APR 29TH, 2014

LAS 3-EigenVectors and Eigenvalues

线性代数那些事 Things of Linear Algebra
逸夫图书馆, 2014/4/27

特征值和特征向量

好，我们知道了矩阵就是线性变换，那么矩阵的特征值和特征向量是什么？

wiki在线性代数中，对于一个给定的线性变换，它的特征向量（本征向量或称正规正交向量）v经过这个线性变换之后，得到的新向量仍然与原来的v 保持在同一条直线上，但其长度也许会改变。一个特征向量的长度在该线性变换下缩放的比例称为其特征值（本征值）。如果特征值为正，则表示v 在经过线性变换的作用后方向也不变；如果特征值为负，说明方向会反转；如果特征值为0，则是表示缩回零点。但无论怎样，仍在同一条直线上。

简而言之就是说，对于一个确定的矩阵，如果它的特征向量存在，那么就有下面的现象，当这个矩阵(即这个线性变换)作用在这些特征向量中的任意一个上的时候，得到的向量和原来的特征向量在同一条直线上，只是长度发生了变化，长度的变化量的比例为该特征向量对应的特征值。从这里可以看出，这些特征向量是对这个矩阵的很好的描述！

用《蒙娜丽莎》来理解下：

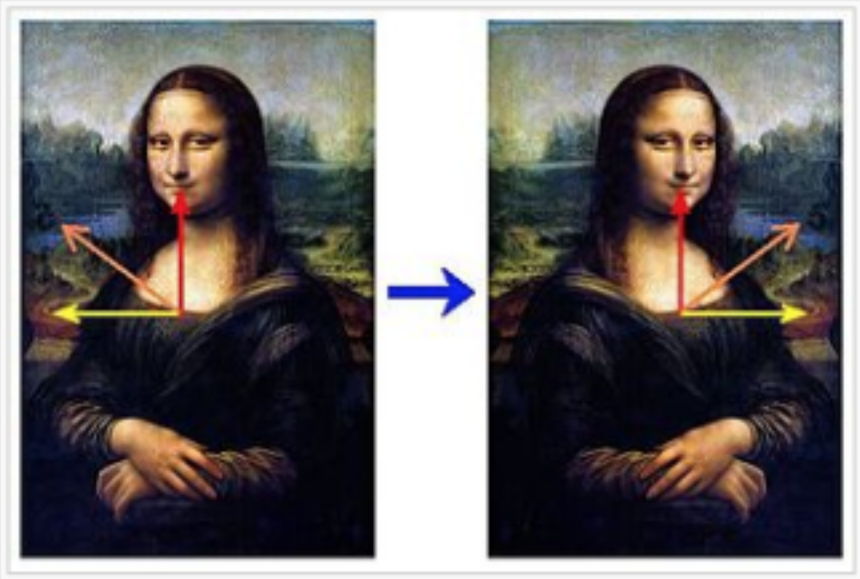


图1.当蒙娜丽莎的图像左右翻转时，中间垂直的红色向量方向保持不变。而水平方向上黄色的向量的方向完全反转，因此它们都是左右翻转变换的特征向量。红色向量长度不变，其特征值为1。黄色向量长度也不变但方向变了，其特征值为-1。橙色向量在翻转后和原来的向量不在同一条直线上，因此不是特征向量。

根据wiki上对特征向量的定义，首先要明确的是这个线性变换(也就是这个矩阵)是向量空间E到自身的一个线性变换，它可以是旋转、反射、拉伸、压缩或者这些变换的组合等等，本来呢，一个向量经过线性变换可以得到任何向量，但是，如果这个向量是这个线性变换的特征向量的话，经过线性变换得到的向量那就一定是和特征向量在同一条直线上！特征向量可能会有多个，特征值最大的特征向量称为主特征。所有具有相同的特征值λ的特征向量和零向量一起，组成了一个向量空间，称为线性变换的一个特征空间。

定义 [编辑]

给定一个向量空间$\mathbf{E}$，从$\mathbf{E}$到$\mathbf{E}$自身的线性变换$\mathbf{T}$是一个保持向量加法和标量乘法的函数，例如旋转、反射、拉伸、压缩，或者这些变换的组合等等[1]。一个线性变换可以通过它们在向量上的作用来可视化。一般来说，一个向量在经过映射之后可以变为任何可能的向量，而特征向量具有更好的性质[2]。

一个线性变换$\mathbf{T}:\mathbf{E}\to\mathbf{E}$的特征向量$v$是在这个线性变换下简单地乘以一个标量$\lambda$的非零向量[2][2]。也就是说$\lambda$满足：

$$\mathbf{T}(v)=\lambda v$$

其中的缩放因子称为这个特征向量的特征值，或者说是线性变换$\mathbf{T}$的特征值。反过来，一个实数λ是线性变换$\mathbf{T}$的一个特征值，当且仅当有一个非零向量v满足上面的式子[2][3]。

