30 特征值和奇异值

30.1 特征值和奇异值定义

在多元统计和时间序列分析中会用到特征值和奇异值, 比如，主成分分析、典型相关分析、对应分析、多元自回归模型等。

先简单回顾线性代数中特征值的定义和性质。 设A为n阶方阵， 若有非零向量$\boldsymbol{\alpha}$和复数λ使得

$$A\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha},$$

则称λ是矩阵A的一个**特征值**， $\boldsymbol{\alpha}$是特征值λ对应的**特征向量。** 特征向量具有某种不变性：矩阵A左乘特征向量， 不改变特征向量的方向（没有正反）。

当A为n阶实对称阵时，A恰有n个实特征值， 记作$\lambda_1, \lambda_2, \dots, \lambda_n$, 相应的特征向量也都是实向量, 且存在$n$阶正交阵$U$使得

$$A = U \Lambda U^T = \sum_{j=1}^n \lambda_j \boldsymbol{u}_{\cdot j} \boldsymbol{u}_{\cdot j}^T,$$

其中$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$是对角线元素为$A$的特征值的对角阵, U的第j列$\boldsymbol{u}_{\cdot j}$是λ_j对应的特征向量。

当A为非负定阵时，所有特征值非负； 当A为正定阵时，所有特征值都是正数。 A为非负定阵时， 令$\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_n^{1/2})$, 令$A^{1/2} = U \Lambda^{1/2} U^T$， 则

$$(A^{1/2})^2 = U \Lambda^{1/2} U^T U \Lambda^{1/2} U^T = U \Lambda^{1/2} \Lambda^{1/2} U^T = U \Lambda U^T = A.$$

如果A是对称阵且$\lambda_1, \dots, \lambda_n$中仅有$\lambda_1, \dots, \lambda_r$绝对值较大， 其余特征值接近于零， 则矩阵$A$有如下的近似：

$$A \sum_{j=1}^r \lambda_j \boldsymbol{u}_{\cdot j} \boldsymbol{u}_{\cdot j}^T,$$

在统计问题如典型相关分析中还会遇到广义的特征值问题: 设A, B为n阶方阵， 若有复数λ和非零向量$\boldsymbol{\alpha}$使得

Loading [MathJax]/extensions/MathEvents.js

$$A\boldsymbol{\alpha} = \lambda B\boldsymbol{\alpha},$$

则称λ和$\boldsymbol{\alpha}$分别为矩阵A相对于矩阵B的**广义特征值**和**广义特征向量**。 实际问题中， B通常是正定阵，A是实对称阵， 这时(30.3)等价于$B^{-1} A\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}$， 可以化为普通特征值问题，并可利用Cholesky分解进行计算。 设B有Cholesky分解$B = LL^T$, 则由$A\boldsymbol{\alpha} = \lambda LL^T\boldsymbol{\alpha}$得$L^{-1}A(L^T)^{-1}(L^T\boldsymbol{\alpha}) = \lambda(L^T\boldsymbol{\alpha})$, 求解普通特征值问题$\left(L^{-1}A(L^T)^{-1}\right)\boldsymbol{\beta} = \lambda\boldsymbol{\beta}$得$\lambda$和$\boldsymbol{\beta}$ 再求解$L^T\boldsymbol{\alpha} = \boldsymbol{\beta}$即可得广义特征值和广义特征向量。

对n阶非奇异矩阵A, 必存在n阶正交阵U和V， 使得

$$A = V\text{diag}(d_1, d_2, \dots, d_n)U^T,$$

其中$d_1, d_2, \dots, d_n$是正定阵A^TA的n个特征值的算数平方根, 称(30.4)为矩阵A的**奇异值分解**, $d_1, d_2, \dots, d_n$称为A的**奇异值**。

若A是一般的$n \times m$非零矩阵， A的秩为$\text{rank}(A) = r \leq \min(n, m)$, A^TA的非零特征值为$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$, 令$d_i = \sqrt{\lambda_i}$, $i = 1, 2, \dots, r$, 则称d_i为A的**奇异值**, 且一定有m阶正交阵U和n阶正交阵V使得

$$A = VDU^T,$$

其中

$$D = \begin{pmatrix} \text{diag}(d_1, d_2, \dots, d_r) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n\times m},$$

