Introduction to Linear Algebra – 第六章(4)：正定矩阵与奇异值分解

bourneli

2016年12月

正定矩阵是一种特殊的对称的矩阵，很多应用中都会涉及正定矩阵的相关性质。正定矩阵比较有趣的一点是，它可以将线性代数中很多重要的概念串联起来 ，形成一个整体。SVD是正定矩阵的一个主要应用，它可以将任意矩阵分解成正交矩阵和对角矩阵的乘积，可以用于数据压缩。

正定矩阵性质

对称 矩阵A具有下面5个性质中的一个，那么它就具备了剩下的其他性质，并且被称之为正定矩阵：

1. 所有主元大于0
2. 所有左上行列式大于0
3. 所有特征值大于0
4. $x^TAx > 0, x \neq 0$
5. $A = R^TR$并且R的列线性独立

例子

已知矩阵$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$。计算主元，得到$rref(A) = \begin{bmatrix} 2 & 1 \\ 0 & \frac{3}{2} \end{bmatrix}$。可以轻松计算特征值，左上行列式。

二次形式展开 $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2(x_1^2 + x_1x_2 + x_2^2) = 2(x_1 + \frac{1}{2}x_2)^2 + \frac{3}{2}x_2^2 > 0$

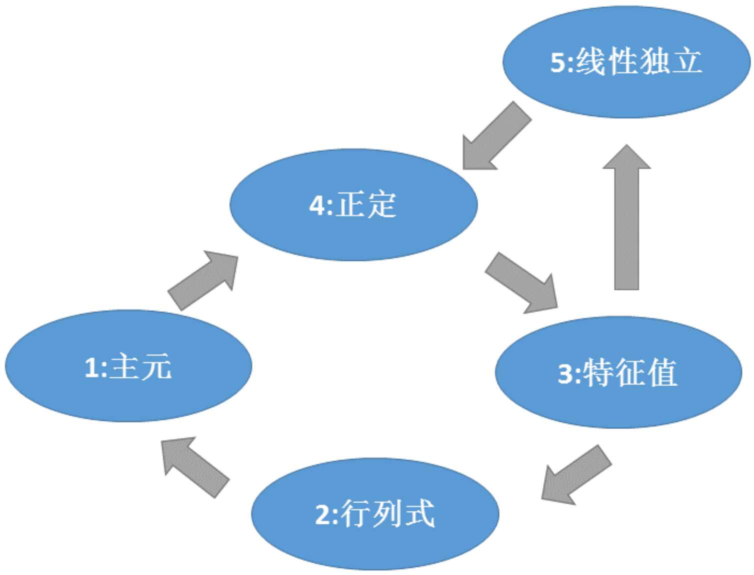
$A = LDL^T$ 分解,然后变成R

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

2乘2矩阵，完全符合上面的性质。

正定矩阵性质证明

上面有5个性质，如果每两个互相证明，那么需要证明20个定理，非常繁琐。其实，可以用图的角度看待这个问题，将每个性质看作图的顶点，如果性质之间具有推导特性，看作一条有向边。只要任意一个点出发，都可以到达其他顶点，那么问题就解决了，并且需要证明的定理也不需要20个。经过观察，可以得到下面的推导图，



根据上面的推导图，现在只需要证明6个定理，即可达到目的，下面逐个证明。

证明：4 $\Rightarrow$ 3

命题 若对称矩阵A有性质$x^TAx > 0, x \neq 0$，证明A的所有特征值均为正数。

证明

A对称，那么其必然可以对角化，有$A = Q\Lambda Q^T$,且将正定性质代入，具有$x^TQ\Lambda Q^Tx > 0$，其中Λ是对角矩阵，对角元素为$\lambda_1, \cdots, \lambda_n$。令$y = Q^Tx = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T$。所以，最后有

