《线性代数》

第六章 线性空间

§6.4 最小二乘问题

2019秋

杨晶 主讲

内容提要

- 问题提出: 最小二乘问题
- 最小二乘问题的求解方法: 法方程
- 最小二乘解的应用
- 投影与投影矩阵

(一) 问题提出: 近似解的标准

回顾: 线性方程组的基本问题之六——解的近似问题

具体地说, 设非齐次线性方程组为

$$A_{m\times n}\vec{X}=\vec{b}$$

当该方程组无解时, 称为矛盾方程组.

在一些实际的问题中, 由于各种因素的干扰和测量误差的影响, 使得方程组不存在精确解, 但又需要求得它的近似解, 以便对实际情况的一般规律进行估计和预测.

于是, 可考虑某种近似解. 那么, 应该采用什么样的近似标准呢？本节我们就来讨论这个问题, 我们仍然利用代数与几何相结合的方法来考虑.

- 从代数上看，矛盾方程组 $A_{m\times n}\vec{X}=\vec{b}$，满足条件 $\mathrm{r}(A)\neq\mathrm{r}(A,\vec{b})$. 进一步设 $A=(\vec{\alpha}_1,\cdots,\vec{\alpha}_n)$，其中 $\vec{\alpha}_1,\cdots,\vec{\alpha}_n\in\mathbb{R}^m$，则有如下向量表示形式

$$A\vec{X}=(\vec{\alpha}_1,\cdots,\vec{\alpha}_n)\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}=x_1\vec{\alpha}_1+x_2\vec{\alpha}_2+\cdots+x_n\vec{\alpha}_n.$$

当 $x_1,\cdots,x_n$ 跑遍全体实数 $\mathbb{R}$ 时，$\vec{X}$ 跑遍 $\mathbb{R}^n$，则上式就生成了 $\mathbb{R}^m$ 的一个子空间:

$$W=L(\vec{\alpha}_1,\cdots,\vec{\alpha}_n)=\left\{A\vec{X}\,\middle|\,\vec{X}\in\mathbb{R}^n\right\}:=\mathrm{C}(A).$$

即由 A 的全体列向量生成的子空间，也就是 A 的列空间 $\mathrm{C}(A)$.

- 因此，从几何上看，矛盾方程组 $A\vec{X}=\vec{b}$ 无解的原因就是 $\vec{b}\notin\mathrm{C}(A)$. 于是，求近似解的问题，就相当于在 $\mathrm{C}(A)$ 中找一个与 $\vec{b}$ 最接近的向量来近似它. 那么，这个近似的向量应该如何找呢?

设$\vec{b}\notin \mathrm{C}(A)=W$, 那么, 直观上看, W中最接近$\vec{b}$的向量就是它在W上的正交投影向量：$\vec{\beta}=\left(\vec{b}\right)_W$. 因 $\vec{\beta}\in W=\mathrm{C}(A)$，可设$\vec{\beta}=A\vec{X}_0$，则$\vec{X}_0$就是所要求的近似解，即

$$\left|\vec{b}-A\vec{X}_0\right|=\min_{\vec{X}\in\mathbb{R}^n}\left|\vec{b}-A\vec{X}\right|$$

两边平方得

$$\left|\vec{b}-A\vec{X}_0\right|^2=\min_{\vec{X}\in\mathbb{R}^n}\left|\vec{b}-A\vec{X}\right|^2$$

从这个意义上讲， $\vec{X}_0$ 是方程组的最优近似解，称为“最小二乘解”. 求最小二乘解的问题就称为“最小二乘问题”.

这里, 所谓的“最小二乘”是源于$\left|\vec{b}-A\vec{X}\right|^2$等于其分量的平方和.

上述讨论是在直观意义下进行的. 下面我们给出严格证明, 说明正交投影向量 $\vec{\beta}=\left(\vec{b}\right)_W$确实是$W$中与$\vec{b}$最近的向量.

