第十五章
奇异值分解

定义与定理

矩阵的奇异值分解是指，将一个非零的 $m \times n$ 实矩阵 A，$A \in \mathbf{R}^{m \times n}$，表示为以下三个实矩阵乘积形式的运算①，即进行矩阵的因子分解:

$$A = U \Sigma V^{\mathrm{T}} \tag{15.1}$$

其中 U 是 m 阶正交矩阵（orthogonal matrix），V 是 n 阶正交矩阵，Σ 是由降序排列的非负的对角线元素组成的 $m \times n$ 矩形对角矩阵（rectangular diagonal matrix），满足

$$UU^{\mathrm{T}} = I$$

$$VV^{\mathrm{T}} = I$$

$$\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_p)$$

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$$

$$p = \min(m, n)$$

定义与定理

- $U\Sigma V^{\mathrm{T}}$： 矩阵A的奇异值分解（singular value decomposition, SVD)
- σ_i ： 矩阵 A的奇异值（singular value)
- U的列向量： 左奇异向量（left singular vector)
- V 的列向量： 右奇异向量（right singular vector)

- 注意奇异值分解不要求矩阵A是方阵，事实上矩阵的奇异值分解可以看作是方阵的对角化的推广。

例

- 给定一个5x4矩阵A

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

例

- 它的奇异值分解由三个矩阵的乘积 $U\Sigma V^{\mathrm{T}}$ 给出

$$U=\begin{bmatrix} 0 & 0 & \sqrt{0.2} & 0 & \sqrt{0.8} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \sqrt{0.8} & 0 & -\sqrt{0.2} \end{bmatrix}, \quad \Sigma=\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad V^{\mathrm{T}}=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

例

- 矩阵 Σ 是对角矩阵，对角线外的元素都是0，对角线上的元素非负，按降序排列。

- 矩阵U和V是正交矩阵，它们与各自的转置矩阵相乘是单位矩阵，即

$$UU^{\mathrm{T}} = I_5, \quad VV^{\mathrm{T}} = I_4$$

例

- 矩阵的奇异值分解不是唯一的。在此例中如果选择U为

$$U=\begin{bmatrix} 0 & 0 & \sqrt{0.2} & \sqrt{0.4} & -\sqrt{0.4} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{0.5} & \sqrt{0.5} \\ 0 & 0 & \sqrt{0.8} & -\sqrt{0.1} & \sqrt{0.1} \end{bmatrix}$$

- 而 Σ 和V不变，那么 $U\Sigma V^{\mathrm{T}}$ 也是A的一个奇异值分解

奇异值分解基本定理

- 若A为一 m x n 实矩阵，$A \in \mathbf{R}^{m\times n}$，，则A 的奇异值分解存在

$$A = U\Sigma V^{\mathrm{T}}$$

- 其中U是m阶正交矩阵，V是n阶正交矩阵，Σ 是 m x n 矩形对角矩阵，其对角线元素非负，且按降序排列。

奇异值分解基本定理

- 证明

- 证明是构造性的，对给定的矩阵A，构造出其奇异值分解的各个矩阵。

- 为了方便，不妨假设$m \geqslant n$，如果$m<n$证明仍然成立。

Tsinghua University

奇异值分解基本定理

- (1）确定V和 Σ

- 首先构造n阶正交实矩阵V和 m x n 矩形对角实矩阵 Σ

- 矩阵A是 m x n 实矩阵，则矩阵A^TA是n阶实对称矩阵。

- 因而A^TA的特征值都是实数，并且存在一个n阶正交实矩阵V实现A^TA的对角化，使得 $V^T(A^TA)V = A$ 成立

- 其中A是n阶对角矩阵，其对角线元素由A^TA的特征值组成。

奇异值分解基本定理

- 而且，A^TA的特征值都是非负的。事实上，令 λ 是A^TA的一个特征值，x是对应的特征向量，则

$$\|Ax\|^2 = x^{\mathrm{T}}A^{\mathrm{T}}Ax = \lambda x^{\mathrm{T}}x = \lambda\|x\|^2$$

- 于是

$$\lambda = \frac{\|Ax\|^2}{\|x\|^2} \geqslant 0$$

奇异值分解基本定理

- 可以假设正交矩阵V的列的排列使得对应的特征值形成降序排列

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$$

- 计算特征值的平方根（实际就是矩阵A的奇异值）

$$\sigma_j = \sqrt{\lambda_j}, \quad j = 1, 2, \cdots, n$$

- 设矩阵A的秩是r, rank(A) = r，则矩阵A^TA的秩也是r

奇异值分解基本定理

- 由于A^TA是对称矩阵， 它的秩等于正的特征值的个数，所以

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > 0, \quad \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$$

