第1章　插值方法

1、插值概念
2、拉格朗日插值
3、牛顿插值
4、埃尔米特插值
5、分段插值
6、样条插值
7、曲线拟合的最小二乘法

1、问题引入

在工程实践和科学实验中，我们经常需要对实际问题建立函数关系，即$y=f(x)$。但是不知道$f(x)$的真实表达式，只得到一系列离散的观测数据（含误差）：

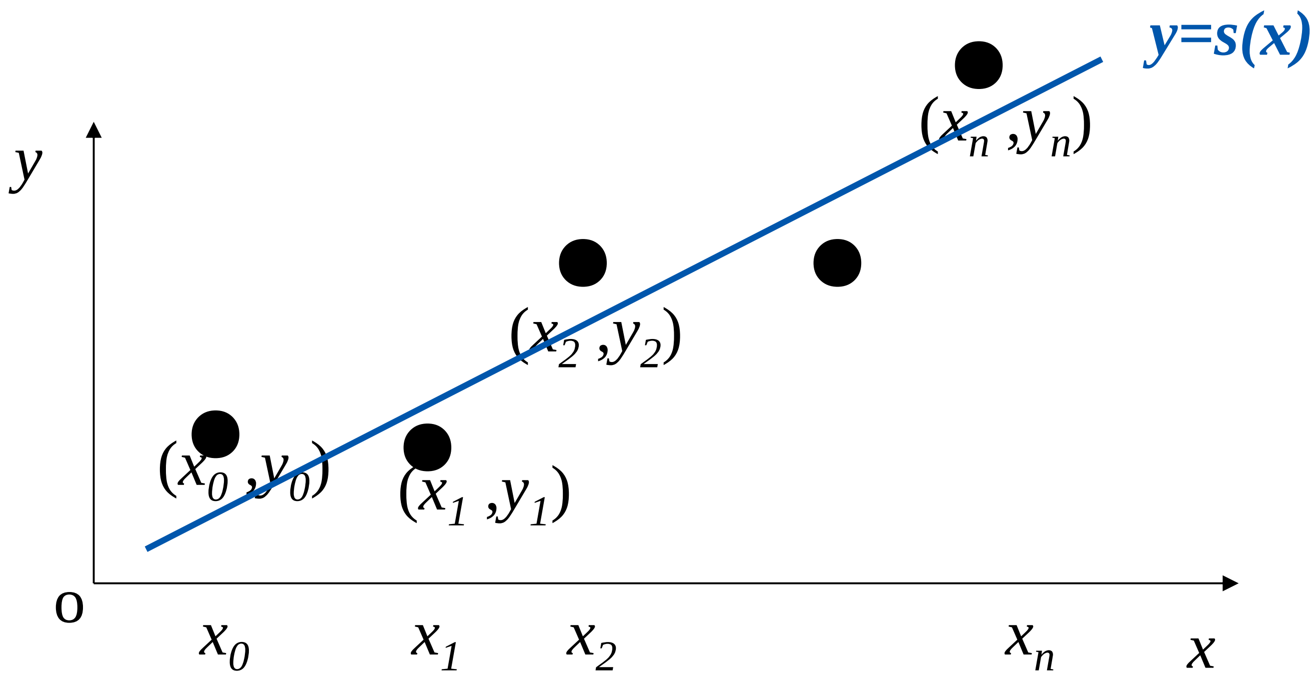
x_i	x_1	x_2	…	x_n
$f(x_i)$	$f(x_1)$	$f(x_2)$	…	$f(x_n)$