$$y\Lambda y^T > 0 \Rightarrow \sum_{i=1}^{n} \lambda_i y_i^2 > 0$$

上述不等式需要对任意不等于0向量的x成立，由于Q^T是标准正交矩阵，也就是对任意不等于0向量的y成立，所以$\lambda_i > 0$必须横成立。

证毕

证明：3 $\Rightarrow$ 5

命题 若对称矩阵A所有特征值大于0，证明存在矩阵R，其列线性独立，且$A = R^TR$。

证明

A对称，所以有$A = Q\Lambda Q^T$。因为条件$\lambda_i > 0$，所以有$A = Q\Lambda^{\frac{1}{2}}(\Lambda^{\frac{1}{2}})^TQ^T$，令$R = (Q\Lambda^{\frac{1}{2}})^T$,且$R^{-1} = Q\Lambda^{-\frac{1}{2}}$，那么有R可逆且$A = R^TR$。

证毕

证明: 5 $\Rightarrow$ 4

命题 对称矩阵A有$A = R^TR$并且R的列线性独立，证明 $x^TAx > 0, x \neq 0$

证明

因为R列线性独立，$x \neq 0$，所以$Rx \neq 0$。

直接按照条件展开$x^TAx = x^TR^TRx = (\|Rx\|)^2 > 0$恒成立。

证毕

证明：$3 \Rightarrow 2$

命题 对称矩阵A的所有特征值大于0，证明A的所有左上行列式大于0

证明

因为A所有特征值$\lambda_i > 0$，所以有$\|A\| = \prod_{i=1}^{n} \lambda_i > 0$。

假设A_k是矩阵A保留左上角k行和k列。因为对任意$x \neq 0$，均有$x^TAx > 0$,所以对于特殊形式的$x = (x_1, x_2, \cdots, x_k, 0, \cdots, 0) \neq 0, k = 1, 2, \cdots, n$仍然成立，则$A_k$与A有如下关系

$$\begin{aligned} 0 <& (x_1, x_2, \cdots, x_k, 0, \cdots, 0)A(x_1, x_2, \cdots, x_k, 0, \cdots, 0)^T \\ =& (x_1, x_2, \cdots, x_k)A_k(x_1, x_2, \cdots, x_k) \end{aligned}$$

所以A_k所有特征值大于0（利用$4 \Rightarrow 3$），所以$\|A_k\| > 0$。

证毕

证明：$2 \Rightarrow 1$

命题 对称矩阵A所有左上行列式大于0,证明A的所有主元大于0

证明

因为$\|A_k\| > 0$，所以A可逆。那么，可以只通过基础消元（特征值不变），而不需要基础换行或基础乘法（这些导致特征值变化），将A化简为只有**主对角线**的矩阵，此时对角线上的元素就是主元p_i，并且此过程做上角行列式保持不变。所以有$\|A_k\| = \prod_{i=1}^{k} p_i, \|A_{k-1}\| = \prod_{i=1}^{k-1} p_i$,推导出$p_k = \frac{\|A_k\|}{\|A_{k-1}\|} > 0$。

证毕

证明：$1 \Rightarrow 4$

命题 对称矩阵A所有主元大于0，证明$x^TAx > 0, x \neq 0$

证明

使用LDU分解，由于A对称，所以$A = LDL^T$，其中D是对角矩阵，且每个元素p_i是A的主元。代入x，有x^TLDL^Tx。令$y = L^Tx = (y_1, y_2, \cdots, y_n)$。所以$x^TLDL^Tx = y^TDy = \sum_{i=1}^{n} p_i y_i^2 > 0$。

证毕

至此，所以必要的定理已经证明完成，得到了完成的证明图，最终的问题得到了证实。

正定应用

- 假设C是正定，且A列线性独立，证明A^TCA正定，工程应用中的重要定理。

证明：$R = \Lambda^{\frac{1}{2}}QA$,易证明R线性独立。因为只有x=0时，$RX = 0$才成立，否则不行。

- 二元矩阵中，正定的二次型是椭圆形。
- 很多概率模型需要其中的矩阵正定。

奇异值分解SVD

假设矩阵A为任意维度，A的秩为r，如果不为方正，那么自然是得不到特征值和特征向量的。但是A^TA与AA^T却是对称的，而且根据正定性，它们是半正定矩阵。所以，可以计算相关的特征值和特征向量。那么，这两个矩阵的特征向量是否有什么关系呢？

