第三十四讲：左右逆和伪逆

前面我们涉及到的逆（inverse）都是指左、右乘均成立的逆矩阵，即$A^{-1}A=I=AA^{-1}$。在这种情况下，$m\times n$矩阵A满足$m=n=rank(A)$，也就是满秩方阵。

左逆（left inserve）

记得我们在最小二乘一讲（第十六讲）介绍过列满秩的情况，也就是列向量线性无关，但行向量通常不是线性无关的。常见的列满秩矩阵A满足$m>n=rank(A)$。

列满秩时，列向量线性无关，所以其零空间中只有零解，方程$Ax=b$可能有一个唯一解（b在A的列空间中，此特解就是全部解，因为通常的特解可以通过零空间中的向量扩展出一组解集，而此时零空间只有列向量），也可能无解（b不在A的列空间中）。

另外，此时行空间为$\mathbb{R}^n$，也正印证了与行空间互为正交补的零空间中只有列向量。

现在来观察A^TA，也就是在$m>n=rank(A)$的情况下，$n\times m$矩阵乘以$m\times n$矩阵，结果为一个满秩的$n\times n$矩阵，所以A^TA是一个可逆矩阵。也就是说$\underbrace{(A^TA)^{-1}A^T}A=I$成立，而大括号部分的$(A^TA)^{-1}A^T$称为长方形矩阵$A$的左逆

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

顺便复习一下最小二乘一讲，通过关键方程$A^TA\hat{x}=A^Tb$，A^{-1}_{left}被当做一个系数矩阵乘在b向量上，求得b向量投影在A的列空间之后的解$\hat{x}=(A^TA)^{-1}A^Tb$。如果我们强行给左逆左乘矩阵A，得到的矩阵就是投影矩阵$P=A(A^TA)^{-1}A^T$，来自$p=A\hat{x}=A(A^TA)^{-1}A^T$，它将右乘的向量$b$投影在矩阵$A$的列空间中。

再来观察AA^T矩阵，这是一个$m\times m$矩阵，秩为$rank(AA^T)=n<m$，也就是说AA^T是不可逆的，那么接下来我们看看右逆。

右逆（right inverse）

可以与左逆对称的看，右逆也就是研究$m \times n$矩阵A行满秩的情况，此时$n > m = rank(A)$。对称的，其左零空间中仅有零向量，即没有行向量的线性组合能够得到零向量。

行满秩时，矩阵的列空间将充满向量空间$C(A) = \mathrm{R}^m$，所以方程$Ax = b$总是有解集，由于消元后有$n - m$个自由变量，所以方程的零空间为$n - m$维。

与左逆对称，再来观察AA^T，在$n > m = rank(A)$的情况下，$m \times n$矩阵乘以$n \times m$矩阵，结果为一个满秩的$m \times m$矩阵，所以此时AA^T是一个满秩矩阵，也就是AA^T可逆。所以$A\underbrace{A^T\left(AA^T\right)} = I$，大括号部分的$A^T\left(AA^T\right)$称为长方形矩阵的右逆

$$A_{right}^{-1} = A^T\left(AA^T\right)$$

同样的，如果我们强行给右逆右乘矩阵A，将得到另一个投影矩阵$P = A^T\left(AA^T\right)A$，与上一个投影矩阵不同的是，这个矩阵的$A$全部变为$A^T$了。所以这是一个能够将右乘的向量$b$投影在$A$的行空间中。

前面我们提及了逆（方阵满秩），并讨论了左逆（矩阵列满秩）、右逆（矩阵行满秩），现在看一下第四种情况，$m \times n$矩阵A不满秩的情况。

伪逆（pseudo inverse）

有$m \times n$矩阵A，满足$rank(A) < min(m,\ n)$，则

- 列空间$C(A) \in \mathrm{R}^m,\ \dim C(A) = r$，左零空间$N\left(A^T\right) \in \mathrm{R}^m,\ \dim N\left(A^T\right) = m - r$，列空间与左零空间互为正交补；
- 行空间$C\left(A^T\right) \in \mathrm{R}^n,\ \dim C\left(A^T\right) = r$，零空间$N(A) \in \mathrm{R}^n,\ \dim N(A) = n - r$，行空间与零空间互为正交补。