所有具有相同的特征值λ的特征向量和零向量一起，组成了一个向量空间，称为线性变换的一个特征空间，一般记作$\mathbf{E}_\lambda(\mathbf{T})$[4]。这个特征空间如果是有限维的，那么它的维数叫做λ的几何重数[5]。

变换的主特征向量是模最大的特征值对应的特征向量[6]。有限维向量空间上的一个变换的谱是其所有特征值的集合[7]。

特征向量也可以看作是关于系数λ的方程：

$$\mathbf{T}(x)=\lambda x$$

的非零解。显然只有在λ是变换$\mathbf{T}$的特征值之时，方程才有非零解[8]。

在一定条件下（例如实对称矩阵形式的线性变换），一个变换可以由其特征值和特征向量完全表述。一个特征空间是具有相同特征值的特征向量与一个同维数的零向量的集合，可以证明该集合是一个线性子空间。 [TODO：特征空间我还不太理解，若读者明白，请留言告知，谢谢]

一般来说，2×2的非奇异矩阵如果有两个相异的特征值，就有两个线性无关的特征向量。在这种情况下，对于特征向量，线性变换仅仅改变它们的长度，而不改变它们的方向（除了反转以外），而对于其它向量，长度和方向都可能被矩阵所改变。如果特征值的模大于1，特征向量的长度将被拉伸，而如果特征值的模小于1，特征向量的长度就将被压缩。如果特征值小于0，特征向量将会被翻转。

重复了这么多次，我想你也已经认可了什么是特征向量了，下面看看例子。

先看个恒等变换和对角矩阵，注意其中对特征向量和特征空间的分析。

线性变换 [编辑]

最简单的例子是恒等变换$\mathbf{I}$的特征向量。由于对所有的非零向量v，

$$\mathbf{I}(v)=v=1\cdot v$$

所以所有的非零向量都是恒等变换$\mathbf{I}$的特征向量，对应着特征值1。恒等变换的特征空间只有一个，就是整个空间，对应着特征值1。[9]类似地，数乘变换$\lambda\mathbf{I}$的特征向量也是所有非零向量，因为按照定义，对所有的非零向量v，

$$\lambda\mathbf{I}(v)=\lambda\cdot v$$

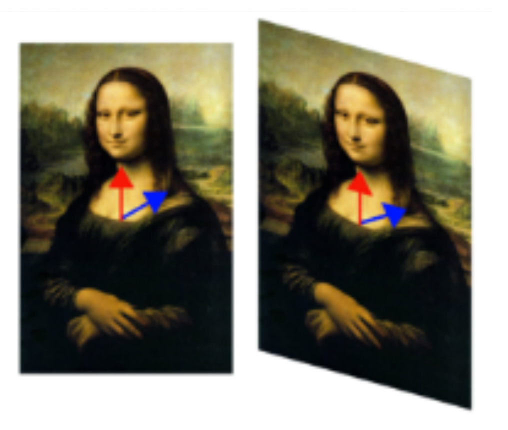
如果一个变换可以写成对角矩阵，那么它的特征值就是它对角线上的元素，而特征向量就是相应的基。例如矩阵：

$$\mathbf{A}=\begin{bmatrix}2&0&0\\0&2&0\\0&0&4\end{bmatrix}$$

的特征值就是2和4。2对应的特征向量是所有形同$(a,b,0)^T$的非零向量，而4对应的特征向量是所有形同$(0,0,c)^T$的非零向量。2对应的特征空间是一个2维空间，而4对应的特征空间是一个1维空间。矩阵$\mathbf{A}$的谱是$\{2,4\}$。

再看个实际的例子，错切变换，这里利用了矩阵行列式的知识来求解特征值。

对于更复杂的矩阵，特征向量和特征值就不是显然的了。右图中的例子是一个二维平面上的错切变换，其矩阵可以表示为：



$$\mathbf{A}=\begin{bmatrix}1&0\\-\frac{1}{2}&1\end{bmatrix}$$

$\mathbf{A}$的特征向量$\mathbf{x}$，按照定义，是在变换$\mathbf{A}$的作用下会得到$\mathbf{x}$自身的若干倍的非零向量。假设在$\mathbf{A}$的作用下$\mathbf{x}$变成了自身的λ倍，也就是

$$\mathbf{Ax}=\lambda\mathbf{x}$$

在等式两边的左侧乘以单位矩阵I，得到

$$\mathbf{IAx}=\mathbf{I}\cdot\lambda\mathbf{x}$$

$$\mathbf{Ax}=(\lambda I)\mathbf{x}$$

因此

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}=0$$

根据线性方程组理论，为了使这个方程有非零解，矩阵$\mathbf{A}-\lambda\mathbf{I}$的行列式必须是零：

$$\det(\mathbf{A}-\lambda\mathbf{I})=0$$

按照行列式的展开定义，上面式子的左端是一个关于λ的多项式，称为特征多项式。这个多项式的系数只和$\mathbf{A}$有关。在这个例子中，可以计算这个特征多项式：

$$\det\left(\begin{bmatrix}1&0\\-\frac{1}{2}&1\end{bmatrix}-\lambda\begin{bmatrix}1&0\\0&1\end{bmatrix}\right)=(1-\lambda)^2$$

