phi_0(x), \phi_1(x), \cdots, \phi_m(x)\}$。

问题变为：找$a_0, a_1, \cdots, a_m$使得

$$G = \sum_{i=0}^{n}(f(x_i) - [a_0\phi_0(x_i) + a_1\phi_1(x_i) + \cdots + a_m\phi_m(x_i)])^2$$

达到最小。$G = G(a_0, a_1, \cdots, a_m)$是一个多元函数。

多元函数的最小值点必须满足

$$\frac{\partial G}{\partial a_j} = 0, \forall j = 0, 1, \cdots, m$$

拟合

最小二乘

由G的表达式，

$$\frac{\partial G}{\partial a_j} = \sum_{i=0}^{n} \left(-2\left(f(x_i) - [a_0\phi_0(x_i) + a_1\phi_1(x_i) + \cdots + a_m\phi_m(x_i)]\right)\phi_j(x_i)\right)$$

$$\frac{\partial^2 G}{\partial a_j \partial a_k} = \sum_{i=0}^{n} 2\phi_j(x_i)\phi_k(x_i)$$

这样，G达到最小时，系数要满足

$$\sum_{i=0}^{n}[a_0\phi_0(x_i) + a_1\phi_1(x_i) + \cdots + a_m\phi_m(x_i)]\phi_j(x_i) = \sum_{i=0}^{n} f(x_i)\phi_j(x_i)$$

即有

拟合

最小二乘

$$
\begin{aligned}
a_0 \sum_{i=0}^{n} \phi_0(x_i)\phi_j(x_i) + a_1 \sum_{i=0}^{n} \phi_1(x_i)\phi_j(x_i) + \cdots \\
+ a_m \sum_{i=0}^{n} \phi_m(x_i)\phi_j(x_i) = \sum_{i=0}^{n} f(x_i)\phi_j(x_i)
\end{aligned}
$$

定义

$$
(\phi, \psi) = \sum_{i=0}^{n} \phi(x_i)\psi(x_i)
$$

则可以简写上式为

$$
a_0(\phi_0, \phi_j) + a_1(\phi_1, \phi_j) + \cdots + a_m(\phi_m, \phi_j) = (f, \phi_j)
$$

$$
j = 0, 1, \cdots, m
$$

拟合

最小二乘

这样，可以得到如下的线性方程组

$$\begin{pmatrix} (\phi_0,\phi_0) & (\phi_1,\phi_0) & \cdots & (\phi_m,\phi_0) \\ (\phi_0,\phi_1) & (\phi_1,\phi_1) & \cdots & (\phi_m,\phi_1) \\ \vdots & \vdots & \cdots & \vdots \\ (\phi_0,\phi_m) & (\phi_1,\phi_m) & \cdots & (\phi_m,\phi_m) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} (f,\phi_0) \\ (f,\phi_1) \\ \vdots \\ (f,\phi_m) \end{pmatrix}$$

称为*法方程*。

拟合

最小二乘

例 3. 有数据

x	-4	-2	1	2	4
y	14.2	2.3	1.4	2.0	4.3

试对数据分别做如下的拟合：

(1).$y = a + bx$, (2).$y = ax^2 + b$, (3).$y = ax^2 + bx + c$,

(4).$y = ae^{bx}$, (5).$y = ax^b$, (6).$y = \frac{1}{a+bx}$

解. (1) 可以确定基函数为$1, x$，则可以列出法方程

$$\begin{pmatrix} (1,1) & (1,x) \\ (x,1) & (x,x) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} (f,1) \\ (f,x) \end{pmatrix}$$

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

拟合

最小二乘

```
========= a+bx ===========
Coefficients:  [ 5.05392157 -1.06960784]
Matrix:  [[  5.    1.]
 [  1.  41.]]
RHS: [ 24.2 -38.8]
```

(2) 基函数为$x^2, 1$，则有

$$\begin{pmatrix} \sum x_i^4 & \sum x_i^2 \\ \sum x_i^2 & \sum 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i x_i^2 \\ \sum y_i \end{pmatrix}$$

```
========= ax^2+b ===========
Coefficients:  [ 0.55632184  0.27816092]
Matrix:  [[ 545.    41.]
 [  41.     5.]]
RHS: [ 314.6    24.2]
```

