Chapter28奇异值分解

Section1.引入

我们已经学过两种不错的分解方式了:

- 在正定一讲中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，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。

本讲我们介绍将一个矩阵写为$A=U\Sigma V^T$,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。

其中U是正交矩阵，描述的是A的列空间；Σ是对角矩阵；V是正交矩阵，描述的是A的行空间。

Section2.证明

我们首先说明，$v_1,v_2\cdots v_n$既是A^TA的特征向量，也是A的行空间的基:

$$A^T(Av)=\lambda v\rightarrow v\text{在}A^T\text{的列空间中，则}v\text{在}A\text{的行空间中}$$

接下来说明v能够描述A的整个行空间:

$$\begin{gathered}A\text{的零空间为}n-r\text{维，又因为}Av=0\text{实际上是}\lambda=0\text{时候的情况}\\ \therefore\lambda=0\text{时候的特征向量为}n-r\text{维，因此}\lambda\neq0\text{的特征向量为}n-(n-r)=r\text{维}\\ \text{又}\because\text{行空间为}r\text{维，故}v_n\text{可以描述整个行空间}\end{gathered}$$

这里注意到A^TA是实对称矩阵，那么v是特征向量是正交的。

我们令

$$Av_n=u_n$$

，因此u_n位于A的列空间中，又注意到u_n之间的正交性:

$$u_i^Tu_j=(Av_i)^TAv_j=\lambda_i\lambda_jv_i^Tv_j=0$$

又列空间为r维的，因此u_n是列空间的一组正交基。

但是我们要注意到，u_n不是标准正交向量，因为模可能不等于1，因此我们引入$Av_1=\sigma_1 u_1$，$Av_2=\sigma_2 u_2,\cdots,Av_r=\sigma_r u_r$,些$\sigma$是缩放因子，表示在转换过程中有拉伸或压缩。

当然，此时我们只描述了r维的情况，但是A是$n\times n$的矩阵，因此我们考虑算上左零、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

$$A\begin{bmatrix} v_1 & v_2 & \cdots & v_r & v_{r+1} & \cdots & v_m \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_r & u_{r+1} & \cdots & u_n \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & [0] \end{array}\right]$$

，此时U是$m\times m$正交矩阵，Σ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$A=U\Sigma V^T$。

我们通过证明可以知道

- 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。

Section3.实例

$A=\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$，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；
- $\sigma_1>0,\sigma_2>0$。

我们从证明过程可以知道，$Av=\sigma u$，因此我们做奇异值分解的时候，我们关注求解一个正交矩阵，另外一个就可以推出来:

那我们注意到:

$$A^TA=V\Sigma^TU^TU\Sigma V^T=V\Sigma^2V^T$$

即：

$$A^TA=V\begin{bmatrix}\sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2\end{bmatrix}V^T$$

这个式子中V即是A^TA的特征向量矩阵而Σ^2是其特征值矩阵。

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$\sigma_1=32,\ v_1=\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix},\ \sigma_2=18,\ v_2=\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}$。

我们求解u的时候，需要根据v来求，我们在证明的时候就提到过$Av=\sigma u$,因此我们考虑

$$Av_2=\begin{bmatrix}0\\ -\sqrt{18}\end{bmatrix}=u_2\sigma_2=\begin{bmatrix}0\\ -1\end{bmatrix}\sqrt{18}$$

则

$$u_1=\begin{bmatrix}1\\ 0\end{bmatrix},\ u_2=\begin{bmatrix}0\\ -1\end{bmatrix}$$

AB和BA特征值相同的证明

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。

Section4.另一个性质不同的实例

上一个例子中，我们的特征值没有0，因此考虑另一个例子。

$A=\begin{bmatrix}4&3\\8&6\end{bmatrix}$，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。

同样的步骤：
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$v_2=\begin{bmatrix}0.6\\-0,8\end{bmatrix}$, $u_2=\frac{1}{\sqrt{5}}\begin{bmatrix}2\\-1\end{bmatrix}$

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。

通过将矩阵写为$Av_i=\sigma_iu_i$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。