- 第三十讲：奇异值分解

第三十讲：奇异值分解

本讲我们介绍将一个矩阵写为$A=U\Sigma V^T$，分解的因子分别为正交矩阵、对角矩阵、正交矩阵，与前面几讲的分解不同的是，这两个正交矩阵通常是不同的，而且这个式子可以对任意矩阵使用，不仅限于方阵、可对角化的方阵等。

- 在正定一讲中（第二十八讲）我们知道一个正定矩阵可以分解为$A=Q\Lambda Q^T$的形式，由于A对称性其特征向量是正交的，且其Λ矩阵中的元素皆为正，这就是正定矩阵的奇异值分解。在这种特殊的分解中，我们只需要一个正交矩阵Q就可以使等式成立。
- 在对角化一讲中（第二十二讲），我们知道可对角化的矩阵能够分解为$A=S\Lambda S^T$的形式，其中S的列向量由A的特征向量组成，但S并不是正交矩阵，所以这不是我们希望得到的奇异值分解。

我们现在要做的是，在A的**列空间**中找到一组特殊的正交基$v_1,v_2,\cdots,v_r$，这组基在A的作用下可以转换为A的**行空间**中的一组正交基$u_1,u_2,\cdots,u_r$。

用矩阵语言描述为$A\left[v_1\ v_2\ \cdots\ v_r\right]=\left[\sigma_1u_1\ \sigma_2u_2\ \cdots\ \sigma_ru_r\right]=\left[u_1\ u_2\ \cdots\ u_r\right]\begin{bmatrix}\sigma_1&&&\\&\sigma_2&&\\&&\ddots&\\&&&\sigma_n\end{bmatrix}$，即$Av_1=\sigma_1u_1$，$Av_2=\sigma_2u_2,\cdots,Av_r=\sigma_ru_r$，这些$\sigma$是缩放因子，表示在转换过程中有拉伸或压缩。而$A$的左零空间和零空间将体现在$\sigma$的零值中。

另外，如果算上左零、零空间，我们同样可以对左零、零空间取标准正交基，然后写为$A\left[v_1\ v_2\ \cdots\ v_r\ v_{r+1}\ \cdots\ v_m\right]=\left[u_1\ u_2\ \cdots\ u_r\ u_{r+1}\ \cdots\ u_n\right]\left[\begin{array}{ccc|c}\sigma_1&&&\\&\ddots&&\\&&\sigma_r&\\\hline&&&[0]\end{array}\right]$，此时$U$是$m\times m$正交矩阵，$\Sigma$是$m\times n$对角矩阵，$V^T$是$n\times n$正交矩阵。

最终可以写为$AV=U\Sigma$，可以看出这十分类似对角化的公式，矩阵A被转化为对角矩阵Σ，我们也注意到U，V是两组不同的正交基。（在正定的情况下，U，V都变成了Q。）。进一步可以写作$A=U\Sigma V^{-1}$，因为V是标准正交矩阵所以可以写为$A=U\Sigma V^T$

计算一个例子，$A=\begin{bmatrix}4&4\\-3&3\end{bmatrix}$，我们需要找到：

- 行空间$\mathbb{R}^2$的标准正交基v_1,v_2；
- 列空间$\mathbb{R}^2$的标准正交基u_1,u_2；
- $\sigma_1>0,\sigma_2>0$。

在$A=U\Sigma V^T$中有两个标准正交矩阵需要求解，我们希望一次只解一个，如何先将U消去来求V？

这个技巧会经常出现在长方形矩阵中：求A^TA，这是一个对称正定矩阵（至少是半正定矩阵），于是有$A^TA=V\Sigma^TU^TU\Sigma V^T$，由于$U$是标准正交矩阵，所以$U^TU=I$，而$\Sigma^T\Sigma$是对角线元素为$\sigma^2$的对角矩阵。

现在有$A^TA=V\begin{bmatrix}\sigma_1&&&\\&\sigma_2&&\\&&\ddots&\\&&&\sigma_n\end{bmatrix}V^T$，这个式子中$V$即是$A^TA$的特征向量矩阵而$\Sigma^2$是其特征值矩阵。