称(30.5)为A的**奇异值分解**(singular value decomposition, SVD)。 可以看出

$$A = \sum_{i=1}^{r} d_i \boldsymbol{v}_i \boldsymbol{u}_j^T.$$

当r比较小或者奇异值中仅有少数几个较大而其它奇异值都接近于零时， 可以近似A为

$$A = \sum_{i=1}^{r'} d_i \boldsymbol{v}_i \boldsymbol{u}_j^T.$$

其中r'远小于$\max(n, m)$。

详见 高惠璇 (1995) §5.4 和 Monahan (2001) §6.6。

30.2 对称阵特征值分解的Jacobi算法

矩阵A的特征值λ是A的特征多项式$A - \lambda I$的根， 但直接求多项式的根并不容易， 特征值和特征向量的计算一般都通过迭代算法实现。

Loading [MathJax]/extensions/MathEvents.js

§29.3引入的Givens变换是一个旋转变换， 可以仅改变向量中指定的两个元素并使得第二个指定元素变成零。 类似这样仅改变向量中第i, j两个元素的旋转变换矩阵可以写成

$$G_{ij}(\theta) = \begin{pmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \cos\theta & & & & \sin\theta & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & -\sin\theta & & & & \cos\theta & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{pmatrix}.$$

若A为对称阵，适当选取角度θ对A作如下变换

$$A^* = G_{ij}(\theta) A (G_{ij}(\theta))^T$$

可以使得$a^*_{ij} = a^*_{ji} = 0$， 这样的变换叫做**Jacobi**变换。 对A反复地作Jocobi变换可以使得非对角线元素趋于零。

考虑Jacobi变换(30.8)中角度θ的确定。 显然，A^*和A的不同仅体现在第i, j行和第i, j列， 其它元素保持不变; $G_{ij}(\theta)A$与A仅在第i, j行有差别， $A(G_{ij}(\theta))^T$与A仅在第i, j列有差别， A^*与$G_{ij}(\theta)A$仅在第i, j列有差别。 简单推导可得

$$a^*_{ij} = \frac{1}{2}(a_{jj} - a_{ii}) \sin 2\theta + a_{ij} \cos 2\theta,$$

只要取$\theta \in (-\frac{\pi}{4}, \frac{\pi}{4})$使得

$$\tau = \cot 2\theta = \frac{a_{ii} - a_{jj}}{2 a_{ij}}$$

即可使$a^*_{ij} = a^*_{ji} = 0$。

注意到$G_{ij}(\theta)$仅依赖于$\cos\theta$和$\sin\theta$, 设$x = \tan\theta$， 由三角函数公式得

$$x^2 + 2\tau x - 1 = 0.$$

x有两个根，为保证$|\theta| \leq \frac{\pi}{4}$取其中绝对值较小的一个， 为

Loading [MathJax]/extensions/MathEvents.js

$$x=\tan\theta=\text{sgn}(\tau)(-|\tau|+\sqrt{\tau^2+1})=\frac{\text{sgn}(\tau)}{|\tau|+\sqrt{\tau^2+1}},$$

(其中$\text{sgn}(\cdot)$表示符号函数，对非负数取1，对负数取-1)从$x=\tan\theta$再计算出

$$\cos\theta=(1+x^2)^{-\frac{1}{2}},\quad \sin\theta=x\cos\theta.$$

这样求$\cos\theta$和$\sin\theta$避免了三角函数计算并且x的计算方法考虑到了避免两个相近数相减造成精度损失的问题。

设A为实对称阵， J_{ij}是A关于下标(i,j)的的Jacobi变换阵， 令$A^*=J_{ij}AJ_{ij}^T$， 则A^*有如下性质:

$$\begin{aligned}
\text{i)}\ & A^* \text{ 仍为对称阵;} \\
\text{ii)}\ & a^*_{ij}=a^*_{ji}=0, \text{且对} k,t\neq i,j \text{有} a^*_{kt}=a_{kt}\ ; \\
\text{iii)}\ & \text{对} t\neq i,j \text{有} (a^*_{it})^2+(a^*_{jt})^2=a^2_{it}+a^2_{jt}\ ; \\
\text{iv)}\ & \sum_i\sum_j(a^*_{ij})^2=\sum_i\sum_j a^2_{ij}\ ; \\
\text{v)}\ & \text{off}(A^*)=\text{off}(A)-2a^2_{ij}, \text{其中} \text{off}(A) \text{表示} A \text{中非对角线元素的平方和.}
\end{aligned}$$

Jacobi算法从$A^{(0)}=A$和$U^{(0)}=I_n$出发反复作Jacobi变换， 设已有$A^{(k-1)}$, 则在第k步求$A^{(k-1)}$的非对角元素中绝对值最大者， 设为其(i_k,j_k)元素， 用(30.9)和(30.10)针对$A^{(k-1)}$和(i_k,j_k)求出Jacobi变换矩阵$J^{(k)}$, 则令$U^{(k)}=U^{(k-1)}(J^{(k)})^T$, $A^{(k)}=J^{(k)}A^{(k-1)}(J^{(k)})^T$, 如此重复直到$A^{(k)}$的非对角元素的绝对值最大值小于预定的精度$\epsilon$。 这时有$A=U\Lambda U^T$, $U=U^{(k)}$是正交阵，Λ近似为对角阵。