- 对应地有 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0, \quad \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$
- 令 $V_1 = [\nu_1 \quad \nu_2 \quad \cdots \quad \nu_r], \quad V_2 = [\nu_{r+1} \quad \nu_{r+2} \quad \cdots \quad \nu_n]$
- 其中$v_1,\cdots,v_r$为A^TA的正特征值对应的特征向量，$v_{r+1},\cdots,v_n$为0特征值对应的特征向量，则 $V = [V_1 \quad V_2]$ 15.6

- 这就是矩阵A的奇异值分解中的n阶正交矩阵V。

奇异值分解基本定理

- 令

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

- 则 Σ_1 是一个r阶对角矩阵，其对角线元素为按降序排列的正的 $\sigma_1, \cdots, \sigma_r$，于是 m x n 矩形对角矩阵 Σ 可以表为

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \quad 15.7$$

- 这就是矩阵A的奇异值分解中的 m x n 矩形对角矩阵 Σ

奇异值分解基本定理

- 在式（15.6）中，V_2的列向量是A^TA对应于特征值为0的特征向量，因此

$$A^{\mathrm{T}}Av_j=0,\quad j=r+1,\cdots,n$$

- 于是，V_2的列向量构成了A^TA的零空间$N(A^TA)$，而$N(A^TA) = N(A)$。
- 所以V2的列向量构成A的零空间的一组标准正交基。因此，

$$AV_2=0$$

- 由于V是正交矩阵，由式（15.6）可得

$$I=VV^{\mathrm{T}}=V_1V_1^{\mathrm{T}}+V_2V_2^{\mathrm{T}}$$

$$A=AI=AV_1V_1^{\mathrm{T}}+AV_2V_2^{\mathrm{T}}=AV_1V_1^{\mathrm{T}} \qquad 15.11$$

奇异值分解基本定理

- （2）确定U

- 接着构造m阶正交实矩阵
- 令

$$u_j = \frac{1}{\sigma_j} A v_j, \quad j = 1, 2, \cdots, r \qquad 15.12$$

$$U_1 = [u_1 \quad u_2 \quad \cdots \quad u_r]$$

- 则有

$$AV_1 = U_1 \Sigma_1 \qquad 15.14$$

清华大学
Tsinghua University

奇异值分解基本定理

- U_1的列向量构成了一组标准正交集，因为

$$\begin{aligned} u_i^{\mathrm{T}} u_j &= \left(\frac{1}{\sigma_i} v_i^{\mathrm{T}} A^{\mathrm{T}}\right)\left(\frac{1}{\sigma_j} A v_j\right) \\ &= \frac{1}{\sigma_i \sigma_j} v_i^{\mathrm{T}} (A^{\mathrm{T}} A v_j) \\ &= \frac{\sigma_j}{\sigma_i} v_i^{\mathrm{T}} v_j \\ &= \delta_{ij}, \quad i = 1, 2, \cdots, r; \quad j = 1, 2, \cdots, r \end{aligned} \qquad 15.15$$

奇异值分解基本定理

- 由式（15.12）和式（15.15）可知，$u_1, u_2, \cdots, u_r$ 构成A的列空间的一组标准正交基，列空间的维数为r。

- 如果将A看成是从R^n到R^m的线性变换，则A的列空间和A的值域R(A）是相同的。因此$u_1, u_2, \cdots, u_r$ 也是R(A）的一组标准正交基。

- 若 $R(A)^{\perp}$ 表示R(A)的正交补，则有R(A)的维数为r，$R(A)^{\perp}$ 的维数为 m – r，两者的维数之和等于m。而且有 $R(A)^{\perp} = N(A^T)$ 成立

奇异值分解基本定理

- 令 $\{u_{r+1}, u_{r+2}, \cdots, u_m\}$ 为N(A^T) 的一组标准正交基，并令

$$U_2 = [u_{r+1} \quad u_{r+2} \quad \cdots \quad u_m]$$

$$U = [U_1 \quad U_2] \qquad 15.16$$

- 则u_1, u_2,⋯, u_m 构成了R^m的一组标准正交基。因此，U是m阶正交矩阵。

- 这就是矩阵A的奇异值分解中的m阶正交矩阵。

奇异值分解基本定理

- （3）证明 $U\Sigma V^{\mathrm{T}}=A$

- 由式（15.6)、式（15.7)、式（15.11)、式（15.14）和式（15.16）得

$$
\begin{aligned}
U\Sigma V^{\mathrm{T}} &= [U_1 \quad U_2]\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_1^{\mathrm{T}}\\ V_2^{\mathrm{T}}\end{bmatrix}\\
&= U_1\Sigma_1 V_1^{\mathrm{T}}\\
&= AV_1V_1^{\mathrm{T}}\\
&= A
\end{aligned}
$$

