16.投影矩阵和最小二乘

对于投影矩阵 P,

1. 投影矩阵作用特殊量

$$P = A(A^TA)^{-1}A^T$$

1. 如果矩阵 b 在列空间中，那么 $Pb = b$
2. 如果矩阵 b 垂直于列空间，那么 $Pb = O$

对于 （2.） 这是很显然的，从几何角度考虑可以立刻得出结论。
即使是从特殊值的角度来考虑，也是很好想的。（比如一个直线垂直于一平面，那么直线上一点（所有点）做对于平面的投影，它们都会汇于一点。）

但是也可以通过公式角度理解。

$$Pb = A(A^TA)^{-1}A^Tb$$

由于**矩阵 b 垂直于列空间**，同时我们也知道，垂直列空间的就是**转置矩阵对应的零空间**。
那么岂不是就是说 b 在这个零空间里面吗。

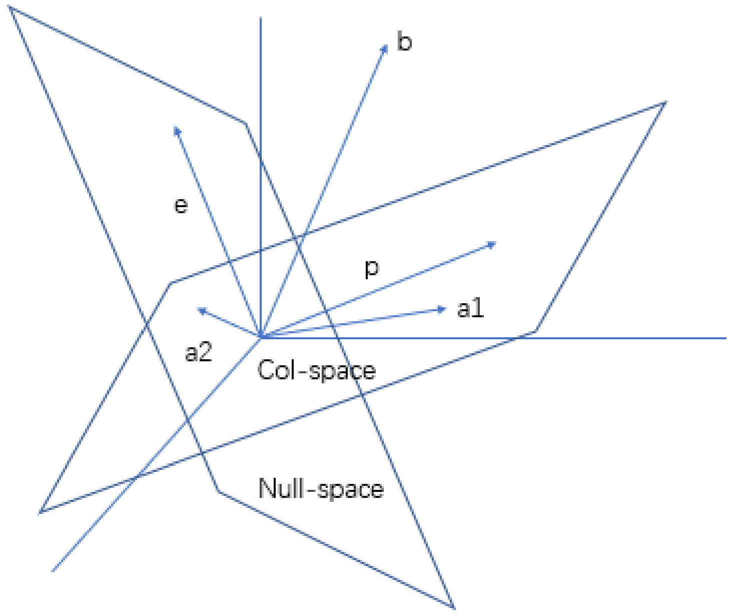
所以就有：

$$A^Tb = 0$$

所以 $Pb = 0$ 才成立。

2. 投影矩阵作用一般情况

i. 线性代数角度——投影矩阵



可以看到对于任意 b， 其投影 p 在 列空间 中，与投影正交的矩阵 e 在 A^T 的零空间中。

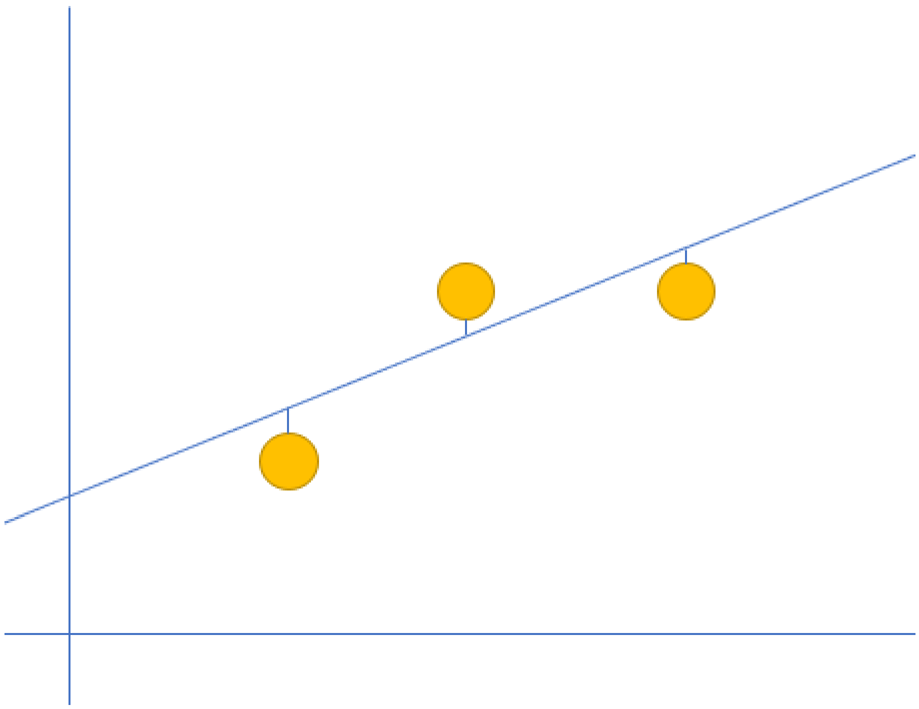
对于上一讲最后的矩阵方程，这个矩阵方程是**没有解的**（因为给的方程数多于参数数目，并且**存在噪声**）。

$$\begin{cases} C+D=1 \\ C+2D=2 \\ C+3D=2 \end{cases} \quad \Rightarrow \quad \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