定理1 (最佳逼近定理) 设W是欧氏空间$\mathbb{R}^n$中的子空间, 对$\mathbb{R}^n$中任意向量$\vec{b}$, 设$\left(\vec{b}\right)_W$为$\vec{b}$在W上的正交投影向量, 则

$$\left|\vec{b}-\left(\vec{b}\right)_W\right|<\left|\vec{b}-\vec{w}\right|,\ (\forall\vec{w}\in W, \text{且}\vec{w}\neq\left(\vec{b}\right)_W).$$

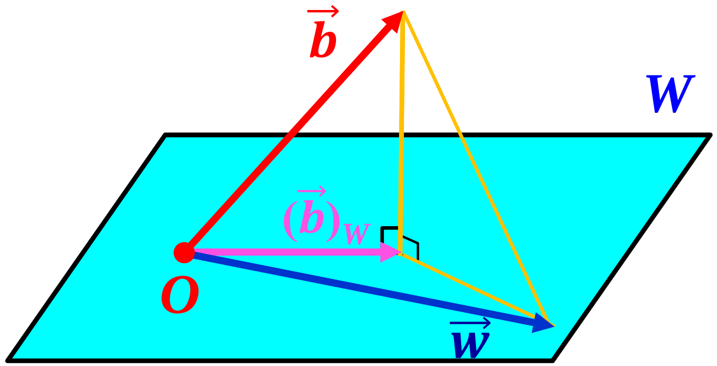
证明 若$\vec{b}\in W$, 则$\vec{b}=\left(\vec{b}\right)_W$, 则 $\vec{w}\neq\vec{b}$, 即$\left|\vec{b}-\vec{w}\right|>0$, 结论成立.

若$\vec{b}\notin W$, 那么, 由正交分解定理, 有$\vec{b}-\left(\vec{b}\right)_W\in W^{\perp}$. 特别有 $\vec{b}-\left(\vec{b}\right)_W$正交于$\left(\vec{b}\right)_W-\vec{w}$, 于是由勾股定理知:

$$\left|\vec{b}-\vec{w}\right|^2=\left|\vec{b}-\left(\vec{b}\right)_W\right|^2+\left|\left(\vec{b}\right)_W-\vec{w}\right|^2$$

(由$\vec{b}$, $\left(\vec{b}\right)_W$, $\vec{w}$的终点构成的直角三角形)

因$\vec{w}\neq\left(\vec{b}\right)_W$, 有$|(\vec{b})_W-\vec{w}|^2>0$, 代入上式即得定理1中的不等式. ■



(二) 最优近似解的求法: 最小二乘法

- 下面，我们来讨论如何求这个最优近似解.

令$\vec{\beta}=\left(\vec{b}\right)_W$, $A=(\vec{\alpha}_1,\cdots,\vec{\alpha}_n)$, 考虑向量$\vec{b}-\vec{\beta}$, 它与Col $A=W$正交, 故有

$$(\vec{b}-\vec{\beta})\perp\vec{\alpha}_i\ (i=1,2,\cdots,n)$$

用标准内积表示为

$$\left(\vec{b}-\vec{\beta},\vec{\alpha}_i\right)=\vec{\alpha}_i{}^T\left(\vec{b}-\vec{\beta}\right)=0,\qquad (i=1,2,\cdots,n)$$

将上述n个式子按行排列得

$$\begin{pmatrix}\vec{\alpha}_1^T\\ \vdots\\ \vec{\alpha}_n^T\end{pmatrix}\left(\vec{b}-\vec{\beta}\right)=A^T\left(\vec{b}-\vec{\beta}\right)=\vec{0}.$$

设$\vec{\beta}=A\vec{X}_0$, 且代入上式展开, 得 $A^TA\vec{X}_0=A^T\vec{b}$. 即近似解$\vec{X}_0$是方程组

$$A^TA\vec{X}=A^T\vec{b}$$

的解. 该方程组通常称为“法方程”.

把上面结果写为结论, 得到如下定理:

定理2 对于矛盾的非齐次线性方程组 $A\vec{X}=\vec{b}$, 有

(i) 可用法方程 $A^TA\vec{X}=A^T\vec{b}$的解作为 $A\vec{X}=\vec{b}$的最小二乘解.

