Chapter13Ax=b无解情况的优化，投影矩阵

Section1.引入(怎样处理$Ax=b$无解时候的情况)

我们在处理$Ax=b$的过程中，会发现有解的情况是非常少的，哪怕是在列满秩的优秀前提下，也需要一定的条件才能满足x是有解的，比如说在A化成$RREF$的过程中，出现的零行对应的b的位置需要也为0，或者朴素的来说，需要b在A的列空间中。

但是显然，有解的情况会显著少于无解的情况，也就是说比如说在A化成$RREF$的过程中，出现的零行对应的b的位置不为0，那么我们要做相应的优化，让这些“坏数据”的b变成能够有解的b，那么我们采取什么样的手段才能解决呢？

一种常用的方法是在方程两侧乘以 A^T,无解方程从而改写成 $A^TA\hat{x}=A^Tb$,求解新方程的解 $\hat{x}$ 即为最优解。原理将在后面介绍。

注意,这个解$\hat{x}$ 并非是 $Ax=b$ 的解,我们已经假设 $Ax=b$ 是无解的,也即符合方程的 x 不存在。在已知 $Ax\neq b$ 的情况下,我们尝试求解 $A^TA\hat{x}=A^Tb$。

$A^TAx=A^Tb$的可解性

一定有解，参见(71 封私信 / 80 条消息) 如何证明ATAX=ATB一定有解？ - 知乎 (zhihu.com)

从几何意义上比较容易理解，把b投影到A的列空间中肯定会有解。

Section2.A^TA的性质

实际上,就算我们不懂原理,我们也大概能发现,乘以 A^T 给方程带来了什么好的变化。

我们在进行科学研究的时候，经常对于几个参数要统计成千上万的数据来去计算拟合，也就是说会形成一个$m>>n$的长方形矩阵，而乘以 A^T 以后,我们得到矩阵 A^TA,这个矩阵是一个对称方阵,至少我们已经避免了长方矩阵的情况。一旦 A^TA 是可逆的,那么解 $\hat{x}$ 就很容易求得了。

但实际上 A^TA 未必是可逆的,当 A 的各列线性相关的时候,A^TA 就不可逆了。

那么为了说明这一点我们考虑A^TA的两个性质:

- $N(A^TA)=N(A)$：事实上我们需要证明的是$A^TAx=0$与$Ax=0$是等价条件，那么证明如下:

$$A^TAx=0 \rightarrow x^TA^TAx=0 \rightarrow (Ax)^T(Ax)=0 \rightarrow Ax=0$$

$$Ax=0 \rightarrow A^TAx=0$$

  关于$(Ax)^T(Ax)=0 \rightarrow Ax=0$的说明:

$$A^TA=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = [0] \rightarrow x_1=x_2=\ldots=x_n=0 \rightarrow A=0$$

- $r(A^TA)=r(A)$：这是显然的，根据$N(A^TA)=N(A)$,说明$dim(N(A^TA))=dim(N(A))=n-r$,且$A^TA$为$n \times n$,$A$为$m \times n$,说明$n$相同，因此$r$相同。