◆插值带来较大误差；

◆想能用一个经验函数（规律性）$y=s(x)$对真实函数$y=f(x)$作近似，从而得到在其它离散点处的函数值。

拟合的基本思想

求一个经验函数$s(x)$，

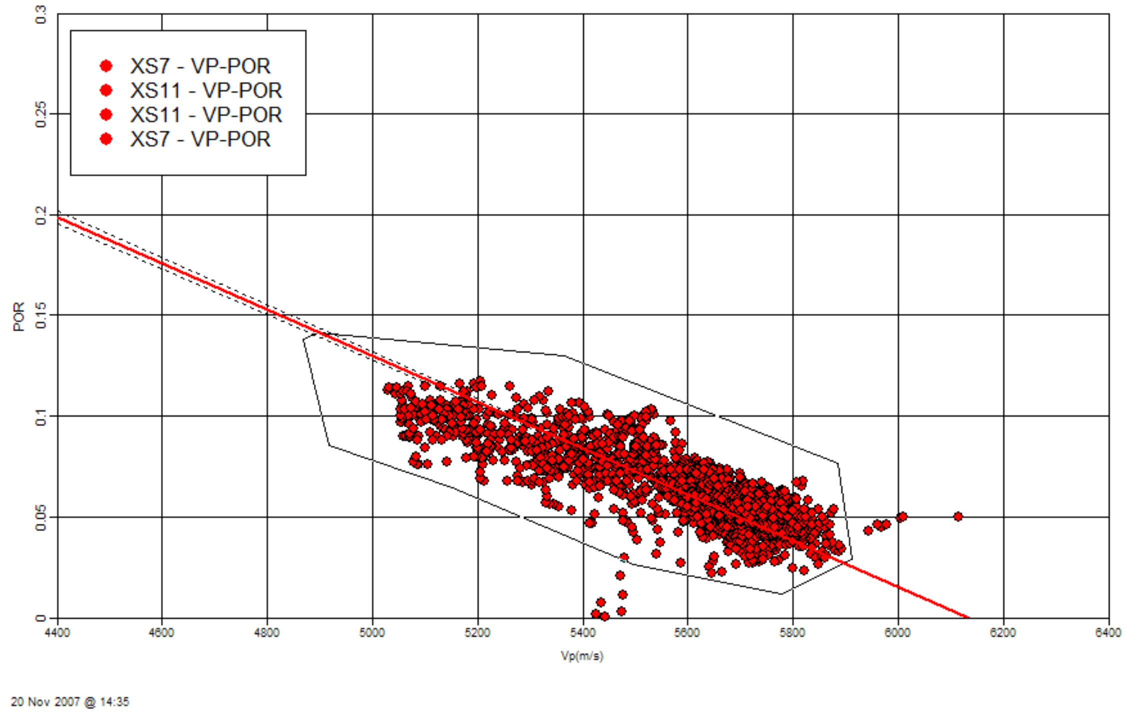