(ii) 法方程 $A^TA\vec{X}=A^T\vec{b}$ 必有解, 且当 $\mathrm{r}(A)=n$ (A列满秩) 时, A^TA可逆, 法方程有唯一解

$$\vec{X}_0=\left(A^TA\right)^{-1}A^T\vec{b}.$$

注: 定理2给出了求最小二乘解的方法, 即求解对应的法方程. 并且在理论上保证了法方程必有解. 在实际计算中, 只须按一般非齐次线性方程组的求解步骤去计算法方程的通解即可.

证明: (i) 之前的推导已经说明原方程的最优近似解是法方程的解. 下面说明法方程的解必为原方程的最优近似解.

设n维向量$\vec{y}$满足法方程, 即$A^TA\vec{y} = A^T\vec{b}$, 从而有

$$A^T\left(\vec{b} - A\vec{y}\right) = \vec{0}.$$

这说明$\vec{b} - A\vec{y}$与A的列向量均正交，即 $(\vec{b} - A\vec{y}) \perp \mathrm{C}(A)$.

从而有分解式:

$$\vec{b} = A\vec{y} + (\vec{b} - A\vec{y}), \text{其中 } A\vec{y} \in \mathrm{Col}\, A, (\vec{b} - A\vec{y}) \in \mathrm{C}(A)^{\perp}.$$

由正交分解的唯一性知,

$$A\vec{y} = \left(\vec{b}\right)_{\mathrm{C}(A)}.$$

因此, $\vec{y}$是原方程的最小二乘解, 即最优近似解.

(ii) 要证法方程必有解，需证: $\mathrm{r}(A^TA)=\mathrm{r}\left(A^TA,A^T\vec{b}\right)$.

一方面，由秩的性质有

$$\mathrm{r}\left(A^TA,A^T\vec{b}\right)=\mathrm{r}\left(A^T\left(A,\vec{b}\right)\right)\leq \mathrm{r}(A^T)=\mathrm{r}(A)$$

另一方面，

$$\mathrm{r}\left(A^TA,A^T\vec{b}\right)\geq \underline{\mathrm{r}(A^TA)=\mathrm{r}(A)}$$

综上得，

$$\mathrm{r}\left(A^TA\right)=\mathrm{r}(A)=\mathrm{r}\left(A^TA,A^T\vec{b}\right).$$

故法方程$A^TA\vec{X}=A^T\vec{b}$必有解.

又因A^TA为n阶方阵, A列满秩时, $\mathrm{r}(A^TA)=\mathrm{r}(A)=n$, 故$A^TA$可逆. 两边左乘$(A^TA)^{-1}$即得法方程的唯一解: ■

$$\vec{X}_0=\left(A^TA\right)^{-1}A^T\vec{b}.$$

例1 给出以下线性方程组 $\begin{cases} x+y+z=1 \\ x+y+z=2 \end{cases}$ 的最优近似解.

解 此方程组比较简单，明显是一个矛盾方程组. 令

$$A=\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}\quad \vec{b}=\begin{pmatrix}1\\2\end{pmatrix}$$

从而, 有 $A^TA=\begin{pmatrix}1&1\\1&1\\1&1\end{pmatrix}\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}=\begin{pmatrix}2&2&2\\2&2&2\\2&2&2\end{pmatrix}$, 而 $A^T\vec{b}=\begin{pmatrix}3\\3\\3\end{pmatrix}$.