清華大學
Tsinghua University

紧奇异值分解与截断奇异值分解

- $A = U\Sigma V^{\mathrm{T}}$ 又称为矩阵的完全奇异值分解（full singular value decomposition)。
- 实际常用的是奇异值分解的紧凑形式和截断形式。
- 紧奇异值分解是与原始矩阵等秩的奇异值分解
- 截断奇异值分解是比原始矩阵低秩的奇异值分解。

Tsinghua University

紧奇异值分解

- 设有 m x n 实矩阵A， 其秩为rank(A)=r, r≤min(m,n)， 则称 $U_r \Sigma_r V_r^{\mathrm{T}}$ 为A的紧奇异值分解（compact singular value decomposition)， 即

$$A = U_r \Sigma_r V_r^{\mathrm{T}}$$

- U_r： m x r 矩阵
- V_r： n x r 矩阵
- Σ_r r阶对角矩阵

- 矩阵U_r 由完全奇异值分解中的U的前r列、矩阵V_r的前r列、矩阵 Σ_r 凡由 Σ 的前r个对角线元素得到。紧奇异值分解的对角矩阵 Σ_r 的秩与原始矩阵A的秩相等。

例

- 矩阵A的秩r = 3

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

$$A = U_r \Sigma_r V_r^{\mathrm{T}}$$

例

- A的紧奇异值分解是 $A = U_r \Sigma_r V_r^{\mathrm{T}}$

$$U_r = \begin{bmatrix} 0 & 0 & \sqrt{0.2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{0.8} \end{bmatrix} \quad \Sigma_r = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \sqrt{5} \end{bmatrix}, \quad V_r^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

截断奇异值分解

- 在矩阵的奇异值分解中，只取最大的k个奇异值（$k<r$, r为矩阵的秩）对应的部分，就得到矩阵的截断奇异值分解。

- 实际应用中提到矩阵的奇异值分解时，通常指截断奇异值分解。

截断奇异值分解

设 A 为 $m\times n$ 实矩阵，其秩 $\operatorname{rank}(A)=r$，且 $0<k<r$，则称 $U_k\Sigma_kV_k^{\mathrm{T}}$ 为矩阵 A 的截断奇异值分解（truncated singular value decomposition）

$$A\approx U_k\Sigma_kV_k^{\mathrm{T}} \tag{15.19}$$

其中 U_k 是 $m\times k$ 矩阵，V_k 是 $n\times k$ 矩阵，Σ_k 是 k 阶对角矩阵；矩阵 U_k 由完全奇异值分解中 U 的前 k 列、矩阵 V_k 由 V 的前 k 列、矩阵 Σ_k 由 Σ 的前 k 个对角线元素得到。对角矩阵 Σ_k 的秩比原始矩阵 A 的秩低。

例

- 矩阵A的秩为3

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

- 若取k=2，则其截断奇异值分解是 $A \approx A_2 = U_2 \Sigma_2 V_2^{\mathrm{T}}$

$$U_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \quad V_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad A_2 = U_2 \Sigma_2 V_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

几何解释

- 从线性变换的角度理解奇异值分解， m x n 矩阵A表示从n维空间R^n到m维空间R^m的一个线性变换，

$$T : x \to Ax \qquad x \in \mathbf{R}^n, \ Ax \in \mathbf{R}^m$$

- x和Ax分别是各自空间的向量。

- 线性变换可以分解为三个简单的变换：
  - 一个坐标系的旋转或反射变换
  - 一个坐标轴的缩放变换
  - 另一个坐标系的旋转或反射变换
- 奇异值定理保证这种分解一定存在。这就是奇异值分解的几何解释。