(3) 基函数为$x^2, x, 1$，则有

拟合

最小二乘

(4) 依据最小二乘问题的定义，问题变为找a, b使得

$$\sum_i (y_i - ae^{bx_i})^2$$

达到最小。这看起来**比较困难**。
将函数两边取对数，有

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

- 先对数据$(x_i, \ln y_i)$做基函数为$1, x$的拟合，得到 $\tilde{y} = \tilde{a} + \tilde{b}x$,
- 然后，取

$$y = e^{\tilde{y}} = e^{\tilde{a}} e^{\tilde{b}x}$$

注. 这样得到的系数并不能满足最小二乘的定义，但它确实满足某种范数意义下的最小。

拟合

最小二乘

法方程为

$$\begin{pmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \tilde{a} \\ \tilde{b} \end{pmatrix} = \begin{pmatrix} \sum \tilde{y}_i \\ \sum \tilde{y}_i x_i \end{pmatrix}$$

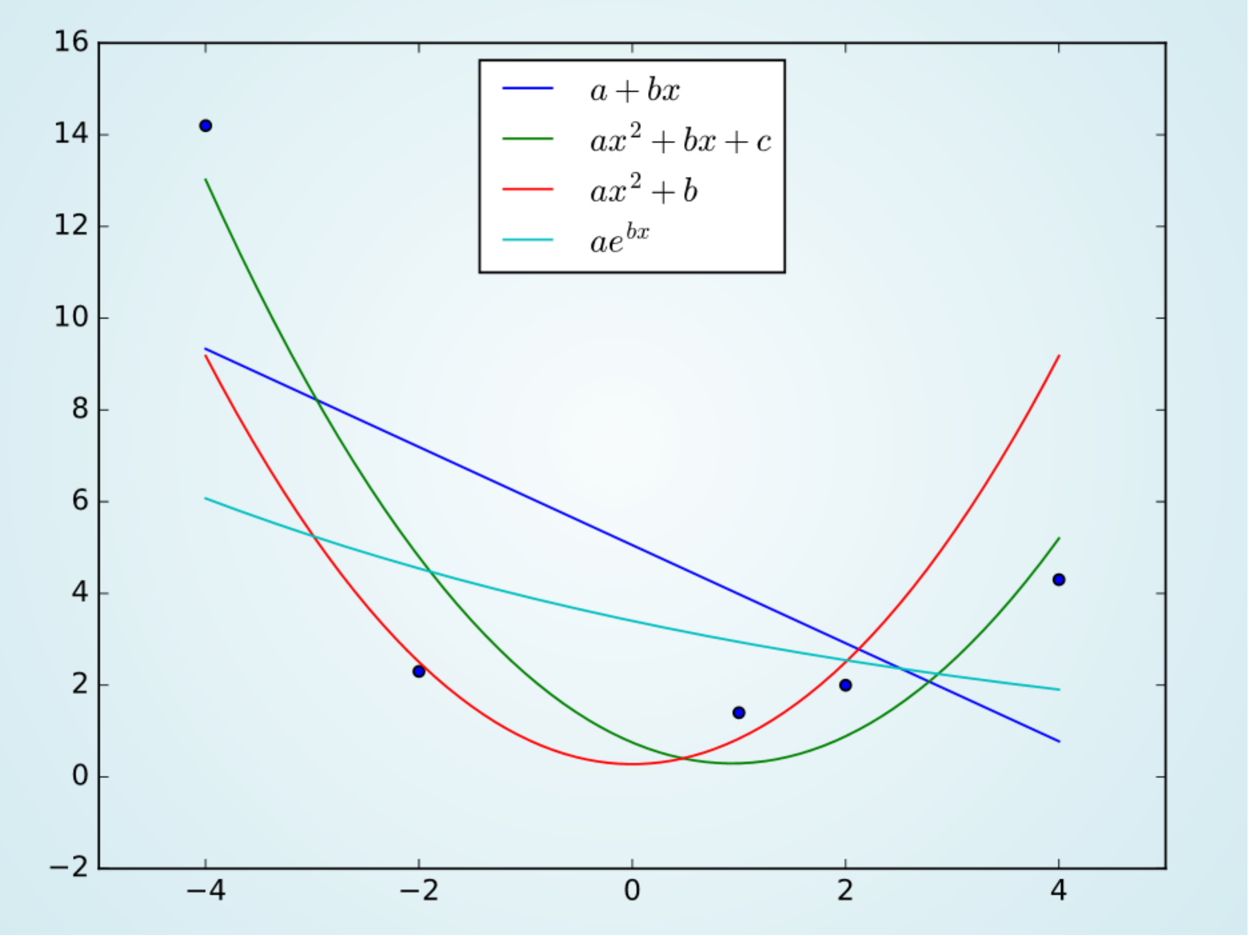
其中$\tilde{y}_i = \ln y_i$。

```
========= a e^{bx} ===========
Coefficients:  [ 1.22387925 -0.1450107 ]
Matrix:  [[  5.    1.]
 [  1.   41.]]
RHS: [ 5.97438553 -4.72155942]
```