综合这两个性质就可以说明，如果A的列不满秩，那么A^TA作为一个$n \times n$的矩阵就无法行列满秩，因此不可逆。

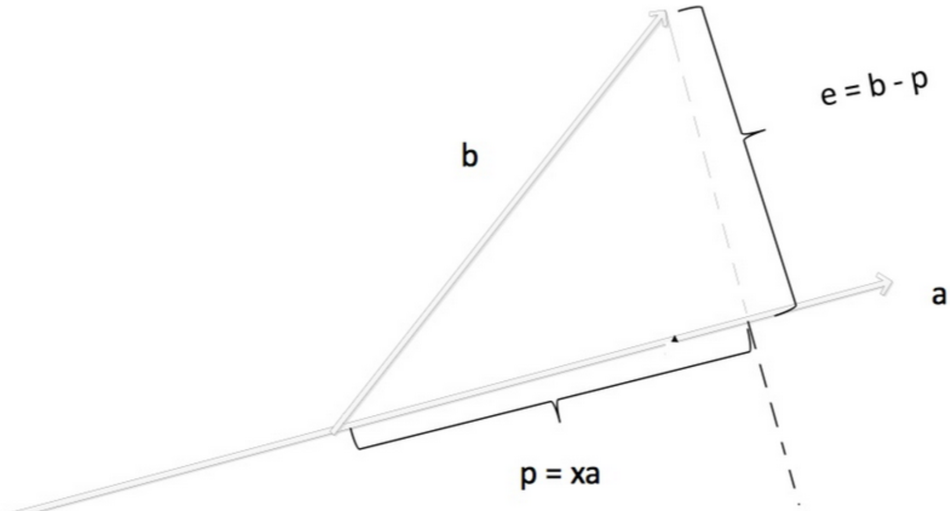
同时还要注意到A^TA是一个对称阵，即$(A^TA)^T=A^TA$,转置等于其本身。

Section3.对于该方法原理的解释

事实上，我们对于$Ax=b$的解的解释是:b在A的列空间中，那么如果b不在A的列空间中时，我们把它投影到A的列空间中，就可以实现一个最优解的过程，实际上,解的最优性就体现在b的变化最小上，而变化最小是通过投影来实现的，这就是原理的解释。那么现在我们就要考察，为什么利用投影，结果就会是乘以A^T呢?

Subsection1.基础知识:投影的矩阵表示

Subsubsection1.简单的二维投影



这里p为投影向量(projection)，显然为投影上的x倍。我们需要用已知的条件a,b来构成方程，那么这里显然有:

$$
\begin{aligned}
&a^T \times e = 0 \\
&a^T \times (b - xa) = 0 \\
&\text{因为}a^Ta\text{是一个数, 故}x = \frac{a^Tb}{a^Ta} \\
&\text{同样的，因为}x\text{是一个标量} \\
&\therefore p = xa = ax = \frac{aa^T}{a^Ta}b
\end{aligned}
$$

注意,其中的 a^Ta 是一个标量,而 aa^T 是一个矩阵。假设 b 变为原来的 2 倍,那么显然投影 p 也变为原来的 2 倍。但如果 a 变为原来的 2 倍,整个 $\frac{aa^T}{a^Ta}$ 并没有发生变化,所以投影 p 不变。

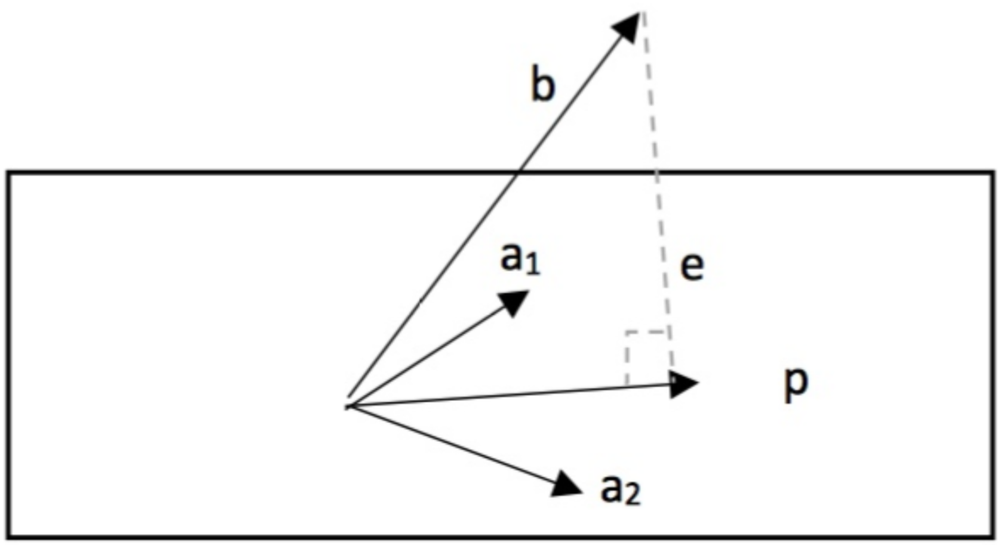
到这里,我们大概感觉到 $\frac{aa^T}{a^Ta}$ 此式的妙用了,该式是一个矩阵,令 $P = \frac{aa^T}{a^Ta}$,则有 $p = Pb$,我们将 P 称为投影矩阵,当向量与 P 相乘后,即可得到向量对应的到 a 上的投影。