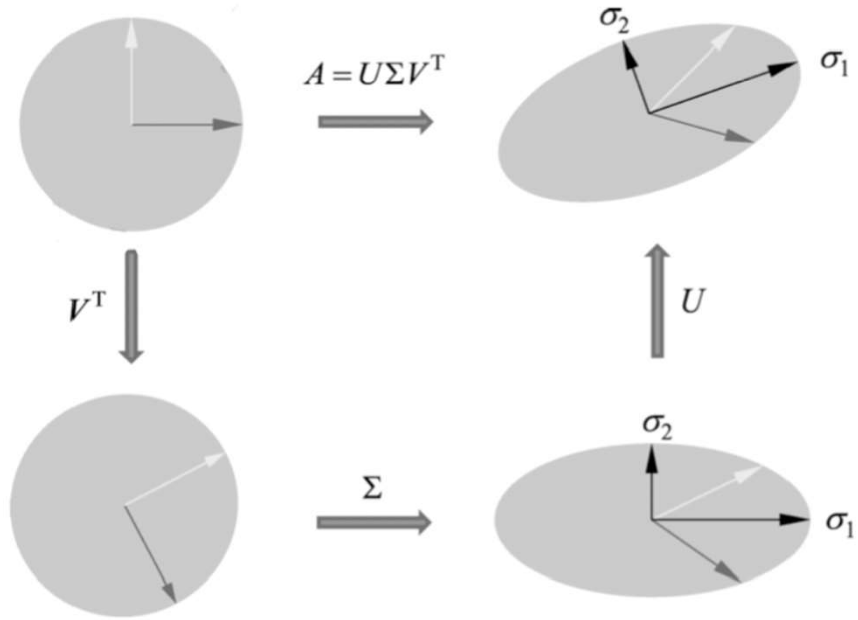
几何解释

- 对矩阵A进行奇异值分解，得到 $A = U\Sigma V^{\mathrm{T}}$
- ,V和U都是正交矩阵
- V的列向量$v_1, v_2, \cdots, v_n$构成R^n空间的一组标准正交基，表示R中的正交坐标系的旋转或反射变换
- U的列向量$u_1, u_2, \cdots, u_m$构成R^m空间的一组标准正交基，表示 Rm 中的正交坐标系的旋转或反射变换
- Σ 的对角元素 $\sigma_1, \sigma_2, \cdots, \sigma_n$ 是一组非负实数，表示R^n中的原始正交坐标系坐标轴的 $\sigma_1, \sigma_2, \cdots, \sigma_n$ 倍的缩放变换。

几何解释

- 任意一个向量 $x \in \mathbf{R}^n$ ，经过基于 $A = U\Sigma V^{\mathrm{T}}$ 的线性变换，等价于经过坐标系的旋转或反射变换V^{T}，坐标轴的缩放变换 Σ ，以及坐标系的旋转或反射变换U，得到向量 $Ax \in \mathbf{R}^m$ 。

- 原始空间的标准正交基，
  经过坐标系的旋转变换V^{T}、
  坐标轴的缩放变换刃、
  坐标系的旋转变换U，
  得到和经过线性变换A等价的结果。



例

- 给定一个2阶矩阵

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

- 其奇异值分解为

$$U = \begin{bmatrix} 0.8174 & -0.5760 \\ 0.5760 & 0.8174 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3.8643 & 0 \\ 0 & 0.2588 \end{bmatrix}, \quad V^{\mathrm{T}} = \begin{bmatrix} 0.9327 & 0.3606 \\ -0.3606 & 0.9327 \end{bmatrix}$$

例

- 观察基于矩阵A的奇异值分解将R^2的标准正交基

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- 进行线性转换的情况
- 首先，V^T表示一个旋转变换，将标准正交基e_1, e_2旋转，得到向量$V^T e_1$, $V^T e_2$:

$$V^T e_1 = \begin{bmatrix} 0.9327 \\ -0.3606 \end{bmatrix}, \quad V^T e_2 = \begin{bmatrix} 0.3606 \\ 0.9327 \end{bmatrix}$$

例

- 其次，Σ 表示一个缩放变换，将向量$V^T e_1$, $V^T e_2$ 在坐标轴方向缩放 σ_1 倍和 σ_2 倍，得到向量 $\Sigma V^{\mathrm{T}} e_1$, $\Sigma V^{\mathrm{T}} e_2$：

$$\Sigma V^{\mathrm{T}} e_1 = \begin{bmatrix} 3.6042 \\ -0.0933 \end{bmatrix}, \quad \Sigma V^{\mathrm{T}} e_2 = \begin{bmatrix} 1.3935 \\ 0.2414 \end{bmatrix}$$

- 最后，U表示一个旋转变换，再将向量 $\Sigma V^{\mathrm{T}} e_1$, $\Sigma V^{\mathrm{T}} e_2$ 旋转，得到向量 $U\Sigma V^{\mathrm{T}} e_1$, $U\Sigma V^{\mathrm{T}} e_2$，也就是向量$Ae_1$，$Ae_2$：

$$Ae_1 = U\Sigma V^{\mathrm{T}} e_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad Ae_2 = U\Sigma V^{\mathrm{T}} e_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