大胆假设

我们来做一些大胆的设想，假设下面条件统统成立：

1. $v_1, \cdots, v_r \in R(A)$,$u_1, \cdots, u_r \in C(A)$
2. $v_1, \cdots . v_r$**是正交单位矩阵**, $u_1, \cdots, u_r$**是正交单位矩阵**
3. $Av_1 = \rho_1 u_1, Av_2 = \rho_2 u_2, \cdots, Av_r = \rho_r u_r$
4. $\rho_i > 0, i = 1, 2, \cdots, r$

读者可能不禁要想，对于任意矩阵A，这种假设是不是太过苛刻！先不管那么多，假设成立，可以得到

$$A\begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

由于v_i, u_i均是正交的，所以可以使用Gram-Schimidt方法将其扩展到整个行空间和列空间，得到下面

$$A\underbrace{\begin{bmatrix} v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix}}_{U}\underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}}_{\Sigma}$$

简化上面的表达式

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

也就是说，如果矩阵A满足上面这些条件，就可以将A分解成两个正交矩阵和一个对角矩阵，是不是非常优美！由于Σ后面对角线元素都为0，所以可以进一步简化：

$$\underbrace{A}_{m\times n} = \underbrace{U_r}_{m\times r}\underbrace{\Sigma_r}_{r\times r}\underbrace{V_r^T}_{r\times n}$$

U_r是U保留前r列，V_r^T是V^T保留前r行，Σ_r是保留前r行与r列。

小心求证

首先从A^TA开始，由于半正定性，所以可以将其特征值写成如下形式

$$A^TAv_i = \sigma_i^2 v_i \qquad (1)$$

由A^TA半正定性，可得v_i标准正交，且令$\sigma_i > 0 (i = 1, 2, \cdots, r)$称之为奇异值 (注意:不要与特征值混淆)。对公式(1)两边乘以v_i^T得到

$$v_i^T A^T A v_i = v_i^T \sigma_i^2 v_i \Rightarrow \|Av_i\|^2 = \sigma_i^2 \Rightarrow \sigma_i = \|Av_i\| \qquad (2)$$

从公式(2)可以了解到，奇异值σ_i是A^TA单位特征向量v_i在矩阵A列空间的投影向量的模长（很绕口，看看就好，不用太在意）。然后，继续对公式(1)两边乘以A,构造出AA^T

$$AA^TAv_i = A\sigma_i^2 v_i \Rightarrow AA^T(Av_i) = \sigma_i^2(Av_i) \qquad (3)$$

线性代数中有很多重要证明利用了括号变化，上面又是一个重要的例子。根据公式(3),可以发现AA^T与A^TA具有相同的特征值σ_i^2。并且特征向量具有关系，令u_i是AA^T的特征向量，明显有u_i正交，且由公式(3), $u_i = Av_i$。但是正交特征矩阵一般需要为单位矩阵，而且由公式(1)可以容易得到模长，所以最终令$u_i = \frac{Av_i}{\sigma_i}$。上面的假设得到了证明。

事实证明，上面的假设并不苛刻，对于任意A都可以找到这些向量和奇异值将其分解，不得不佩服发现SVD的数学家们。

计算策略

由于A^TA与AA^T的特征值相同，计算策略显而易见，计算维度较小的那个矩阵的特征值。因为特征值的复杂度是$O(n^3)$,n是矩阵阵的行数，当行数很高时，根本无法计算。spark中的SVD也是按这种策略实现。

奇异值的权重

奇异值都是大于0的实数，可以认为是权重，所以将奇异值按照从大到小的方式延对角线排列，可以得到唯一的分解。将SVD分解按代数形式展开，如下

$$A = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

A是r个秩为1的矩阵的线性组合，且每个矩阵室由单位向量乘法构成模，奇异值是系数。如果最后的奇异值比较小，可以忽略，对A的整体影响比较小，这样可以做有损压缩。