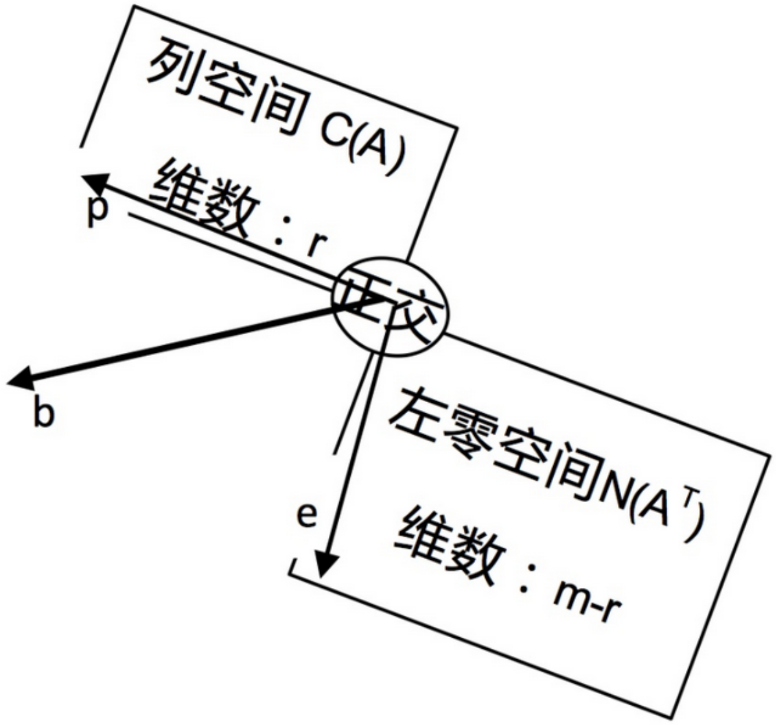
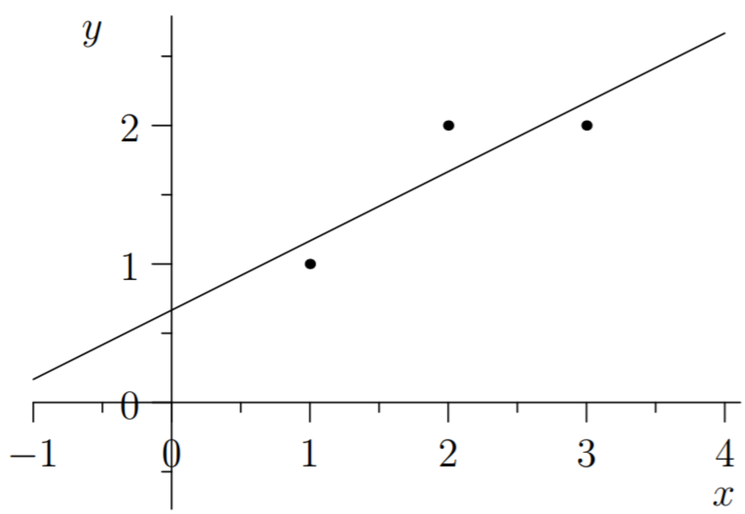
Chapter14基于投影矩阵的最小二乘法原理，标准正交向量组

Section1.本章两张图中其中一张



这张图说明了,向量b总可以分为两个分量,一个分量在A的列空间中,另一个分量垂直于A的列空间(也即在A的左零空间中)。而上述投影矩阵的作用就是保留列空间中的分量p,去掉左零空间中的分量e。

Section2.一个实例(最小二乘法)



图中三个点 为$(1,1)$ $(2,2)$ $(3,2)$，我们设这个直线为$y=C+Dx$，根据以上条件可以得到方程组$\begin{cases} C+D & =1 \\ C+2D & =2 \\ C+3D & =2 \end{cases}$,写作矩阵形式有$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$，也就是我们的$Ax=b$,显然该方程组无解。

那么我们就要考虑找到一条最优的直线 $y=C+Dx$ 来拟合图中的三个点。我们需要注意到的一点就是，**最优**是指什么?
在寻求最优解之前,我们需要先定义总误差是什么,因为总误差能够衡量直线是否是更优的,定义了总误差我们才能通过最小化这个量,来找到最好的 C 和 D(也即最优的直线)。
这里,我们定义误差为$Ax-b=e$ 的模长的平方来作为误差,也即
$|Ax-b|^2=|e|^2=e_1^2+e_2^2+e_3^2$,我们要求其最小平方和(也即最小二乘)。

Subsection1.利用微积分的偏导来求最优解

将误差展开用C和D的二元函数如下:
$|e|^2=e_1^2+e_2^2+e_3^2=(C+D-1)^2+(C+2D-2)^2+(C+3D-2)^2=3C^2+14D^2+9-10C-22$
(这里说明误差的分析是拟合的曲线与原本的点的纵坐标差)