曲线拟合也是函数逼近问题。

岩石物理应用

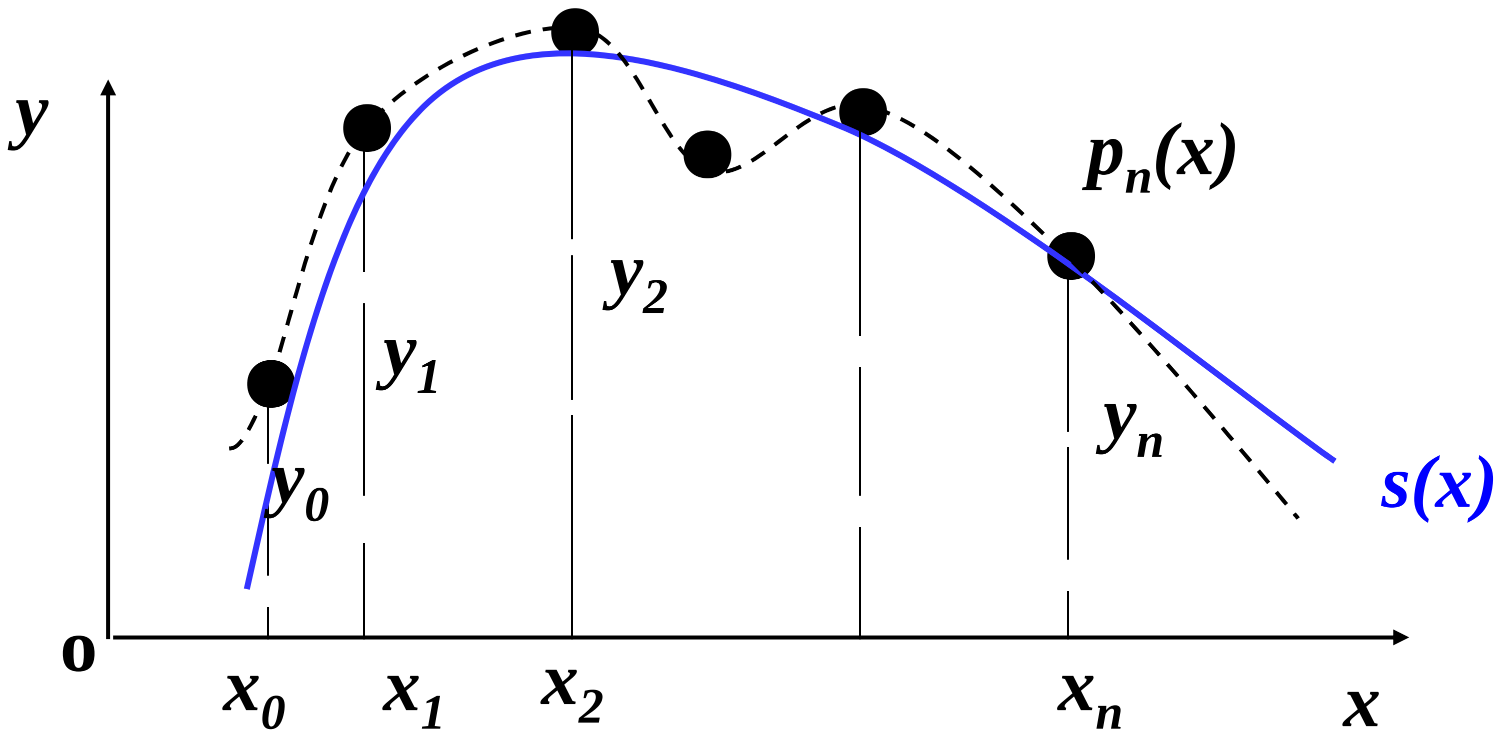
一般根据岩石测量参数建立岩石物理模型，钻孔提供孔隙度数据可以与岩石储层的声波速度建立关系，如下图所示，其关系为：