Subsubsection2.投影矩阵P的性质

- $rank(P) = 1$,因为a^Ta是一个标量，而aa^T中我们根据秩1矩阵的模式，可以知道aa^T是秩1矩阵。
- 向量 a 是 P 的列空间的基,向量 a 所在的直线是 P 的列空间。
  这个可以通过几何方向与代数方向来考虑。
  - 几何方向:P 乘以任意向量 b 所得到的结果即为投影 p,而 p 就在 a 所在的直线上,这也即意味着,对 P 的各列进行任意的 b 对应的线性组合的结果都在 a 所在的直线上,所以向量 a 是 P 的列空间的基,向量 a 所在的直线是 P 的列空间。
  - 代数方向:可以知道aa^T是对a这一列向量做一个a^T的线性组合，因此实际上aa^T还是一个对a的线性组合，因此P的列空间就是a所在的列空间，而向量 a 是 P 的列空间的基。

- P 是对称的。显然,因为 $P = \frac{aa^T}{a^Ta}$ 中的 aa^T 是一个对称矩阵,所以 P 也是一个对称矩阵。
- $P^2 = P$。
  仍然可以通过几何方向与代数方向来考虑。
  - 几何方向:P进行的工作是把任意的b投影到a上，因此乘以一个P已经把b投影到a上，再乘以一个P,就是再投影一次，事实上已经没有作用了。
  - 代数方向:$P^2 = aa^T/a^Ta \times aa^T/a^Ta = a(a^Ta)a^T/(a^Ta)^2 = aa^T/a^Ta = P$

Subsection3.复杂的三维投影



如上图,a_1,a_2 为构成平面的一组基,p 为 b 在平面上的投影,故p处在平面上。于是有: $p = \hat{x_1}a_1 + \hat{x_2}a_2$,那么我们利用矩阵形式写为$p = A\hat{x}$,其中$A = [a_1 \ \ a_2]$,$x = [x_1 \ \ x_2]^T$。这里平面就是矩阵$A$的列空间,由于$b$不在平面上,所以$Ax = b$无解,那么$A\hat{x} = p$中的$\hat{x}$就是我们想要的最优解。

由图可知,$e = b - p = b - Ax$垂直于平面A,那么容易得到两个方程:$\begin{cases} a_1^T(b - Ax) = 0 \\ a_2^T(b - Ax) = 0 \end{cases}$,将方程写成矩阵形式有$\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}(b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,即 $A^T(b - A\hat{x}) = 0$。

该式与向量上的投影方程$a^T(b - xa) = 0$很相似,对于向量来说,矩阵A只有一列,也即一个小写的a,本质都是$A^Te = 0$。由$A^Te = 0$可知所有可能的e对应着A的左零空间,从图像可知e垂直于平面(A的列空间),由此例我们很直观地看到了左零空间与列空间的正交。