误差对C求偏导为$6C-10+12D=0$,说明单看C的话,随着C的增长,总误差的斜率先为负数后为正数,也即总误差先下降后上升。误差对D求偏导为$28D-22+12C=0$,说明单看D的话,随着D的增长,总误差的斜率先为负数后为正数,也即总误差先下降后上升。因此,总误差的驻点显然也即总误差的最小值(最优值)

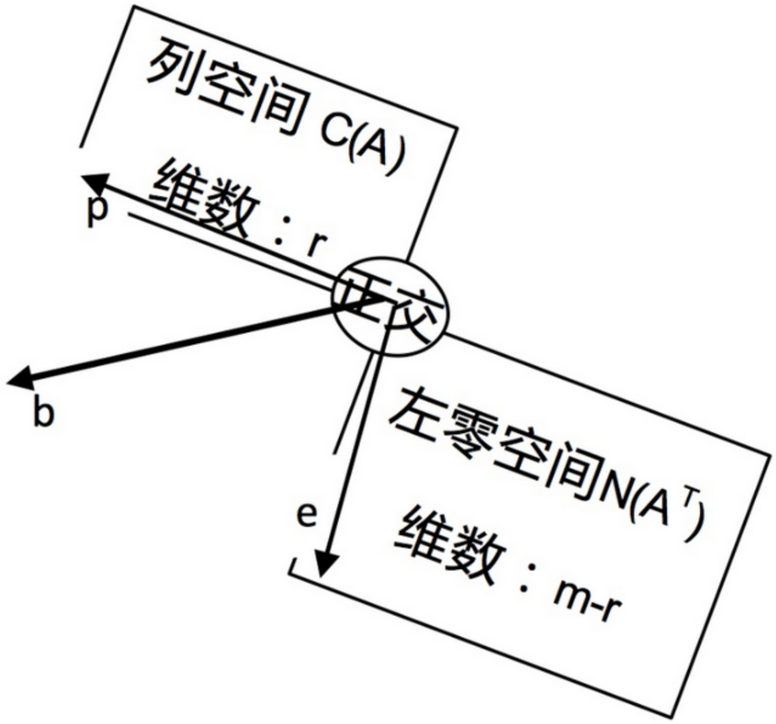
求解方程组$\begin{cases} 3C-5+6D=0 \\ 14D-11+6C=0 \end{cases}$得$\hat{C}=2, \hat{D}=\frac{1}{2}$,因此最优直线为$y=\frac{2}{3}+\frac{1}{2}x$。

代入x可求得$p_1=\frac{7}{6}, p_2=\frac{5}{3}, p_3=\frac{13}{6}$,自然$e_1=-\frac{1}{6}, e_2=\frac{1}{3}, e_3=-\frac{1}{6}$。
于是我们得到$p=\begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix}, \quad e=\begin{bmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix}$,易看出$b=p+e$,且 $p^Te=0$(也即$p\perp e$)。
容易知道

$$p^Te=e^Ta_1=e^Ta_2=0$$

综上可知,我们所求得的误差向量e确实垂直于整个列空间,如$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$(投影向量$p$也在$A$的列空间中)。

Subsection2.利用线性代数的投影来求解



还是关注这幅图，此时我们已知$A\hat{x}=p$与b，b我们已知一切信息，而$A\hat{x}=p$我们只知道方向(位于列空间中)，我们怎样控制$e=b-A\hat{x}$才能让e最小呢？显然是垂直。

因此我们就会有

$$A^TAx=A^Tb$$

这里我们考虑到:

$$A=\begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}\text{则}A^TA=\begin{bmatrix}3&6\\6&14\end{bmatrix}A^Tb=\begin{bmatrix}5\\11\end{bmatrix}$$

写成方程组形式为: $\begin{cases}3\hat{C}+6\hat{D}=5\\6\hat{C}+14\hat{D}=11\end{cases}$, 也称其为正规方程组(normal equations)。注意到该正规方程组正是先前求偏导所得的方程组。故所求得的结果也都是一样的:$\hat{C}=2,\hat{D}=\frac{1}{2}$。我们现在所做的运算实际上也称为线性回归(linear regression)。

此外,还需要补充说明一点,如果在上述例题中,还有另外一个点如 $(0,100)$,那么最小二乘法就很容易被这个明显离群的值影响,通常使用最小二乘的时候要先去除掉明显离群的点!

Section2.标准正交向量组

- 标准：向量的模为1
- 正交：指向量组的向量之间内积为0

我们在前面一直在探究A的列的线性无关性，有一种线性无关的情况是比较特殊的:互相垂直的各列一定是线性无关的。

更特殊地,我们会要求互相垂直的单位向量(标准正交),比如 $\begin{bmatrix}1\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\end{bmatrix}$, $\begin{bmatrix}0\\0\\1\end{bmatrix}$, 这些向量所组成的向量组一般被称为标准正交向量组, 标准正交向量组中的向量互相垂直(正交)且为单位向量(标准)!

同样的标准正交向量组还有: $\begin{bmatrix}\cos\theta\\sin\theta\end{bmatrix}$, $\begin{bmatrix}-\sin\theta\\cos\theta\end{bmatrix}$。