主要性质

- （1）设矩阵A的奇异值分解为 $A = U\Sigma V^{\mathrm{T}}$， 则一下关系成立：

$$A^{\mathrm{T}}A = (U\Sigma V^{\mathrm{T}})^{\mathrm{T}}(U\Sigma V^{\mathrm{T}}) = V(\Sigma^{\mathrm{T}}\Sigma)V^{\mathrm{T}}$$

$$AA^{\mathrm{T}} = (U\Sigma V^{\mathrm{T}})(U\Sigma V^{\mathrm{T}})^{\mathrm{T}} = U(\Sigma\Sigma^{\mathrm{T}})U^{\mathrm{T}}$$

- 也就是说，矩阵$A^{T}A$和AA^{T}的特征分解存在，且可以由矩阵A的奇异值分解 的矩阵表示。

- V的列向量是$A^{T}A$的特征向量
- U的列向量是AA^{T}的特征向量
- Σ 的奇异值是$A^{T}A$和AA^{T}的特征值的平方根。

主要性质

- （2）在矩阵A的奇异值分解中，奇异值、左奇异向量和右奇异向量之间存在对应关系。

- 由 $A = U\Sigma V^{\mathrm{T}}$ 易知 $AV = U\Sigma$
- 比较这一等式两端的第j列，得到 $Av_j = \sigma_j u_j, \quad j = 1, 2, \cdots, n$
- 这是矩阵A的右奇异向量和奇异值、 左奇异向量的关系

- 类似地，由 $A^{\mathrm{T}}U = V\Sigma^{\mathrm{T}}$，
- 得到 $A^{\mathrm{T}}u_j = \sigma_j v_j, \quad j = 1, 2, \cdots, n$ $A^{\mathrm{T}}u_j = 0, \quad j = n+1, n+2, \cdots, m$
- 这是矩阵A的左奇异向量和奇异值、右奇异向量的关系。

主要性质

- （3）矩阵A的奇异值分解中，奇异值 $\sigma_1, \sigma_2, \cdots, \sigma_n$ 是唯一的，而矩阵U和V不是唯一的。

- （4）矩阵A和 Σ 的秩相等，等于正奇异值 σ_i 的个数r（包含重复的奇异值）。

主要性质

- （5）

- 矩阵A的r个右奇异向量$v_1, v_2, \cdots, v_r$构成A^T的值域$R(A^T)$的一组标准正交基。

- 因为矩阵A^T是从R^m映射到矽的线性变换，则A^T的值域$R(A^T)$和A^T的列空间是相同的，$v_1, v_2, \cdots, v_r$是A^T的一组标准正交基，因而也是$R(A^T)$的一组标准正交基。

标准性质

- 矩阵A的n-r个右奇异向量v_{r+1},v_{r+2}, ⋯,v_n构成A的零空间N(A)的一组标准正交基。

- 矩阵A的r个左奇异向量u_1, u_2, ⋯, u_r构成值域R(A）的一组标准正交基。

- 矩阵A的m-r个左奇异向量u_{r+1},u_{r+2}, ⋯,u_m构成A^T的零空间N(A^T）的一组标准正交基。

Tsinghua University

奇异值分解的计算

- 矩阵A的奇异值分解可以通过求对称矩阵A^TA的特征值和特征向量得到。
- A^TA的特征向量构成正交矩阵V的列
- A^TA的特征值 λ_j 的平方根为奇异值 σ_i， 即

$$\sigma_j = \sqrt{\lambda_j}, \quad j = 1, 2, \cdots, n$$

- 对其由大到小排列作为对角线元素，构成对角矩阵 Σ
- 求正奇异值对应的左奇异向量，再求扩充的A^T的标准正交基，构成正交矩阵U的列
- 从而得到A的奇异值分解 $A = U\Sigma V^{\mathrm{T}}$

奇异值分解的计算

- （1）求A^TA的特征值和特征向量
- 计算对称矩阵$W=A^TA$
- 求解特征方程 $(W-\lambda I)x=0$
- 得到特征值 λ_i，并将特征值由大到小排列 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$
- 将特征值 $\lambda_i\ (i=1,2,\cdots,n)$ 代入特征方程求得对应的特征向量

- （2）求n阶正交矩阵V
- 将特征向量单位化，得到单位特征向量$v_1,v_2,\cdots,v_n$，构成n阶正交矩阵V: $V=\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$