再化简方程可得$A^TA\hat{x} = A^Tb$,这个式子恰好就是我们上节课所介绍的式子,至此,我们应该彻底明白了乘以A^T的意义何在。

最核心的式子就是:

$$A^TAX=A^Tb$$

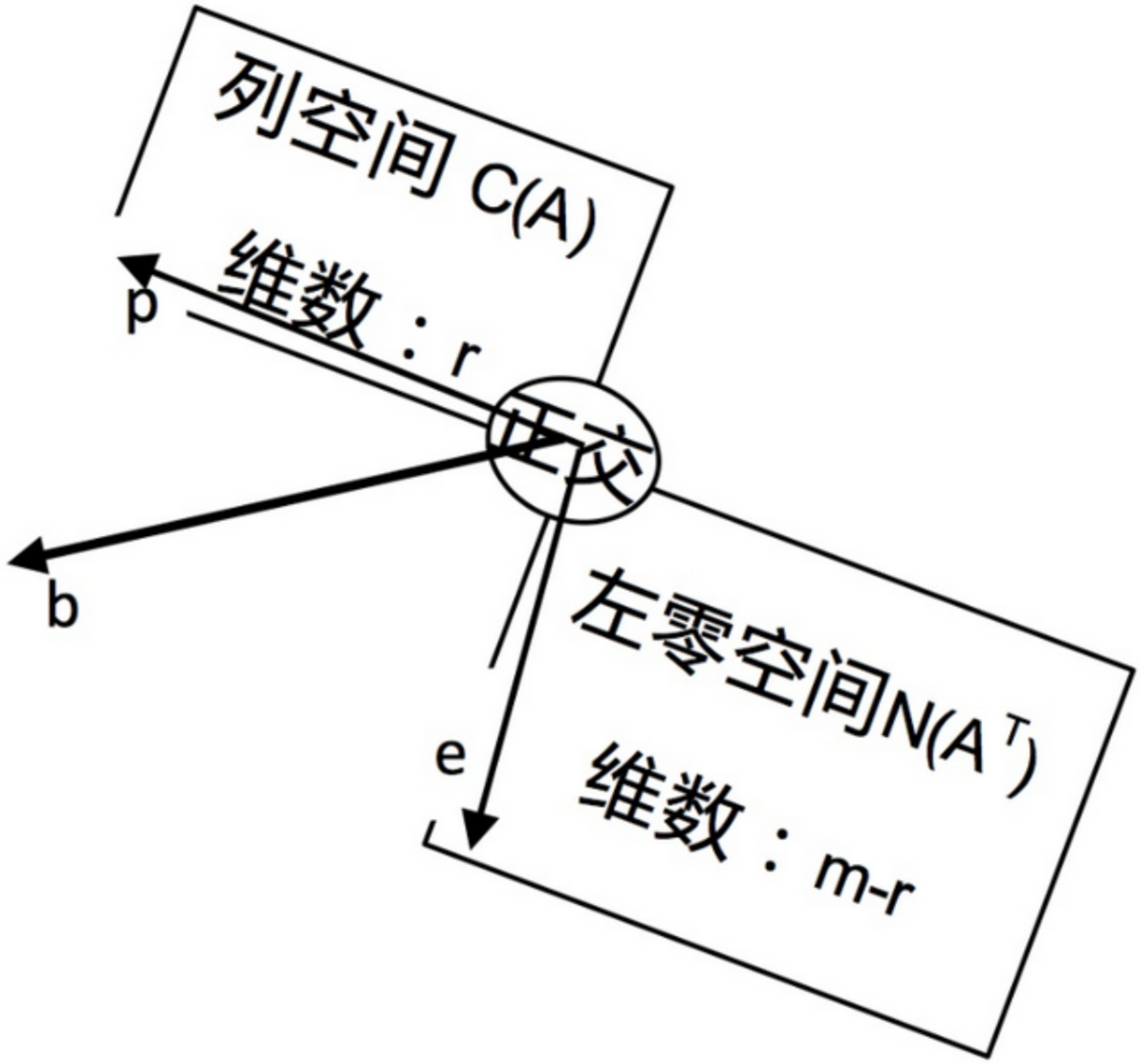
现在,我们需要再次考虑:最优解$\hat{x}$、投影p和投影矩阵P:

- $\hat{x}=(A^TA)^{-1}A^Tb$。可以看出来A^TA需要是可逆的才能求出最优解(这个说法并不正确,$\hat{x}$始终有解)。由N(A^TA)=N(A)我们可以知道,当A的各列线性无关的时候,A^TA才是可逆的。
- 从$x=A^{-1}b$到$\hat{x}=(A^TA)^{-1}A^Tb$,我们可以看到,x有解的条件是A是可逆的,而$\hat{x}$有解的条件是A的各列是线性无关的,其中对A的要求是有放宽的,因此,在$Ax=b$无解的情况下尝试$A^TA\hat{x}=A^Tb$以获得最优解的思路是可行的正确的。
- *$p=A\hat{x}=A(A^TA)^{-1}A^Tb$。回忆在二维空间的情况下,$A(A^TA)^{-1}A^T$对应着原来的$\frac{aa^T}{a^Ta}$,因为向量情况下的$a^Ta$是常数，所以能够放到分母处
- $P=A(A^TA)^{-1}A^T$。
  - P不是任何时候都存在，显然由N(A^TA)=N(A)我们可以知道,当A的各列线性无关的时候,A^TA才是可逆的。
  - 这个矩阵能否做$P=A(A^TA)^{-1}A^T=AA^{-1}(A^T)^{-1}A^T=I$的化简？显然是否定的，因为$A$都不一定是方阵，不存在可逆性，但是如果是可逆方阵，那么说明A本身是一个满秩的矩阵，那么他的列空间就是一个n维的满空间，这样的话b一定就会在列空间中，那么映射就是它本身，所以投影矩阵是I也是正确的。

Subsection2.两个极端的例子帮助加深对投影矩阵的理解

- 如果b在矩阵A的列空间里,那么$Pb=b$
  - 代数角度:如果b在矩阵A的列空间里,那么必然有$Ax=b$，$Pb=A(A^TA)^{-1}A^Tb=A(A^TA)^{-1}ATAx=AIx=Ax=b$
  - 几何角度:b在A的列空间中，那么显然b往A列空间投影的向量就是自己本身。
- 如果 b垂直于矩阵A的列空间,那么$Pb=0$
  - 代数角度:b垂直于A的列空间,根据正交补的概念可知b是左零空间中的向量$\therefore A^Tb=0\therefore Pb=A(A^TA)^{-1}A^Tb=A(A^TA)^{-1}0=0$
  - 几何角度:显然的,b垂直于A的列空间，那么投影到A的列空间就是零向量。

通过上面两个极端的例子,我们可以看出来,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。



一个问题的解决与探讨

这个问题我花了很长时间去想与理清，现在有一点点感触:

问题1.$A^TAx=A^Tb$是否一定有解?

答案是肯定的，因为

$$rank(A)=rank(A^TA)\le rank(A^TA,A^Tb)=rank(A^T(A,b))\le rank(A^T)$$

因此说明

$$rank(A^TA,A^Tb)=rank(A^TA)$$

因此说明其一定有解。

问题2.投影矩阵$P=A(A^TA)^{-1}A^T$是否一定存在?

显然是不一定的，我们提到过$r(A^TA)=r(A)$,如果A不满秩，那么显然A^TA是不存在逆矩阵的。
但是这并不意味着b在$C(A)$中没有投影，我们仍然可以解出$\hat{x}$，$A\hat{x}$就是$\hat{b}$，只不过b与$\hat{b}$中的关系不能再用P来描述了。

那么我们考虑问题1衍生的一个问题，如果$A^TAx=A^Tb$一定要A^TA可逆才能得到x的解吗？(就是x的表示)，显然并不是，哪怕我们在做$Ax=b$的解的时候，A也可以线性相关方程却仍然有解，所以我认为鲁莽的用逆矩阵来表示，如果逆矩阵不存在解就不存在这种想法是错误的。