

# 矩阵代数部分

## 1. 特征值与特征向量

特征值与特征向量的定义： $Ax = \lambda x$ ，其中  $A_{n \times n}$ ， $x_{n \times 1} \neq 0$ 。在复数域上  $A_{n \times n}$  一定对应  $n$  个特征值，故有如下等式成立：

$$Aw_i = \lambda_i w_i, \quad i = 1, 2, \dots, n$$

于是可得

$$\begin{aligned} A[w_1, w_2, \dots, w_n] &= [w_1, w_2, \dots, w_n] \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \dots \\ & & & \lambda_n \end{bmatrix} \\ &\implies AW = W\Sigma \implies A = W\Sigma W^{-1} \end{aligned}$$

其中， $W$  的  $n$  个特征向量为标准正交基，满足  $W^T W = I$ ，即**正交矩阵**或**酉矩阵**。注意到以上的特征分解，矩阵  $A$  必须为方阵。如果不满足方阵条件，即  $A_{m \times n}$ ，此时对矩阵进行分解需要 SVD 了。

## 2. 奇异值分解

矩阵分解有很多方法，例如特征分解（Eigendecomposition）、LU分解（LU decomposition）、QR分解（QR decomposition）和极分解（Polar decomposition）等。**奇异值分解**（Singular Value Decomposition, SVD）是其中的一种矩阵分解方法：

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

其中， $U$  和  $V$  都是**正交矩阵**，在复数域内的话就是**酉矩阵**，即  $U^T U = I_{m \times m}$ ， $V^T V = I_{n \times n}$ ， $U$  和  $V$  的列分别叫做  $A$  的**左奇异向量**和**右奇异向量**。 $\Sigma$  就是一个非负**实对角矩阵**，主对角线上的每个元素都称为奇异值。其中我们发现  $AA^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T$ ，可知  $AA^T$  的特征值矩阵等于  $A$  的奇异值矩阵  $\Sigma$  的平方，于是特征值  $\lambda$  和 奇异值  $\sigma$  的关系即  $\sigma_i = \sqrt{\lambda_i}$ 。

求解 SVD 的步骤：

1. 求  $AA^T$  的特征值和特征向量，用单位化的特征向量构成  $U$
2. 求  $A^T A$  的特征值和特征向量，用单位化的特征向量构成  $V$
3. 将  $AA^T$  或者  $A^T A$  的特征值求平方根，然后构成  $\Sigma$

例如

$$A = \begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

则有

$$AA^T = \begin{pmatrix} 20 & 14 & 0 & 0 \\ 14 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$AA^T x = \lambda x \implies (AA^T - \lambda I)x = 0$$

求得 4 个特征值，对应的单位化过的特征向量整理为左奇异矩阵：

$$\lambda_1 \approx 29.86606875, \lambda_2 \approx 0.13393125, \lambda_3 = \lambda_4 = 0$$

$$U = \begin{pmatrix} 0.81741556 & -0.57604844 & 0 & 0 \\ 0.57604844 & 0.81741556 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

同样的过程求解  $A^T A$  的特征值和特征向量，得到右奇异矩阵：

$$\lambda_1 \approx 29.86606875, \lambda_2 \approx 0.13393125,$$

$$V = \begin{pmatrix} 0.40455358 & -0.9145143 \\ 0.9145143 & 0.40455358 \end{pmatrix}$$

矩阵  $\Sigma$  根据上面说的为特征值的平方根构成的对角矩阵：

$$\begin{pmatrix} 5.4649857 & 0 \\ 0 & 0.36596619 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

最后得到：

$$A_{4 \times 2} = U_{4 \times 4} \Sigma_{4 \times 2} V_{2 \times 2}^T$$

$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.81741556 & -0.57604844 & 0 & 0 \\ 0.57604844 & 0.81741556 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5.4649857 & 0 \\ 0 & 0.36596619 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0.40455358 & -0.9145143 \\ 0.9145143 & 0.40455358 \end{pmatrix}^T$$

对于奇异值，它跟我们特征分解中的特征值类似，在奇异值矩阵中也是按照从大到小排列，而且奇异值的减少特别的快，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% 以上的比例。

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \approx U_{m \times k} \Sigma_{k \times k} V_{k \times n}^T$$

SVD 可以用于 PCA 降维，来做数据压缩和去噪。也可以用于推荐算法，也可以用于 NLP 中的算法，比如潜在语义索引 (LSI)。

```
[U,S,V] = svd(A)
```

对于复矩阵  $A_{m \times n}$ ，称  $A^H A$  的  $n$  个特征根  $\lambda_i$  的算术平方根  $\sigma_i = \sqrt{\lambda_i}$  为  $A$  的奇异值。若记

$$\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$$

其中， $\sigma_1, \dots, \sigma_r$  是  $A$  的非零的全部的奇异值。则称  $m \times n$  维矩阵

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$$

为矩阵  $A$  的奇异值矩阵。奇异值分解：对于  $m \times n$  维矩阵  $A$ ，则分别存在一个  $m \times m$  维酉矩阵  $U$  和一个  $n \times n$  维酉矩阵  $V$ ，有