$$POR = (6130 - V_P) / 10790$$



火山岩储层声波速度与孔隙度的关系

拟合与插值的区别



1、直线拟合

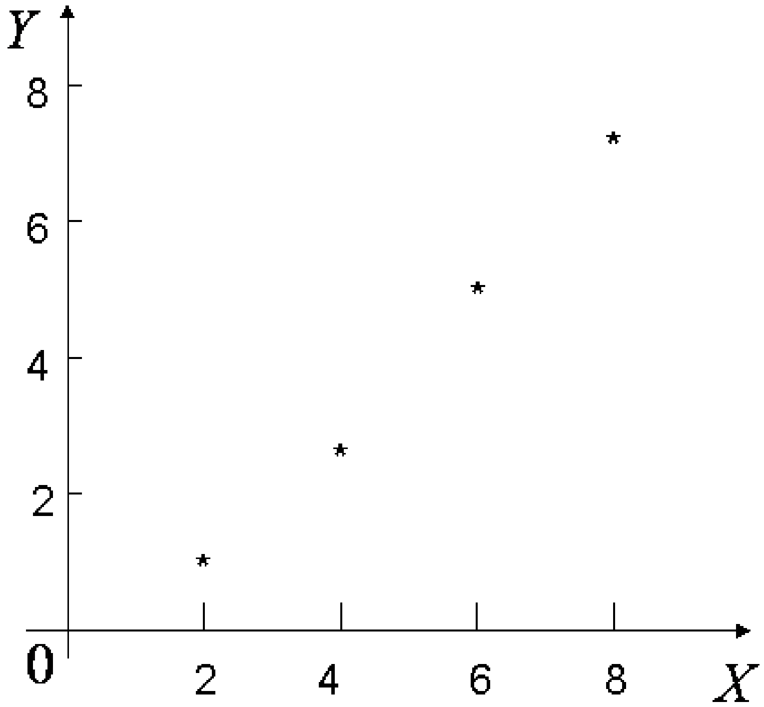
这些点的分布接近一条直线，

$$y = a + bx$$

➢这是一个直线方程，无论怎么选择a, b，直线都不可能同时过全部数据点。

➢怎样选取a, b才能使直线“最好”地反映数据点的基本趋势？

➢首先要建立好坏标准；



残差（误差）衡量 ：

如果某个直线能很好逼近数据规律，即a, b确定，

$$y_i^* = a + bx_i (i = 1, 2, 3, 4)$$

所以根据方程可以计算近似值，他与实测值y_i之差

$$e_i = y_i - y_i^* = y_i - a - bx_i \quad (i = 1, 2, 3, 4)$$

称为残差。残差的大小可作为衡量近似函数好坏的标准。

准则（判断标准）：

（1）使残差的最大绝对值最小，即

$$\max_i |e_i| = \min$$

（2）使残差的绝对值之和最小，即

$$\sum_i |e_i| = \min$$

函数的最佳一致逼近

（3）使残差的平方和最小，即

$$\sum_i e_i^2 = \min$$

函数最佳平方逼近

数据拟合的最小二乘法问题

根据给定的数据组（$x_{\mathrm{i}}, y_{\mathrm{i}}$）(i=1,2,…$N$), 选取近似函数形式，即给定函数类$H$, 求函数$\varphi(x) \in H$，使，

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - \varphi(x_i)]^2 --> \min$$

该函数称为这组数据的最小二乘函数。通常H取为一些比较简单函数的集合，一般为低次多项式 $\varphi = span\{1, x, \cdots, x^n\}$ ，如1次多项式（a, b待定）。 $\varphi(x) = a + bx$

据函数极值方法，下式成立

$$\frac{\partial Q}{\partial a}=0; \qquad \frac{\partial Q}{\partial b}=0$$

将Q代入可以得到线性方程组(式42)

$$\begin{cases} aN+b\sum x_i=\sum y_i \\ a\sum x_i+b\sum x_i^{\,2}=\sum x_i y_i \end{cases}$$

例:给定一组实验数据如下

i	1	2	3	4
xi	2	4	6	8
yi	1.1	2.8	4.9	7.2

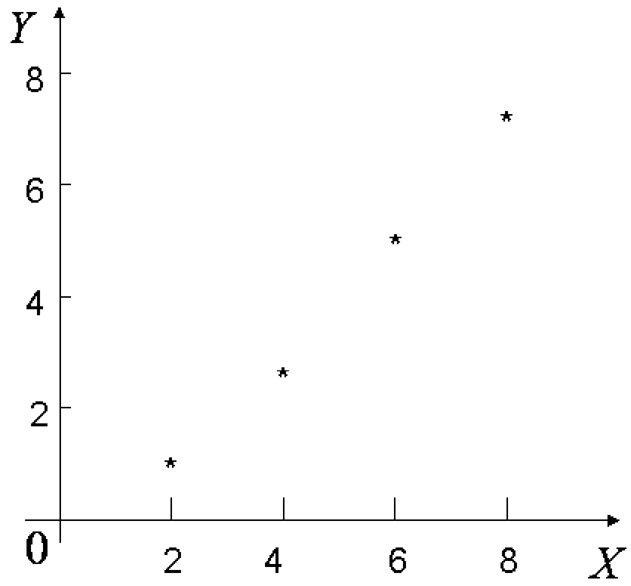
求$y=a+bx$的一次拟合曲线

解： $N=4$, 带入公式

$$\begin{cases} 4a+20b=16 \\ 20a+120b=100.4 \end{cases}$$

得到:

$a=-1.1, b=1.02$ $y=1.02x-1.1$



2、多项式拟合

根据给定的数据组（x_i, y_i）(i=1, 2,⋯,N)，求一个m次多项式（m<N）

$$y = p_m(x) = a_0 + a_1 x + \cdots + a_m x^m = \sum_{j=0}^{m} a_j x^j$$

使所有采样点的误差满足：

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - p_m(x_i)]^2 = Q(a_0, a_1, \cdots, a_m) --> \min$$

Q是一个多元函数，所以求这个多元函数的极值问题。

2、多项式拟合

由多元函数取极值的必要条件，得方程组

$$Q(a_0,a_1,\cdots,a_m)=\sum_{i=1}^{N}\left[y_i-(a_0+a_1x_i+\cdots+a_jx_i^j+\cdots+a_mx_i^m)\right]^2$$

求导

$$\frac{\partial Q}{\partial a_j}=-2\sum_{i=1}^{N}\left[y_i-\sum_{k=0}^{m}a_kx_i^k\right]x_i^j=0$$

移项得

$$\sum_{k=0}^{m}a_k\left(\sum_{i=1}^{N}x_i^{k+j}\right)=\sum_{i=1}^{N}y_ix_i^j \qquad (j=0,1,2,\cdots,m)$$

2、多项式拟合

写成方程组：

$$
\begin{cases}
a_0 N + a_1 \sum_{i=1}^{N} x_i + a_2 \sum_{i=1}^{N} x_i^2 + \cdots + a_m \sum_{i=1}^{N} x_i^m = \sum_{i=1}^{N} y_i \\
a_0 \sum_{i=1}^{N} x_i + a_1 \sum_{i=1}^{N} x_i^2 + a_2 \sum_{i=1}^{N} x_i^3 + \cdots + a_m \sum_{i=1}^{N} x_i^{m+1} = \sum_{i=1}^{N} y_i x_i \\
\qquad \cdots\cdots \\
a_0 \sum_{i=1}^{N} x_i^m + a_1 \sum_{i=1}^{N} x_i^{m+1} + a_2 \sum_{i=1}^{N} x_i^{m+2} + \cdots + a_m \sum_{i=1}^{N} x_i^{2m} = \sum_{i=1}^{N} y_i x_i^m
\end{cases}
$$

这是最小二乘拟合多项式的系数a_j所满足的方程组，称为正则方程组或法方程组。

2、多项式拟合

正则方程组解的存在

1、由函数组$\{1, x, x^2 \ldots x^m\}$的线性无关性，可以证明，方程组存在唯一解；

2、反证法 （定理7 P39）；

3、m比较大时，正则方程组是病态方程组。

2、多项式拟合

非线性拟合模型

有时根据给定数据图形，其拟合函数$y=f(x)$表面上不是线性或多项式的形式，但通过变换仍可化为线性模型。

例如，$S(x)=ae^{bx}$，　若两边取对数得

$$\ln S(x)=\ln a+bx,$$

此时，若令 $\overline{S}(x)=\ln S(x),\quad A=\ln a,\quad B=b,$

则

$$\overline{S}(x)=A+Bx,$$

这样就变成了形如的线性模型 .

2、多项式拟合

例　求数据表的最小二乘法拟合的二次多项式函数

xi	-1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	1
yi	50	40	25	20	18	21	35	56	66

在MATLAB命令窗口实现

```
x=-1:0.25:1;
y=[50,40,25,20,18,21,35,56,66]
p=polyfit(x,y,2);
  xi=-1:0.01:1;
  yi=polyval(p,xi);
  plot(xi,yi,x,y, 'o');
```