故法方程为:

$$x+y+z=\frac{3}{2}.$$

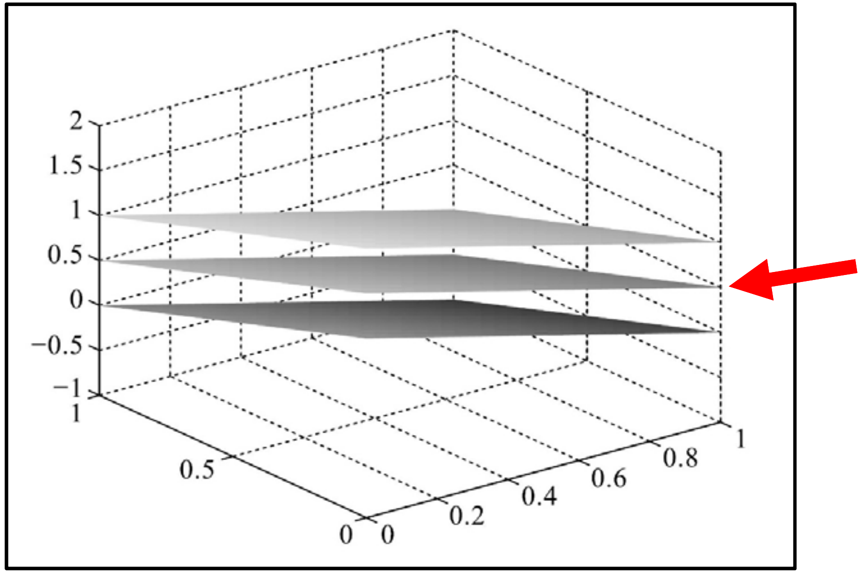
用高斯消元法, 可解得法方程的通解, 也即原方程的最小二乘解为

$$\vec{X}_0=\begin{pmatrix}3/2\\0\\0\end{pmatrix}+k_1\begin{pmatrix}-1\\1\\0\end{pmatrix}+k_2\begin{pmatrix}-1\\0\\1\end{pmatrix}\quad (k_1,k_2\in\mathbb{R}).$$

- 下面从几何方面来看例1的结果.

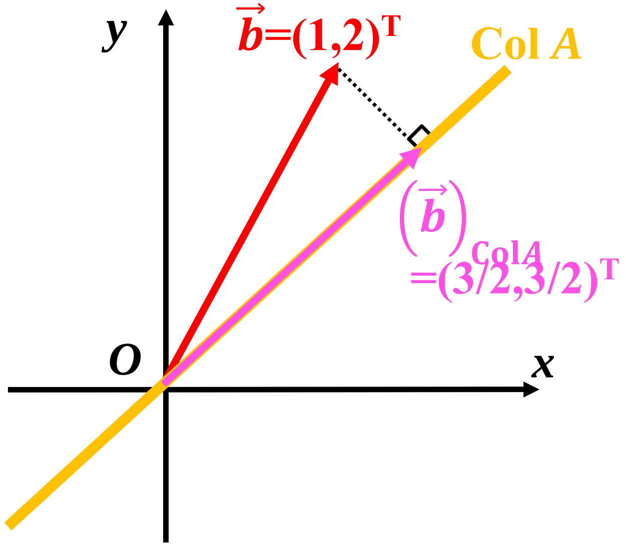
➢ 行图的观点下

最小二乘解的集合是一个平面, 与原来已知的两个平面平行, 落在它们中间, 且距离相等.



➢ 列图的观点下

最小二乘解左乘A后, 等于向量$\vec{b}=(1,2)^T$的正交投影向量$(3/2, 3/2)^T$, 也即Col A中与$\vec{b}$最接近的向量.



(三) 最小二乘解的应用

- 在一些实际问题里, 人们想知道某些变量之间的函数关系, 可通过实验获得数据, 再从实验数据求出函数关系, 这就是统计学中所谓“曲线拟合”或“回归”问题. 我们从最简单的例子出发来说明.

例2 设有以下实验数据

x_i	1	2	3	4	5
y_i	1.2	1.5	2.3	2.4	3.3

求形如 $y = a + bx$ 的函数, 其中a, b为待定参数，使它与实验数据的误差平方和最小.

解 假设a,b已确定, 则当$x=1,2,3,4,5$时, y应得到如下理论值，即

$$\begin{cases} a+b=y_1^* \\ a+2b=y_2^* \\ a+3b=y_3^* \\ a+4b=y_4^* \\ a+5b=y_5^* \end{cases} \Rightarrow \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}}_{A} \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} y_1^* \\ \vdots \\ y_5^* \end{pmatrix}}_{\vec{y}^*} \quad \text{记为 } A\begin{pmatrix} a \\ b \end{pmatrix} = \vec{y}^*.$$

它们与实验数据的误差平方和为: $\sum_{i=1}^{5}(\vec{y}_i^* - \vec{y}_i)^2$.