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

### 3. 广义特征值/特征向量

令  $A, B \in \mathbb{C}^{n \times n}$ ,  $\nu \in \mathbb{C}^n$ , 若标量  $\lambda$  和非零向量  $\nu$  满足:

$$A\nu = \lambda B\nu, \quad \nu \neq 0.$$

则称  $\lambda$  是矩阵  $A$  相对与矩阵  $B$  的广义特征值,  $\nu$  是与  $\lambda$  对应的广义特征向量。如果矩阵非满秩, 那么  $\lambda$  就有可能任意值包括零。当矩阵为单位阵时上式就成为普通的特征值问题, 因此可看作是对普通特征值问题的推广。当且仅当  $\xi$  是  $\{A, B\}$  的广义特征值时,  $A - \xi B$  奇异, 广义特征值是广义特征多项式  $|A - \xi B| = 0$  的根。

### 4. Toeplitz 矩阵

定义一个具有  $(2n - 1)$  个元素的  $n$  阶矩阵

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-n+2} \\ a_2 & a_1 & a_0 & \cdots & a_{-n+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix}$$

为 Toeplitz 矩阵, 简称 T 矩阵, 简记为  $A = (a_{-j+i})_{i,j=0}^n$ 。其中 T 矩阵完全由第 1 行和第 1 列的  $(2n - 1)$  个元素确定。

可见矩阵中位于任意一条平行于主对角线的直线上的元素全都是相等的, 且关于副对角线对称。矩阵是斜对称矩阵, 一般不是对称矩阵。在数据信息处理过程中, 有限元素法、概率统计以及滤波理论等领域常常遇到 T 矩阵。

### 5. Hankel 矩阵

定义以下的  $(n + 1)$  阶矩阵

$$H = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{bmatrix}$$

为 Hankel 矩阵或正交对称矩阵。可见 H 矩阵完全由其第 1 行和第  $n + 1$  列的  $(2n + 1)$  个元素确定。其中沿着所有垂直于主对角线的直线上有相同的元素。

### 6. Vandermonde 矩阵

定义具有以下形式的  $m \times n$  阶矩阵

$$V(a_1, \dots, a_n) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \cdots & a_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^{m-1} & a_2^{m-1} & a_3^{m-1} & \cdots & a_n^{m-1} \end{bmatrix}$$

为 Vandermonde 矩阵或范德蒙矩阵。范德蒙矩阵的转置也称为范德蒙矩阵。其中沿着所有**垂直于主对角线**的直线上有相同的元素。范德蒙行列式如下

$$A = \begin{pmatrix} 1 & 1 & 1 \\ b & c & d \\ b^2 & c^2 & d^2 \end{pmatrix} \Rightarrow |A| = -(b-c)(b-d)(c-d)$$

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ b & c & d & e \\ b^2 & c^2 & d^2 & e^2 \\ b^3 & c^3 & d^3 & e^3 \end{pmatrix} \Rightarrow |B| = (b-c)(b-d)(b-e)(c-d)(c-e)(d-e)$$

如果 V 矩阵满足  $a_i \neq a_j$ , 则 V 矩阵一定是非奇异的。

## 7. Kronecker 积

将  $p \times q$  维矩阵  $A = [a_{i,j}]$ , 和  $m \times n$  维矩阵  $B = [b_{l,k}]$  的克罗内克积记作  $A \otimes B$ , 它是一个  $pm \times qn$  矩阵, 形式为

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1q}B \\ a_{21}B & a_{22}B & \cdots & a_{2q}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}B & a_{p2}B & \cdots & a_{pq}B \end{bmatrix}$$

## 8. Hadamard 积

将  $m \times n$  维矩阵  $A = [a_{i,j}]$ , 和  $m \times n$  维矩阵  $B = [b_{l,k}]$  的哈达玛乘积记作  $A \odot B$ , 也称作 Schur 乘积。它仍然是一个  $m \times n$  矩阵, 形式为

$$A \odot B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{bmatrix}$$

## 9. M-P 广义逆

对于任意矩阵  $A \in \mathbb{C}_{m \times n}$ , 如果存在矩阵  $G \in \mathbb{C}_{m \times n}$ , 满足

$$\begin{aligned} AGA &= A \\ GAG &= G \\ (AG)^H &= AG \\ (GA)^H &= GA \end{aligned}$$

则称  $G$  为  $A$  的 M-P 广义逆, 记为  $A^\dagger$ 。**同时满足上述四个方程的广义可逆矩阵具有唯一性。**部分满足上述四个方程的矩阵不唯一, 每一种广义逆矩阵都包含着一类矩阵。

若  $A \in \mathbb{C}_n^{m \times n}$ , 则  $A^\dagger = (A^H A)^{-1} A^H$ 。若  $A \in \mathbb{C}_m^{m \times n}$ , 则  $A^\dagger = A^H (A^H A)^{-1}$ 。