奇异值分解的计算

- （3）求 m x n 对角矩阵 Σ

- 计算A的奇异值

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2, \cdots, n$$

- 构造 m x n 矩形对角矩阵 Σ，主对角线元素是奇异值，其余元素是零

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$$

奇异值分解的计算

- （4） 求m阶正交矩阵U
- 对A的前r个正奇异值，令 $u_j = \frac{1}{\sigma_j} A v_j, \quad j = 1, 2, \cdots, r$
- 得到 $U_1 = [\ u_1 \quad u_2 \quad \cdots \quad u_r\ ]$
- 求A^T的零空间的一组标准正交基$\{u_{r+1}, u_{r+2}, \cdots, u_m\}$， 令

$$U_2 = [\ u_{r+1} \quad u_{r+2} \quad \cdots \quad u_m\ ]$$

- 并令 $U = [\ U_1 \quad U_2\ ]$

- （5）得到奇异值分解 $A = U \Sigma V^{\mathrm{T}}$

例

- 试求矩阵

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

- 的奇异值分解

例

- （1）求A^TA的特征值和特征向量

$$A^{\mathrm{T}}A=\begin{bmatrix}1&2&0\\1&2&0\end{bmatrix}\begin{bmatrix}1&1\\2&2\\0&0\end{bmatrix}=\begin{bmatrix}5&5\\5&5\end{bmatrix}$$

$$(A^{\mathrm{T}}A-\lambda I)x=0$$

- 得到齐次线性方程组

$$\begin{cases}(5-\lambda)x_1+ & 5x_2=0\\ 5x_1+ & (5-\lambda)x_2=0\end{cases}$$

例

- 该方程有非零解的充要条件是

$$\begin{vmatrix} 5-\lambda & 5 \\ 5 & 5-\lambda \end{vmatrix} = 0$$

$$\lambda^2 - 10\lambda = 0$$

- 解此方程，得矩阵A^TA的特征值 $\lambda_1 = 10$ 和 $\lambda_2 = 0$。
- 将特征值代入线性方程组，得到对应的单位特征向量

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

例

- （2）求正交矩阵V
- 构造正交矩阵V

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

- （3）求对角矩阵 Σ
- 奇异值为 $\sigma_1 = \sqrt{\lambda_1} = \sqrt{10}$ 和 $\sigma_2 = 0$
- 构造对角矩阵

$$\Sigma = \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

例

- （4）求正交矩阵U

- 基于A的正奇异值计算得到列向量u_1

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}$$

- 列向量u_2, u_3是A^T的零空间$N(A^T)$的一组标准正交基

例

- 求解

$$A^{\mathrm{T}}x = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 + 2x_2 + 0x_3 = 0$$
$$x_1 = -2x_2 + 0x_3$$

- 分别取(x_2, x_3)为(1,0)和(0,1)，得到N(A^T)的基 $(-2,1,0)^{\mathrm{T}}, \quad (0,0,1)^{\mathrm{T}}$
- N(A^T)的一组标准正交基是 $u_2 = \left(-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right)^{\mathrm{T}}, \quad u_3 = (0,0,1)^{\mathrm{T}}$
- 构造正交矩阵U

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

例

- （5）矩阵A的奇异值分解

$$A = U\Sigma V^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Tsinghua University

弗罗贝尼乌斯范数

- 奇异值分解也是一种矩阵近似的方法，这个近似是在弗罗贝尼乌斯范数（Frobenius norm）意义下的近似。

- 矩阵的弗罗贝尼乌斯范数是向量的LZ范数的直接推广，对应着机器学习中的平方损失函数。

- 设矩阵 $A \in \mathbf{R}^{m\times n}$，$A=[a_{ij}]_{m\times n}$，定义矩阵A的弗罗贝尼乌斯范数为

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij})^2\right)^{\frac{1}{2}}$$

弗罗贝尼乌斯范数

- 引理15.1

设矩阵 $A \in \mathbf{R}^{m\times n}$，$A$ 的奇异值分解为 $U\Sigma V^{\mathrm{T}}$，其中 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$，则

$$\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$$

弗罗贝尼乌斯范数

- 证明:
- 一般地，若Q是m阶正交矩阵，则有 $\|QA\|_F = \|A\|_F$
- 因为 $$\|QA\|_F^2 = \|(Qa_1, Qa_2, \cdots, Qa_n)\|_F^2$$ $$= \sum_{i=1}^{n} \|Qa_i\|_2^2 = \sum_{i=1}^{n} \|a_i\|_2^2 = \|A\|_F^2$$
- 同样，若P是n阶正交矩阵，则有 $\|AP^{\mathrm{T}}\|_F = \|A\|_F$
- 故 $\|A\|_F = \|U\Sigma V^{\mathrm{T}}\|_F = \|\Sigma\|_F$
- 即 $\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$