表示为向量内积的形式且求最小值, 可表示为

$$\min\ |\vec{y}^* - \vec{y}|^2 = \min_{(a,b)^{\mathrm{T}}\in\mathbb{R}^2} \left|\vec{y} - A\begin{pmatrix} a \\ b \end{pmatrix}\right|^2 .$$

因此求误差平方和的最小值，就是求关于$A\begin{pmatrix} a \\ b \end{pmatrix} = \vec{y}$ 的最小二乘解.

由于 $A=\begin{pmatrix}1&1\\1&2\\1&3\\1&4\\1&5\end{pmatrix}$，$\vec{y}=(1.2, 1.5, 2.3, 2.4, 3.3)^{\mathrm{T}}$, 且$A$明显为列满秩矩阵, 则由定理2(ii), 分别计算得:

$$A^TA=\begin{pmatrix}5&15\\15&55\end{pmatrix},\ A^T\vec{y}=\begin{pmatrix}10.7\\37.2\end{pmatrix},$$

$$\left(A^TA\right)^{-1}=\frac{1}{50}\begin{pmatrix}55&-15\\-15&5\end{pmatrix}.$$

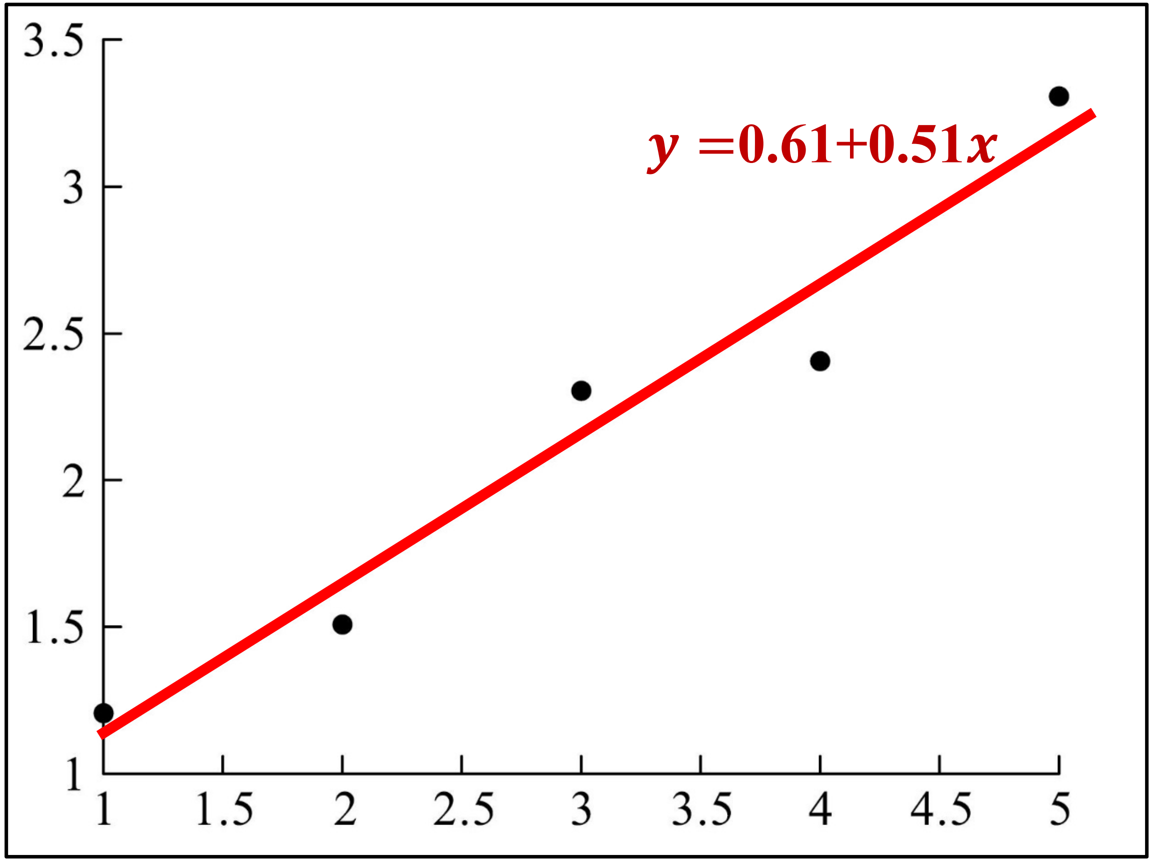
故最小二乘解为:

$$\begin{pmatrix}a_0\\b_0\end{pmatrix}=\left(A^TA\right)^{-1}A^Ty=\frac{1}{50}\begin{pmatrix}55&-15\\-15&5\end{pmatrix}\begin{pmatrix}10.7\\37.2\end{pmatrix}=\begin{pmatrix}0.61\\0.51\end{pmatrix}$$

故满足条件的一次函数为

$$y=0.61+0.51x.$$

求出以上最小二乘解后，得到拟合图像如下：



例3 设有以下实验数据

x_i	1	2	3	4	5
y_i	6.1	10.5	18.4	26.5	38.2

求二次函数$y = a + bx + cx^2$ (a, b, c为待定参数)，使它与实验数据的误差平方和最小.

解 同上例分析, 可说明使得误差平方和最小的二次函数系数 $(a_0, b_0, c_0)^{\mathrm{T}}$，就是如下线性方程组的最小二乘解：

范德蒙矩阵

$$\begin{cases} a + b + c = 6.1 \\ a + 2b + 4c = 10.5 \\ a + 3b + 9c = 18.4 \\ a + 4b + 16c = 26.5 \\ a + 5b + 25c = 38.2 \end{cases} \qquad 令\ A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix}, \vec{y} = \begin{pmatrix} 6.1 \\ 10.5 \\ 18.4 \\ 26.5 \\ 38.2 \end{pmatrix}.$$

由A是列满秩的, 则由定理2(ii), 分别计算得:

$$A^TA=\begin{pmatrix}5 & 15 & 55\\ 15 & 55 & 225\\ 55 & 225 & 979\end{pmatrix},\quad A^T\vec{y}=\begin{pmatrix}99.8\\ 379.3\\ 1592.7\end{pmatrix},$$

$$\left(A^TA\right)^{-1}=\begin{pmatrix}4.6 & -3.3 & 0.5\\ -3.3 & \frac{18.7}{7} & -\frac{3}{7}\\ 0.5 & -\frac{3}{7} & \frac{1}{14}\end{pmatrix}.$$