注. 代码由python3编写，调用 `numpy.linalg.solve` 来解 法方程

拟合

- 最小二乘



拟合

矛盾方程组

矛盾方程组

问题. 行星的运行轨道是一个椭圆。现在有成百上千的观测数据，如何确定这个行星的运行轨道?

解. 这些观测数据应该满足同一个椭圆方程

$$ax^2 + bxy + cy^2 + dx + ey = 1$$

即有

$$ax_i^2 + bx_iy_i + cy_i^2 + dx_i + ey_i = 1, i = 1, 2, 3, \cdots$$

这是一个5个未知量，很多个方程的线性方程组。方程的个数比未知量的个数要多，这是一个**矛盾方程组**。

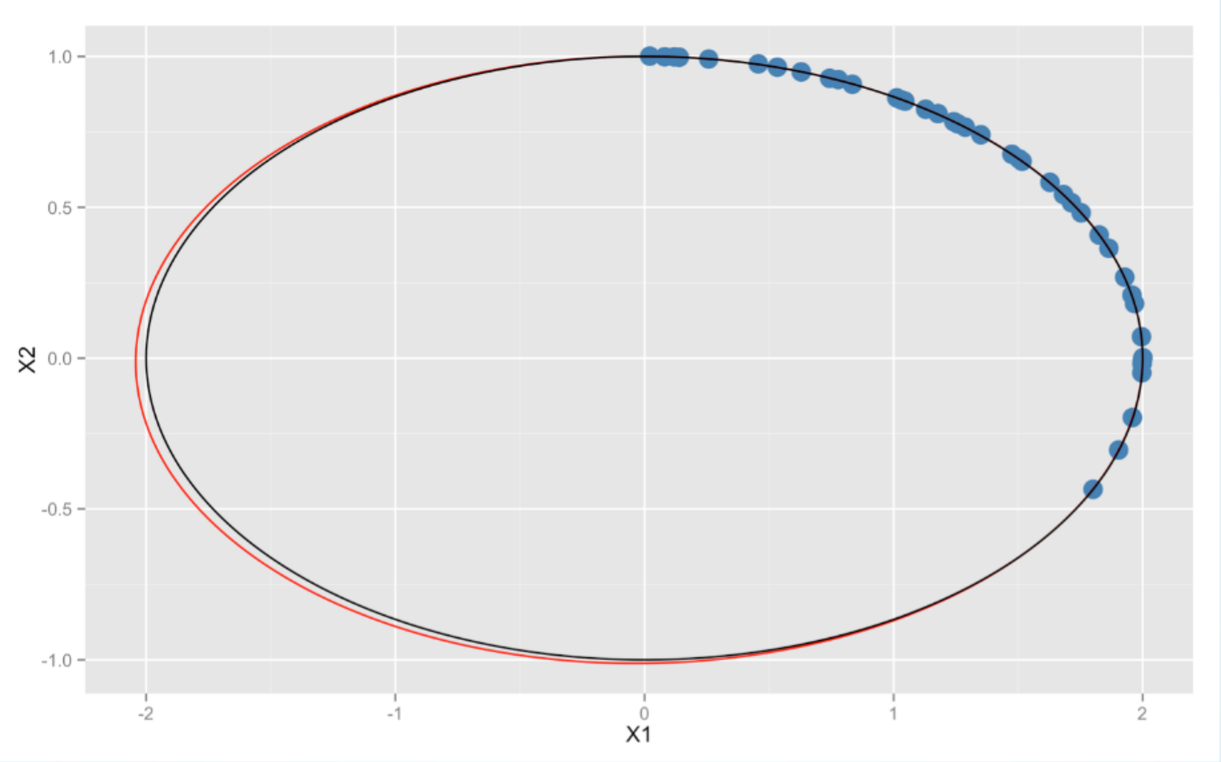
拟合

矛盾方程组

- 1801年，意大利天文学家**朱赛普·皮亚齐**发现了第一颗小行星谷神星。
- 经过40天的跟踪观测后，由于谷神星运行至太阳背后，使得皮亚齐失去了谷神星的位置。
- 随后全世界的科学家利用皮亚齐的观测数据开始寻找谷神星，但是根据大多数人计算的结果来寻找谷神星都没有结果。
- 时年24岁的**高斯**也计算了谷神星的轨道。奥地利天文学家**海因里希·奥尔伯斯**根据高斯计算出来的轨道重新发现了谷神星
- 高斯使用的最小二乘法的方法发表于1809年他的著作《天体运动论》。

拟合

矛盾方程组



黑线是真正的小行星轨道，红线是线性最小二乘法估计出来的轨道。

拟合

矛盾方程组

定理 1.
设线性方程组$Ax=b$的系数矩阵$A=(a_{ij})_{m\times n}$，则

1. $rank(A^TA)=rank(A^T)=rank(A)$
2. 线性方程组$A^TAx=A^Tb$恒有解。

若进步有$rank(A)=n$，则

1. 矩阵A^TA是对称正定矩阵;
2. n阶线性方程组$A^TAx=A^Tb$有唯一的解。

定理 2.
x满足$\min\limits_{y\in\mathbb{R}^n}\|Ay-b\|_2^2$的充要条件是$x$满足$A^TAx=A^Tb$。

拟合

矛盾方程组

如前例中，对于5个数据点，用**解矛盾方程组的方法**来拟合 $y = ax^2 + b$。如前数据，得到矛盾方程组