矩阵的最优近似

- 奇异值分解是在平方损失弗罗贝尼乌斯范数）意义下对矩阵的最优近似，即数据压缩。

设矩阵 $A \in \mathbf{R}^{m\times n}$，矩阵的秩 $\operatorname{rank}(A)=r$，并设 $\mathcal{M}$ 为 $\mathbf{R}^{m\times n}$ 中所有秩不超过 k 的矩阵集合，$0<k<r$，则存在一个秩为 k 的矩阵 $X \in \mathcal{M}$，使得

$$\|A-X\|_F = \min_{S\in\mathcal{M}} \|A-S\|_F \tag{15.31}$$

称矩阵 X 为矩阵 A 在弗罗贝尼乌斯范数意义下的最优近似。

矩阵的最优近似

设矩阵 $A \in \mathbf{R}^{m\times n}$，矩阵的秩 $\mathrm{rank}(A)=r$，有奇异值分解 $A = U\Sigma V^{\mathrm{T}}$，并设 $\mathcal{M}$ 为 $\mathbf{R}^{m\times n}$ 中所有秩不超过 k 的矩阵的集合，$0<k<r$，若秩为 k 的矩阵 $X \in \mathcal{M}$ 满足

$$\|A-X\|_F = \min_{S\in\mathcal{M}} \|A-S\|_F \qquad 15.32$$

则

$$\|A-X\|_F = (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}} \qquad 15.33$$

特别地，若 $A' = U\Sigma' V^{\mathrm{T}}$，其中

$$\Sigma' = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix}$$

则

$$\|A-A'\|_F = (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}} = \min_{S\in\mathcal{M}} \|A-S\|_F$$

矩阵的最优近似

- 证明

- 令 $X \in \mathcal{M}$ 为满足式（15.32）的一个矩阵。由于

$$\|A - X\|_F \leqslant \|A - A'\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$$

- 下面证明 $\|A - X\|_F \geqslant (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$ 于是式（15.33）成立

矩阵的最优近似

- 设X的奇异值分解为 $Q\Omega P^{\mathrm{T}}$;
- 其中

$$\Omega=\begin{bmatrix} \omega_1 & & & & & \\ & \ddots & & & 0 & \\ & & \omega_k & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{bmatrix}=\begin{bmatrix} \Omega_k & 0 \\ 0 & 0 \end{bmatrix}$$

- 若令矩阵B=$Q^{\mathrm{T}}AP$，则A=QBP^{T}。由此得到

$$\|A-X\|_F=\|Q(B-\Omega)P^{\mathrm{T}}\|_F=\|B-\Omega\|_F$$

矩阵的最优近似

- 用 Ω 分块方法对B分块

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

- 其中B_{11}是 k x k 子矩阵，B_{12}是 k x (n-k) 的子矩阵，B_{21}是（m-k) x k 子矩阵，B_{22} 是（m-k) x (n-k)子矩阵。可得

$$\begin{aligned} \|A - X\|_F^2 &= \|B - \Omega\|_F^2 \\ &= \|B_{11} - \Omega_k\|_F^2 + \|B_{12}\|_F^2 + \|B_{21}\|_F^2 + \|B_{22}\|_F^2 \end{aligned}$$

矩阵的最优近似

- 现证$B_{12}=0$， $B_{21}=0$。用反证法。若$B_{12}\neq 0$，令

$$Y=Q\begin{bmatrix} B_{11} & B_{12} \\ 0 & 0 \end{bmatrix}P^{\mathrm{T}}$$

- 则 $Y\in\mathcal{M}$，且 $\|A-Y\|_F^2=\|B_{21}\|_F^2+\|B_{22}\|_F^2<\|A-X\|_F^2$
- 这与X的定义式 $\|A-X\|_F\leqslant\|A-A'\|_F=(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$ 矛盾

- 因此$B_{12}=0$， 同样可证$B_{21}=0$。于是

$$\|A-X\|_F^2=\|B_{11}-\Omega_k\|_F^2+\|B_{22}\|_F^2$$

矩阵的最优近似

- 再证 $B_{11}=\Omega_k$，为此令

$$Z=Q\begin{bmatrix} B_{11} & 0 \\ 0 & 0 \end{bmatrix}P^{\mathrm{T}}$$

- 则 $Z\in\mathcal{M}$，且 $\|A-Z\|_F^2=\|B_{22}\|_F^2\leqslant\|B_{11}-\Omega_k\|_F^2+\|B_{22}\|_F^2=\|A-X\|_F^2$
- 由 $\|A-X\|_F\leqslant\|A-A'\|_F=(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$ 知，$\|B_{11}-\Omega_k\|_F^2=0$，即 $B_{11}=\Omega_k$