同理，我们只想求U时，用AA^T消掉V即可。

我们来计算$A^TA=\begin{bmatrix}4&-3\\4&3\end{bmatrix}\begin{bmatrix}4&4\\-3&3\end{bmatrix}=\begin{bmatrix}25&7\\7&25\end{bmatrix}$，对于简单的矩阵可以直接观察得到特征向量$A^TA\begin{bmatrix}1\\1\end{bmatrix}=32\begin{bmatrix}1\\1\end{bmatrix}$，$A^TA\begin{bmatrix}1\\-1\end{bmatrix}=18\begin{bmatrix}1\\-1\end{bmatrix}$，化为单位向量有$\sigma_1=32,\ v_1=\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}$，$\sigma_2=18,\ v_2=\begin{bmatrix}\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}\end{bmatrix}$。

到目前为止，我们得到$\begin{bmatrix}4&4\\-3&3\end{bmatrix}=\begin{bmatrix}u_?&u_?\\u_?&u_?\end{bmatrix}\begin{bmatrix}\sqrt{32}&0\\0&\sqrt{18}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}$，接下来继续求解$U$。

$AA^T=U\Sigma V^TV\Sigma^TU^T=U\Sigma^2U^T$，求出$AA^T$的特征向量即可得到$U$，$\begin{bmatrix}4&4\\-3&3\end{bmatrix}\begin{bmatrix}4&-3\\4&3\end{bmatrix}=\begin{bmatrix}32&0\\0&18\end{bmatrix}$，观察得$AA^T\begin{bmatrix}1\\0\end{bmatrix}=32\begin{bmatrix}1\\0\end{bmatrix}$，$AA^T\begin{bmatrix}0\\1\end{bmatrix}=18\begin{bmatrix}0\\1\end{bmatrix}$。但是我们不能直接使用这一组特征向量，因为式子$AV=U\Sigma$明确告诉我们，一旦$V$确定下来，$U$也必须取能够满足该式的向量，所以此处$Av_2=\begin{bmatrix}0\\-\sqrt{18}\end{bmatrix}=u_2\sigma_2=\begin{bmatrix}0\\-1\end{bmatrix}\sqrt{18}$，则$u_1=\begin{bmatrix}1\\0\end{bmatrix}$，$u_2=\begin{bmatrix}0\\-1\end{bmatrix}$。（这个问题在本讲的官方笔记中有详细说明。）

- 补充：AB的特征值与BA的特征值相同，证明来自Are the eigenvalues of AB equal to the eigenvalues of BA? (Citation needed!)：

  取$\lambda \neq 0$，v是AB在特征值取λ时的的特征向量，则有$Bv \neq 0$，并有$\lambda Bv = B(\lambda v) = B(ABv) = (BA)Bv$，所以$Bv$是$BA$在特征值取同一个$\lambda$时的特征向量。

  再取AB的特征值$\lambda = 0$，则$0 = \det AB = \det A \det B = \det BA$，所以$\lambda = 0$也是$BA$的特征值，得证。

最终，我们得到$\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$。

再做一个例子，$A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$，这是个秩一矩阵，有零空间。$A$的行空间为$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$的倍数，$A$的列空间为$\begin{bmatrix} 4 \\ 8 \end{bmatrix}$的倍数。

- 标准化向量得$v_1 = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$，$u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$。
- $A^TA = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}$，由于$A$是秩一矩阵，则$A^TA$也不满秩，所以必有特征值0，则另特征值一个由迹可知为125。
- 继续求零空间的特征向量，有$v_2 = \begin{bmatrix} 0.6 \\ -0,8 \end{bmatrix}$，$u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

最终得到$\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 1 & \underline{2} \\ 2 & \underline{-1} \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & \underline{0} \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ \underline{0.6} & \underline{-0.8} \end{bmatrix}$，其中下划线部分都是与零空间相关的部分。

- $v_1, \cdots, v_r$是行空间的标准正交基；
- $u_1, \cdots, u_r$是列空间的标准正交基；
- $v_{r+1}, \cdots, v_n$是零空间的标准正交基；
- $u_{r+1}, \cdots, u_m$是左零空间的标准正交基。

通过将矩阵写为$Av_i = \sigma_i u_i$形式，将矩阵对角化，向量u, v之间没有耦合，A乘以每个v都能得到一个相应的u。