但是对于从最小二乘的角度来看，这样一个方程是有解的：（因为人为地添加了残差项，**给予方程一定的“矫正”**）

$$||AX-b||^2 = ||e||^2$$

从图像来看，是这样的：



称上图的“小线段”为“误差”。现在将“误差”们投影到直线上。三点对应的在直线上的点记作 P_1, P_2, P_3 。

如果所求方程并非 $AX = b$,
而是 $A^TA\hat{X} = p = A^Tb$，那么就可以解出方程。
（要求 $R(A^TA) = R(A^TA, A^Tb) = n$ （根据秩与方程解的关系））

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 5 \\ 11 \end{bmatrix}, \qquad \hat{X} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$$

所以有正规方程：

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

这个方程组也就是：

$$\begin{cases} 3\hat{C} + 6\hat{D} = 5 \\ \\ 6\hat{C} + 14\hat{D} = 11 \end{cases}$$

ii. 高等数学角度——多元函数极值

如果要验证这个方程组就是使得**原本方程组的“误差”最小**的方程，那么可以尝试写出误差函数 $err(C, D) = ||e||^2 = ||AX - b||^2$

$$\begin{aligned} err(C, D) &= (C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2 \\ &= 3C^2 + 14D^2 + 12CD - 10C - 22D + 9 \end{aligned}$$

要求最小值点，肯定就是求偏导数, 并令其为 0：

$$\begin{cases} \frac{\partial err}{\partial C} = 6C + 12D - 10 = 0 \\ \\ \frac{\partial err}{\partial D} = 28D + 12C - 22 = 0 \end{cases}$$

约分之后可以发现，得到的方程组与上面的那个一模一样！

解得：

$$\begin{cases} \hat{C} = \frac{1}{2} \\ \hat{D} = \frac{2}{3} \end{cases}$$

$$\hat{y} = \frac{1}{2}t + \frac{2}{3}$$

从而从两个角度得到了相同的结果。并且这两个角度都是十分清晰的。相较于根据多元函数表达式，求偏导数并使其为 0 的第二种角度，第一种角度的方法以线性代数中几个子空间之间的关系为基础，结合图像进行分析，从另外一种几何角度给出了相同的答案。

iii. 最小二乘中的几何关系

回顾之前的解答，还有几个未知量在最后没有列出：

1. 首先是 p, 作为样本点的投影，表示的也就是预测值。
2. 其次是 b,

由上：

$$p = \begin{bmatrix} \frac{7}{6} \\ \\ \frac{5}{3} \\ \\ \frac{13}{6} \end{bmatrix}, \qquad e = \begin{bmatrix} -\frac{1}{6} \\ \\ \frac{1}{3} \\ \\ -\frac{1}{6} \end{bmatrix}$$

可以看出 $b = p + e$，并且 $p^T e = 0$
同时，由于 e 存在于 转置矩阵的零空间中，所以其正交于列空间中的所有向量。

对于矩阵 $A=\begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}$，其中列向量 $a_1=\begin{bmatrix}1\\1\\1\end{bmatrix}$, 也是列空间中的一个向量，当然它也和 e 呈正交关系。

3. 证明一个小结论

证明：A 为一个 $m*n$ 矩阵，A^TA 可逆当且仅当 A 列线性无关。

1. 解法一：
   那么也就等价于证明：矩阵方程 $A^TAX=O$ 只有零解。
   这句话等同于矩阵 A^TA 的零空间维数为 0。
   等同于证明矩阵 A^TA 的行 / 列空间维数为 n。
   如果 A 列线性相关，那么说明至少存在一列可以被其他列表示。
   也就是说列空间维数不为 n。
   反证成立。
2. 解法二：
   左右同乘 X^T, 有: $X^TA^TAX=O$
   矩阵乘法满足结合律，所以令 $Y=AX$，可以将方程看作：$Y^TY=O$。
   而此方程左边 = $||Y||^2$
   所以可知 $AX=O$，自然就是成立的。
   而这代表 A 列线性无关。

这个小结论也可以进一步推出：$R(A^TA)=R(A)$, 教授说这个结论很重要，因为这是最小二乘法的基础。
如果 A^TA 是不可逆的，也就无法将最小二乘算法进行下去了。

教授没有说到，如果这里的 A^TA 还真的是不可逆的，那会如何。
实际上，A^TA 不可逆就意味着 $R(A^TA)=R(A^TA,A^Tb)<n$。整个矩阵方程组即使左右同乘 A^T, 也还是无法得到一个确切的解。
在这个情况下，矩阵有无穷多解。