- 最后看B_{22}。若（m-k) x (n-k)子矩阵B_{22}有奇异值分解 $U_1\Lambda V_1^{\mathrm{T}}$，则

$$\|A-X\|_F=\|B_{22}\|_F=\|\Lambda\|_F$$

矩阵的最优近似

- 证明 Λ 的对角线元素为A的奇异值。为此，令

$$U_2=\begin{bmatrix} I_k & 0 \\ 0 & U_1 \end{bmatrix}, \quad V_2=\begin{bmatrix} I_k & 0 \\ 0 & V_1 \end{bmatrix}$$

- 其中I_k是k阶单位矩阵，U_2，V_2的分块与B的分块一致注意到B及B_{22}的奇异值分解，即得

$$U_2^{\mathrm{T}}Q^{\mathrm{T}}APV_2=\begin{bmatrix} \Omega_k & 0 \\ 0 & \Lambda \end{bmatrix} \qquad A=(QU_2)\begin{bmatrix} \Omega_k & 0 \\ 0 & \Lambda \end{bmatrix}(PV_2)^{\mathrm{T}}$$

- 由此可知 Λ 的对角线元素为A的奇异值，故有

$$\|A-X\|_F=\|\Lambda\|_F \geqslant (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$$

- 可证

$$\|A-X\|_F=(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}=\|A-A'\|_F$$

矩阵的最优近似

- 在秩不超过k的 m x n 矩阵的集合中，存在矩阵A的弗罗贝尼乌斯范数意义下的最优近似矩阵X

- $A' = U\Sigma' V^{\mathrm{T}}$ 是达到最优值的一个矩阵

- 紧奇异值分解是在弗罗贝尼乌斯范数意义下的无损压缩
- 截断奇异值分解是有损压缩
- 截断奇异值分解得到的矩阵的秩为k，通常远小于原始矩阵的秩r，所以是由低秩矩阵实现了对原始矩阵的压缩。

矩阵的外积展开式

- 矩阵A的奇异值分解 $U\Sigma V^{\mathrm{T}}$ 也可以由外积形式表示

- 若将A的奇异值分解看成矩阵 $U\Sigma$ 和V^{T}的乘积，将 $U\Sigma$ 按列向量分块，将V^{T}按行向量分块，即得

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_n u_n \end{bmatrix}$$

$$V^{\mathrm{T}} = \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix}$$

矩阵的外积展开式

- 则 $A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_n u_n v_n^{\mathrm{T}}$ ⟸ A的外积展开式

- 即

$$u_i v_j^{\mathrm{T}} = \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{mi} \end{bmatrix} \begin{bmatrix} v_{1j} & v_{2j} & \cdots & v_{nj} \end{bmatrix} = \begin{bmatrix} u_{1i}v_{1j} & u_{1i}v_{2j} & \cdots & u_{1i}v_{nj} \\ u_{2i}v_{1j} & u_{2i}v_{2j} & \cdots & u_{2i}v_{nj} \\ \vdots & \vdots & & \vdots \\ u_{mi}v_{1j} & u_{mi}v_{2j} & \cdots & u_{mi}v_{nj} \end{bmatrix}$$

- A的外积展开式也可写为 $A = \sum_{k=1}^{n} A_k = \sum_{k=1}^{n} \sigma_k u_k v_k^{\mathrm{T}}$

清華大学 Tsinghua University

矩阵的外积展开式

- 由矩阵A的外积展开式知，若A的秩为n，则

$$A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_n u_n v_n^{\mathrm{T}}$$

- 设矩阵 $A_k = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_k u_k v_k^{\mathrm{T}}$
- 则A_k的秩为k，并且A_k是秩为k矩阵在弗罗贝尼乌斯范数意义A的最优近似矩阵
- 矩阵A_k就是A的截断奇异值分解

- 由于通常奇异值 σ_i 递减很快，所以k取很小值时，A_k也可以对A有很好的近似。

例

- 给出矩阵A

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

- 的秩为3， 求A的秩为2的最优近似

例

- 从前列已知

$$u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\sigma_1 = 4, \quad \sigma_2 = 3$$

- 于是得到 $A_2 = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ ，以此矩阵为A的最优近似