可以证明上述Jacobi算法的$A^{(k)}$收敛到一个对角阵且对角线元素为A的特征值。 此算法收敛较快， 但每次寻找非对角元素中绝对值最大的一个比较耗时。 改进的Jacobi算法从A和$U=I$出发， 在第k步时基于上一步的A计算一个界限 $\epsilon_k=\sqrt{\text{off}(A^{(k)})}/[n(n-1)]$, 然后对$A$的每个严格上三角元素都做一次Jacobi变换， 用变换后的矩阵代替原来的A并更新矩阵U， 但是若该严格上三角元素绝对值小于ϵ_k就跳过该元素。 所有严格上三角元素都处理过一遍才进入第$k+1$步并计算新的ϵ_{k+1}, 重复运算直到ϵ_{k+1}小于预先指定的误差限ϵ为止。

在R软件中， 用 `eigen()` 函数计算特征值和特征向量。

30.3 用QR分解方法求对称矩阵特征值分解

计算实对称矩阵特征值分解的一种较好的方法是利用Householder变换和Givens变换。 首先，若A为n阶实对称矩阵， 可以用$n-2$个Householder变换把它变成对称三对角矩阵: 设H_1为一个分块对角矩阵，主对角线的第一块为1阶单位阵， 第二块是把A的第一列最后$n-1$个元素中后$n-2$个元素变成零的Householder变换阵， 则H_1A第一列为$(a_{11},a^{(1)}_{21},0,\dots,0)^T$, 其中

Loading [MathJax]/extensions/MathEvents.js

$a_{21}^{(1)} = \sqrt{a_{21}^2 + \cdots + a_{n1}^2}$，且$H_1 A$的第一行与$A$的第一行完全相同，于是，$H_1 A H_1$的第一行为$(a_{11}, a_{21}^{(1)}, 0, \dots, 0)^T$, 注意到$H_1$的对称性与正交性，$H_1 A H_1$仍为对称阵，但是第一列和第一行的最后$n-2$个元素已经变成了零。在第二步，可以构造一个分块对角矩阵H_2，对角线第一块为I_2，第二块是把矩阵$H_1 A H_1$的第二列中最后$n-3$个元素变成零的Householder变换矩阵，把$H_1 A H_1$变成$H_2 H_1 A H_1 H_2$, 易见第一列和第一行不变，第二列和第二行的最后$n-3$个元素变成了零。如此进行下去得到$A^{(0)} = H_{n-2} \cdots H_1 A H_1 \cdots H_{n-2}$，使得$A^{(0)}$为三对角对称矩阵。

得到三对角对称矩阵$A^{(0)}$以后，进行QR迭代。设经过$k-1$次迭代后得到矩阵$A^{(k-1)}$，在第k步，先选一个平移量t_k，对矩阵$A^{(k-1)} - t_k I$用Givens变换方法作QR分解得到$A^{(k-1)} - t_k I = Q_k R_k$, 把得到的上三角阵$R_k$右乘$Q_k$再反向平移，得到

$$A^{(k)} = R_k Q_k + t_k I = Q_k^T (A^{(k-1)} - t_k I) Q_k + t_k I = Q_k^T A^{(k-1)} Q_k,$$

这样的$A^{(k)}$仍是三对角对称矩阵，如此迭代直到$A^{(k)}$变成对角形。收敛时，$A^{(k)}$的对角线元素为各个特征值，$H_1 \cdots H_{n-2} Q_1 \cdots Q_k$的各列为相应特征向量。

具体的算法比较复杂，详见 Monahan (2001) §6.5, Gentle (2007)} §7.4。

30.4 奇异值分解的计算

设A为任意非零$n \times m$实值矩阵，先说明A的奇异值分解的存在性。以下用$\|\cdot\|$表示向量的长度，即$\|\cdot\|_2$。首先，非负定阵$A^T A$和$A A^T$有共同的正特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$且个数$r$为矩阵$A$的秩。设$\boldsymbol{u}^{(i)}$是$A^T A$的特征值$\lambda_i$对应的单位特征向量（长度为1），则$A^T A \boldsymbol{u}^{(i)} = \lambda_i \boldsymbol{u}^{(i)}$, 由此式可得$\|A\boldsymbol{u}^{(i)}\|^2 = \lambda_i$。在$A^T A \boldsymbol{u}^{(i)} = \lambda_i \boldsymbol{u}^{(i)}$两边左乘矩阵$A$得到$(A A^T)(A\boldsymbol{u}^{(i)}) = \lambda_i (A\boldsymbol{u}^{(i)})$, 即$A\boldsymbol{u}^{(i)}$是矩阵$A A^T$的对应于特征值$\lambda_i$的特征向量，长度为$d_i = \sqrt{\lambda_i}$。设$A A^T$的所有特征向量组成的正交阵为$V$, $\boldsymbol{v}^{(j)}$是V的第j列，适当构造的V可使得

$$\boldsymbol{v}^{(j)} A \boldsymbol{u}^{(i)} = d_i \delta_{i-j},\ i = 1, 2, \dots, m,\ j = 1, 2, \dots, n,$$