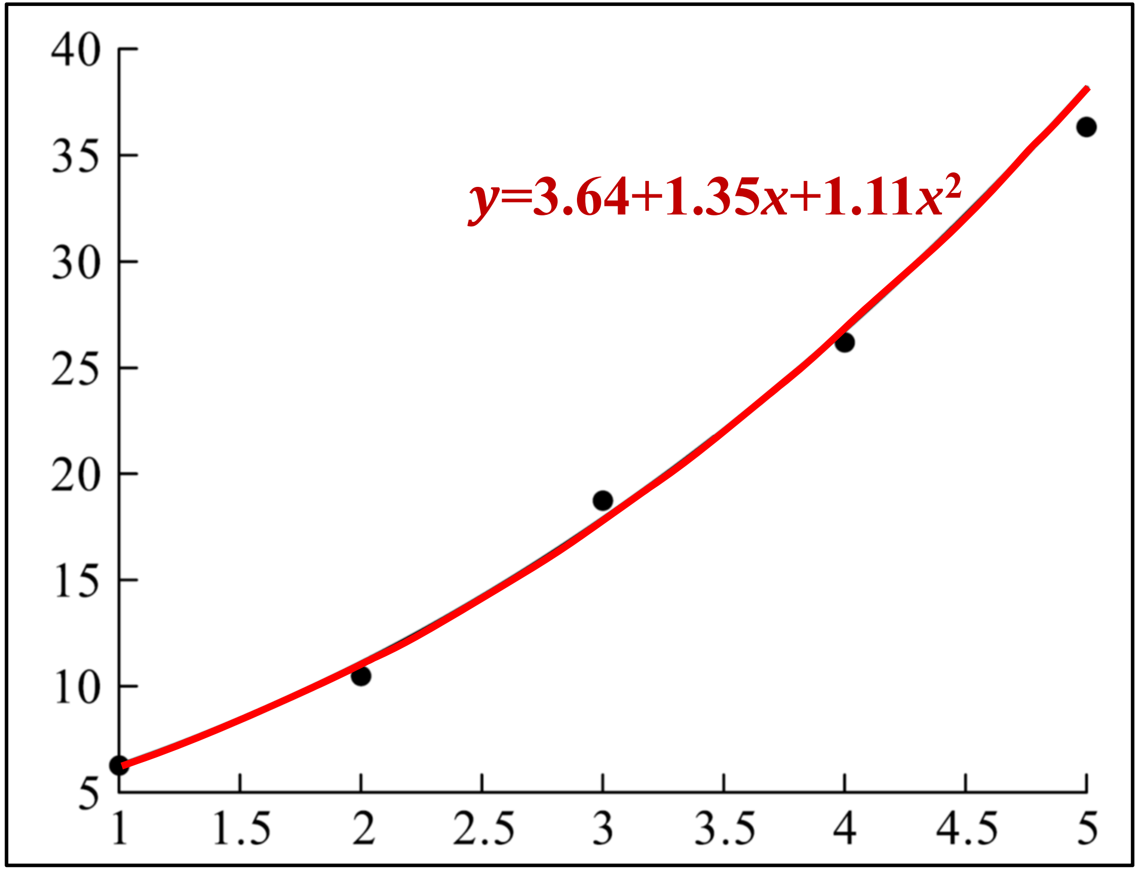
于是最小二乘解为

$$\vec{X}_0=\begin{pmatrix}a_0\\ b_0\\ c_0\end{pmatrix}=\left(A^TA\right)^{-1}A^T\vec{b}=\begin{pmatrix}3.64\\ 1.35\\ 1.11\end{pmatrix}.$$

所以误差平方和最小的二次函数为

$$y=3.64+1.35x+1.11x^2.$$

求出以上最小二乘解后，得到拟合图像如下：



(四) 投影与投影矩阵

问题: (1) 给定$\vec{v} \in \mathbb{R}^n$与子空间$W \subseteq \mathbb{R}^n$, 求向量$\vec{v}$到子空间W的正交投影向量$(\vec{v})_W$.

(2) 把上述投影的过程看作一种“运动”, 该“运动”是否可通过一个矩阵来表示?

- 当$\dim W=1$时, 称为线投影. 设$W=L(\vec{\alpha})$, 则$(\vec{v})_{\vec{\alpha}} = \dfrac{(\vec{v},\vec{\alpha})}{(\vec{\alpha},\vec{\alpha})}\vec{\alpha} = \dfrac{(\vec{\alpha}^{\mathrm{T}}\vec{v})\vec{\alpha}}{\vec{\alpha}^{\mathrm{T}}\vec{\alpha}} = \dfrac{\vec{\alpha}\,\vec{\alpha}^{\mathrm{T}}}{\vec{\alpha}^{\mathrm{T}}\vec{\alpha}}\vec{v}$.

  令$P_{\vec{\alpha}} = \dfrac{\vec{\alpha}\,\vec{\alpha}^{\mathrm{T}}}{\vec{\alpha}^{\mathrm{T}}\vec{\alpha}}$为$n$阶方阵, 则有$(\vec{v})_{\vec{\alpha}}=P_{\vec{\alpha}}\,\vec{v}$.