在这种情况下特征多项式的方程变成$(1-\lambda)^2=0$。它的唯一的解是：$\lambda=1$。这就是矩阵$\mathbf{A}$的特征值。

找到特征值$\lambda=1$后，就可以找出

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}=0$$

的非零解，也就是特征向量了。在例子中：

$$\begin{bmatrix}1-\lambda&0\\-\frac{1}{2}&1-\lambda\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=0$$

将$\lambda=1$代入，就有

$$\begin{bmatrix}0&0\\-\frac{1}{2}&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=0$$

解这个新矩阵方程，得到如下形式的解：

$$\mathbf{x}=\begin{bmatrix}0\\c\end{bmatrix}$$

这里的 c 是任意非零常量。因此，矩阵 $\mathbf{A}$ 的特征向量就是所有竖直方向的向量（比如图中红色箭头代表的向量）。

特征值的代数重数和几何重数(TODO：后者我没有看懂，若读者明白，请留言告知，谢谢)

A的一个特征值λ的代数重数是λ作为A的特征多项式的根的次数；换句话说，若r是该多项式的一个根，它是一次多项式因子（λ - r）在特征多项式中在因式分解后中出现的次数。如果将代数重次计算在内的话，一个$n\times n$矩阵有n个特征值，因为其特征多项式次数为n。

一个代数重次1的特征值为“单特征值”。

在关于矩阵理论的条目中，可能会遇到如下的表示方法：

“一个矩阵A的特征值为4,4,3,3,3,2,2,1,”

表示4的代数重次为二，3的是三，2的是二，而1的是1。这样写是因为代数重次对于矩阵理论中的很多数学证明很重要而被大量使用。

和代数重数相对的是特征值的几何重数：特征值相对应的特征空间（也就是$\lambda I - A$的零空间）的维数。代数重次也可以视为一种维数：它是相应广义特征空间的维数，也就是当自然数k足够大的时候矩阵（$\lambda I - A$）^{k}的零空间。也就是说，它是所有“广义特征向量”组成的空间，其中一个广义特征向量是任何一个如果$\lambda I - A$作用连续作用足够多次就“最终”会变0的向量。任何特征向量都是一个广义特征向量，以此任一个特征空间都被包含于相应的广义特征空间。这给了一个几何重次总是小于或等于代数重次的简单证明。

例如：

$$A=\begin{bmatrix}1&1\\0&1\end{bmatrix}.$$

它只有一个特征值，也就是$\lambda = 1$。其特征多项式是$(\lambda-1)^2$，所以这个特征值代数重次为2。但是，相应特征空间是通常称为x轴的数轴，由向量$\begin{bmatrix}1\\0\end{bmatrix}$线性生成，所以几何重次只是1。

广义特征向量可以用于计算一个矩阵的若尔当标准型（参看下面的讨论）。若尔当块通常不是对角化而是幂零的这个事实与特征向量和广义特征向量之间的区别直接相关。

特征值的计算，简单的矩阵可以使用解特征方程的方法，但是一般情况下都是采用数值计算的方法，其中基于迭代技术的幂法可以用来计算矩阵的主特征值，反幂法类似，不过计算的是模最小的特征值，实际中常用的是QR分解。详情请参考请后面的矩阵分解部分。

数值计算 [编辑]

更多资料：特征值算法

在实践中，大型矩阵的特征值无法通过特征多项式计算。计算该多项式本身相当费资源，而根的精确表达式对于高次的多项式来说很难计算和表达：阿贝尔-鲁菲尼定理显示五次或更高次的多项式的根无法用n次方根来简单表达。对于估算多项式的根的有效算法是有的，但特征值中的微小误差可以导致特征向量的巨大误差。因此，寻找特征多项式和特征值的一般算法，是迭代法。最简单的方法是幂法：取一个随机向量v，然后计算如下的一系列单位向量

$$\frac{Av}{||Av||},\frac{A^2v}{||A^2v||},\frac{A^3v}{||A^3v||},\ldots$$

这个序列几乎总是收敛于最大绝对值的特征值所对应的特征向量。这个算法很简单，但是本身不是很有用。但是，象QR算法这样的算法正是以此为基础的[10]。

此外以豪斯霍尔德变换结合LU分解，可以得到比QR分解更快速的收敛特征值矩阵。[11]

到此，我觉得特征向量应该是清晰了，关于特征值和奇异值分解以及代码实现请参考后面的矩阵分解部分，也可以直接看我写的另一份总结Numerical Methods Using Matlab 第三章 矩阵特征值和奇异值分解。

矩阵特征值的应用特别广，例如因子分析，特征脸，PageRank等等算法都是基于特征值分解，若有时间和精力，我后续会一一介绍(有链接的是已经完成的部分)。

还想看看其他的介绍？这篇文章介绍的不错What are eigenvectors and eigenvalues?

Posted by hujiawei • Apr 29th, 2014 • math

« LAS 2-Matrix
LAS 4-Similarity Matrix »