其中δ_{i-j}是Kronecker记号，当$i = j$时表示1，当$i \neq j$时表示0, 当$i > r$时令$d_i = 0$。把(30.11)式写成矩阵形式即$V^T A U = D$，$A = V D U^T$ (D的定义见(30.6))，说明任何非零矩阵A均有奇异值分解。

以上的证明给出了求奇异值分解的一种方法：先选$A^T A$和$A A^T$中阶数较低一个, 不妨设是$A^T A$，求其特征值分解得到A的所有奇异值和矩阵U，然后利用上面的关系得到$A A^T$的对应于非零特征值的特征向量，如果需要再补充适当列向量组成正交方阵V即可。这种方法比较简单，但是计算$A^T A$会造成累积误差。

Loading [MathJax]/extensions/MathEvents.js

当A为$n \times m$的列满秩矩阵时， 可以用类似§30.3的QR分解方法来求A的奇异值分解。 方法描述如下。 仿照正交三角分解把一个对称矩阵变成三对角矩阵的做法， 我们可以先在A的左边乘以$n-2$个对称正交阵把A变成上三角形， 然后在此上三角形矩阵右边乘以$n-1$个对称正交阵将其变成上双对角阵， 即除对角线和上副对角线外的元素都是零的矩阵。 总之，存在n阶正交阵P和m阶正交阵Q使得

$$PAQ = \begin{pmatrix} B_{m \times m} \\ \mathbf{0}_{(n-m) \times m} \end{pmatrix},$$

其中B的元素满足当$i > j$或$j > i+1$时$b_{ij} = 0$, 且B满秩。

得到上双对角阵B后， 先求$B^T B$的特征值分解。 $B^T B$是一个三对角对称阵， 可以用的QR方法求特征值分解。 设$B^T B = U_1 D_m^2 U_1^T$, 其中$D_m = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$, $\lambda_1 \geq \dots \geq \lambda_m > 0$为$B^T B$的所有特征值， U_1各列为$B^T B$的特征向量。 令$V_1 = B U_1 D_m^{-1}$， 则$V_1^T V_1 = I_m$, 于是$B = V_1 D_m U_1^T$是B的奇异值分解。

这时，

$$\begin{aligned} PAQ =& \begin{pmatrix} B \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} V_1 D_m U_1^T \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} V_1 & \mathbf{0} \\ \mathbf{0} & I_{n-m} \end{pmatrix}_{n \times n} \begin{pmatrix} D_m \\ \mathbf{0} \end{pmatrix}_{n \times m} U_1^T \\ \triangleq& V_2 \begin{pmatrix} D_m \\ \mathbf{0} \end{pmatrix}_{n \times m} U_1^T, \end{aligned}$$

则V_2为n阶正交阵， 可得A有奇异值分解

$$A = (P^T V_2) D (Q U_1)^T \triangleq V D U^T,$$

其中$V = P^T V_2$为n阶正交阵， $U = Q U_1$为m阶正交阵， D为$n \times m$对角阵， 对角线元素为$B^T B$的特征值的算数平方根。

习题

习题1

编写关于实对称矩阵A用Jacobi方法求特征值分解的程序。

习题2

编写关于实对称矩阵A用改进的Jacobi方法求特征值分解的程序。

Loading [MathJax]/extensions/MathEvents.js

设A为n阶实对称方阵， 编写程序用正交相似变换$B = QAQ$ 把A变换为三对角对称矩阵B，其中Q为n阶对称正交阵, 输出B和Q的值。

习题4

设n方阵A为上双对角矩阵: 当$j \neq i, i+1$时总有$a_{ij} = 0$。 证明$B = A^T A$为三对角矩阵。

习题5

设A为n阶非负定对称矩阵， 用拉格朗日乘子法求$\boldsymbol{x}^T \boldsymbol{x} = 1$条件下 $\boldsymbol{x}^T A \boldsymbol{x}$的最大值点。

习题6

设A为n阶非负定对称矩阵， 用A的特征值分解求$\boldsymbol{x}^T \boldsymbol{x} = 1$条件下 $\boldsymbol{x}^T A \boldsymbol{x}$的最大值点。

References

Gentle, James E. 2007. *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer.

Monahan, John F. 2001. *Numerical Methods of Statistics*. Cambridge University Press.

高惠璇. 1995. 统计计算. 北京大学出版社.

Loading [MathJax]/extensions/MathEvents.js