- 当$\dim W=k<n$时, 设$\vec{\alpha}_1,\cdots,\vec{\alpha}_k$为$W$的一组基, 令$A_{n\times k}=(\vec{\alpha}_1,\cdots,\vec{\alpha}_k)$, 则$W=C(A)$, 相当于将$\vec{v}$投影到的列空间$C(A)$上. 按上述最小二乘的讨论, 即可求得$(\vec{v})_W$.

  具体是: 求解法方程$A^TA\vec{x} = A^T\vec{v}$, 由A列满秩, 得 $\vec{x}=\left(A^TA\right)^{-1}A^T\vec{v}$, 再左乘$A$, 得

  $(\vec{v})_W = A\left(A^TA\right)^{-1}A^T\vec{v}$, 故投影矩阵为 $P_{C(A)}=A\left(A^TA\right)^{-1}A^T$.

思考: 当$\dim W = n$时, 投影矩阵$P_W = ?$

回答: 此时, $W=\mathbb{R}^n$, A为可逆方阵, $P_W=AA^{-1}\left(A^T\right)^{-1}A^T = I_n$, $P_W\vec{v} = \vec{v}$.

- 投影矩阵为 $P=P_{\mathrm{C}(A)}=A\left(A^{T}A\right)^{-1}A^{T}$, 其中$A$列满秩.

- 易验证: $P^2=P$, $P^{\mathrm{T}}=P$. 从几何意义来看, 连续做两次投影, 第二次就不变化了.
- 根据正交分解定理, $I-P=P_{W^{\perp}}$, 即$I-P$是到$W^{\perp}=\mathrm{N}(A^{\mathrm{T}})$的投影矩阵.
- 若将$\vec{\alpha}_1,\cdots,\vec{\alpha}_k$标准正交化, 即$A=Q_{n\times k}R_{k\times k}$, 代入上式得

$$P_{\mathrm{C}(A)}=QR\left(R^{\mathrm{T}}Q^{\mathrm{T}}QR\right)^{-1}R^{\mathrm{T}}Q^{\mathrm{T}}=QRR^{-1}\left(R^{\mathrm{T}}\right)^{-1}R^{\mathrm{T}}Q^{\mathrm{T}}=QQ^{\mathrm{T}}.$$

定义: 若方阵A满足 $A^2=A$, $A^{\mathrm{T}}=A$, 则称A为**投影阵 (projection matrix)**.

其中第二个条件是为了保证, $\forall\vec{\alpha},\vec{\beta}\in\mathbb{R}^n$, 有$A\vec{\alpha}\perp(I-A)\vec{\beta}$, $\iff A^{\mathrm{T}}(I-A)=(I-A)^{\mathrm{T}}A=O$

$\iff A^{\mathrm{T}}=A$.

命题: 设P为n阶投影阵, 则$\mathrm{C}(P)=\mathrm{N}(I-P)$, $\mathrm{C}(I-P)=\mathrm{N}(P)$.

证明: 因为$P^2=P$, 则$P(I-P)=O$. $\Rightarrow \mathrm{C}(I-P)\subseteq\mathrm{N}(P)$

另一方面, $\forall\vec{\alpha}\in\mathrm{N}(P)$, 有$\vec{\alpha}=(I-P)\vec{\alpha}\in\mathrm{C}(I-P)$, $\Rightarrow \mathrm{N}(P)\subseteq\mathrm{C}(I-P)$

$\mathrm{N}(P)=\mathrm{C}(I-P)$

另一个等式同理可证. (留作习题)

本讲小结

- 最小二乘解：求近似最优解$\vec{X}_0$, s.t.$\left|\vec{b}-A\vec{X}_0\right|^2=\min\limits_{\vec{X}\in\mathbb{R}^n}\left|\vec{b}-A\vec{X}\right|^2$.
- 最小二乘问题的求解方法： 法方程$A^TA\vec{X}=A^T\vec{b}$, 必有解
- 最小二乘解的应用：实验数据拟合初步
- 子空间投影与投影矩阵： $P_{C(A)}=A\left(A^TA\right)^{-1}A^T$,

  $P^2=P$, $P^T=P$.