$$\begin{cases} ax_0^2 + b = y_0 \\ ax_1^2 + b = y_1 \\ ax_2^2 + b = y_2 \\ ax_3^2 + b = y_3 \\ ax_4^2 + b = y_4 \end{cases} \Rightarrow \begin{pmatrix} x_0^2 & 1 \\ x_1^2 & 1 \\ x_2^2 & 1 \\ x_3^2 & 1 \\ x_4^2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

拟合

矛盾方程组

求解最小二乘问题

$$\begin{pmatrix} x_0^2 & x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \left[\begin{pmatrix} x_0^2 & 1 \\ x_1^2 & 1 \\ x_2^2 & 1 \\ x_3^2 & 1 \\ x_4^2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \right]$$

可以得到

$$\begin{pmatrix} \sum x_i^4 & \sum x_i^2 \\ \sum x_i^2 & \sum 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum x_i^2 y_i \\ \sum y_i \end{pmatrix}$$

与前面得到的法方程是一样的

拟合

矛盾方程组

在拟合问题中，用m次多项式来拟合$n+1$个数据点时：

1. $m<n$时，可以得到矛盾方程组的矩阵为

$$A=\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^m \\ 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix}$$

根据Vandermonde行列式的特性知，$rank(A)=m$。此时，矛盾方程组的解存在唯一。

拟合

矛盾方程组

2. $m=n$时，可以得到唯一的一条插值多项式$L_n(x)$。由

$$\|f(x)-L_n(x)\|_2^2=\sum_{i=0}^{n}(f(x_i)-L_n(x_i))^2=0$$

知，$L_n(x)$是$f(x)$的拟合函数。

3. $m>n$时，通过$n+1$个点的m次多项式有无穷多条，这无穷条插值多项式都是拟合函数。此时，拟合函数存在不唯一。

拟合

- **作业：**

  无

群名称：数值计算方法
群　号：1132838842

矩阵特征值

THE END

谢谢张瑞老师的PPT