现在任取一个向量x，乘上A后结果Ax一定落在矩阵A的列空间$C(A)$中。而根据维数，$x \in \mathrm{R}^n,\ Ax \in \mathrm{R}^m$，那么我们现在猜测，输入向量$x$全部来自矩阵的行空间，而输出向量$Ax$全部来自矩阵的列空间，并且是一一对应的关系，也就是R^n的r维子空间到R^m的r维子空间的映射。

而矩阵A现在有这些零空间存在，其作用是将某些向量变为零向量，这样R^n空间的所有向量都包含在行空间与零空间中，所有向量都能由行空间的分量和零空间的分量构成，变换将零空间的分量消除。但如果我们只看行空间中的向量，那就全部变换到列空间中了。

那么，我们现在只看行空间与列空间，在行空间中任取两个向量$x,\ y \in C(A^T)$，则有$Ax \neq Ay$。所以从行空间到列空间，变换A是个不错的映射，如果限制在这两个空间上，A可以说“是个可逆矩阵”，那么它的逆就称作伪逆，而这个伪逆的作用就是将列空间的向量一一映射到行空间中。通常，伪逆记作A^+，因此$Ax = (Ax),\ y = A^+(Ay)$。

现在我们来证明对于$x,y \in C\left(A^T\right),\ x \neq y$，有$Ax, Ay \in C(A),\ Ax \neq Ay$：

- 反证法，设$Ax = Ay$，则有$A(x-y) = 0$，即向量$x-y \in N(A)$；
- 另一方面，向量$x,y \in C\left(A^T\right)$，所以两者之差$x-y$向量也在$C\left(A^T\right)$中，即$x-y \in C\left(A^T\right)$；
- 此时满足这两个结论要求的仅有一个向量，即零向量同时属于这两个正交的向量空间，从而得到$x=y$，与题设中的条件矛盾，得证。

伪逆在统计学中非常有用，以前我们做最小二乘需要矩阵列满秩这一条件，只有矩阵列满秩才能保证A^TA是可逆矩阵，而统计中经常出现重复测试，会导致列向量线性相关，在这种情况下A^TA就成了奇异矩阵，这时候就需要伪逆。

接下来我们介绍如何计算伪逆A^+：

其中一种方法是使用奇异值分解，$A = U\Sigma V^T$，其中的对角矩阵型为$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_2 & \\ \hline & & & [0] \end{array}\right]$，对角线非零的部分来自$A^TA,\ AA^T$比较好的部分，剩下的来自左/零空间。

我们先来看一下Σ矩阵的伪逆是多少，这是一个$m \times n$矩阵，$rank(\Sigma) = r$，$\Sigma^+ = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & \\ & & \frac{1}{\sigma_r} & \\ \hline & & & [0] \end{array}\right]$，伪逆与原矩阵有个小区别：这是一个$n \times m$矩阵。则有

$$\Sigma\Sigma^+ = \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline & & & [0] \end{array}\right]_{m\times m},\ \Sigma^+\Sigma = \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline & & & [0] \end{array}\right]_{n\times n}。$$

观察$\Sigma\Sigma^+$和$\Sigma^+\Sigma$不难发现，$\Sigma\Sigma^+$是将向量投影到列空间上的投影矩阵，而$\Sigma^+\Sigma$是将向量投影到行空间上的投影矩阵。我们不论是左乘还是右乘伪逆，得到的不是单位矩阵，而是投影矩阵，该投影将向量带入比较好的空间（行空间和列空间，而不是左/零空间）。

接下来我们来求A的伪逆：

$$A^+ = V \Sigma^+ U^T$$