SVD分解

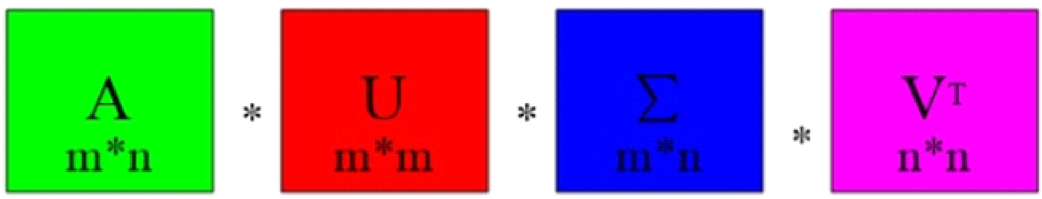
2018年12月26日 星期三　　上午10:56

特征值分解最大的问题是只能针对方阵，即n*n的矩阵。而在实际的应用中，我们分解的大部分都不是方阵。

奇异值分解是一个能适用于任意矩阵的一种分解的方法，对于任意矩阵A总是存在一个奇异值分解：

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

假设A是一个m*n的矩阵，那么得到的U是一个m*m的方阵，U里面的正交向量被称为**左奇异向量**。Σ是一个m*n的矩阵，Σ除了对角线其它元素都为0，对角线上的元素称为奇异值。V^T是v的转置矩阵，是一个n*n的矩阵，它里面的正交向量被称为**右奇异值向量**。而且一般来讲，我们会将Σ上的值按从大到小的顺序排列。



虽说上面奇异值分解等式成立，但是**如何求得左奇异向量、右奇异向量和奇异值**呢？

答案：由上面的奇异值分解等式，我们是不知道如何拆分矩阵A的。我们可以把奇异值和特征值联系起来。

首先，我们用矩阵A的转置乘以A，得到一个方阵，用这样的方阵进行特征分解，得到的特征值和特征向量满足下面的等式：

$$(A^T A) v_i = \lambda_i v_i$$

这里的v_i就是我们要求的**右奇异向量**。

其次，我们将A和A^T做矩阵的乘法，得到一个方阵，用这样的方阵进行特征分解，得到的特征和

特征向量满足下面的等式：

$$(AA^T)u_i = \lambda_i u_i$$

这里的u_i就是**左奇异向量。**

上面我们说A^TA的特征向量组成的矩阵就是我们SVD中的V矩阵，而AA^T的特征向量组成的就是我们SVD中的U矩阵，这有什么根据么?我们来证明一下，以V矩阵的证明为例。

$$A = U\Sigma V^T \Rightarrow A^T = V\Sigma^T U^T \Rightarrow A^TA = V\Sigma^T U^T U\Sigma V^T = V\Sigma^2 V^T (*)$$

上式证明中使用了$U^TU = I$，$\Sigma^T\Sigma = \Sigma^2$。可以看出，$A^TA$的特征向量组成的矩阵就是我们SVD中的V矩阵，而$AA^T$的特征向量组成的就是我们SVD中的U矩阵。

补充定义：

$$U \epsilon M_n(R)\text{满足}U^TU = I\text{，则}U\text{是实正交矩阵。}$$

此外，我们还可以得到奇异值，奇异值求法有两种：

a）第一种：

$$A = U\Sigma V^T \Rightarrow AV = U\Sigma V^T V \Rightarrow AV = U\Sigma \Rightarrow Av_i = \sigma_i u_i \Rightarrow \sigma_i = \frac{Av_i}{u_i}$$

b)第二种：

通过上面*式的证明，我们还可以看出，特征值矩阵等于奇异值矩阵的平方，也就是说特征值和奇异值满足如下关系：

$$\sigma_i = \sqrt{\lambda_i}$$

这里的σ_i就是奇异值，奇异值σ_i跟特征值类似，在矩阵Σ中也是从大到小排列。

在奇异值分解矩阵中Σ里面的奇异值按从大到小的顺序排列，奇异值
σ_i
从大到小的顺序减小的特别快。**在很多情况下，前10%甚至1%的奇异值的和就占了全部的奇异值之和的99%以上。也就是说，剩下的90%甚至99%的奇异值几乎没有什么作用。**

因此，我们可以用前面r个大的奇异值来近似描述矩阵，于是奇异值分解公式可以写成如下：

$$A_{m\times n} \approx U_{m\times r} \Sigma_{r\times r} V^T_{r\times n}$$

其中r是一个远远小于m和n的数，右边的三个矩阵相乘的结果将会使一个接近A的矩阵。如果r越接近于n，则相乘的结果越接近于A。如果r的取值远远小于n，从计算机内存的角度来说，右边三个矩阵的存储内存要远远小于矩阵A的。**所以在奇异值分解中r的取值很重要，就是在计算精度和时间空间之间做选择。**

SVD计算举例：

这里我们用一个简单的矩阵来说明奇异值分解的步骤。我们的矩阵A定义为：

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

首先，我们先求出 A^TA 和 AA^T：

$$A^TA = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

然后，求出 A^TA和AA^T 的特征值和特征向量：

A^TA 的特征值和特征向量：

$$\lambda_1=3;\quad v_1=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix};\quad \lambda_2=1;\quad v_2=\begin{pmatrix}\frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix};$$

AA^T 的特征值和特征向量：

$$\lambda_1=3;\quad u_1=\begin{pmatrix}\frac{1}{\sqrt{6}}\\ \frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{6}}\end{pmatrix};\quad \lambda_2=1;\quad u_2=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ 0\\ \frac{-1}{\sqrt{2}}\end{pmatrix};\quad \lambda_3=0;\quad u_3=\begin{pmatrix}\frac{1}{\sqrt{3}}\\ \frac{-1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{pmatrix};$$

其次，我们利用 $Av_i=\sigma_i u_i$， $i=1,2$,求奇异值：

$$\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}=\sigma_1\begin{pmatrix}\frac{1}{\sqrt{6}}\\ \frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{6}}\end{pmatrix}\Rightarrow\sigma_1=\sqrt{3}$$

$$\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}\begin{pmatrix}\frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}=\sigma_2\begin{pmatrix}\frac{1}{\sqrt{2}}\\ 0\\ \frac{-1}{\sqrt{2}}\end{pmatrix}\Rightarrow\sigma_2=1$$

当然，这一步也可以用 $\sigma_i=\sqrt{\lambda_i}$ 直接求出奇异值为 $\sqrt{3}$ 和1。

最后，我们得到A的奇异值分解为：

$$A=U\Sigma V^T=\begin{pmatrix}\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{6}}&0&\frac{-1}{\sqrt{3}}\\ \frac{1}{\sqrt{6}}&\frac{-1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\end{pmatrix}\begin{pmatrix}\sqrt{3}&0\\0&1\\0&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}$$

本文来源于：
https://mp.weixin.qq.com/s/Dv51K8JETakIKe5dPBAPVg