奇异值分解

奇异值分解是一种矩阵因子分解方法，是线性代数的内容，被引入到机器学习中。

定义与性质

定义

矩阵的奇异值分解是指，将一个非0的$m \times n$的实矩阵A，$A \in R^{m \times n}$，表示为以下三个实矩阵的乘积：

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

其中U是m**阶正交矩阵**，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**。也即：

$$\begin{aligned}
&UU^T = I \\
&VV^T = I \\
&\Sigma = diag(\sigma_1, \sigma_2 \cdots, \sigma_p) \\
&\sigma_1 \geqslant \sigma_2 \cdots \geqslant \sigma_p \geqslant 0 \\
&p = min(m, n)
\end{aligned}$$

奇异值分解并不要求矩阵A是方阵，事实上矩阵的奇异值分解可以看做是方阵对角化的推广。

奇异值分解基本定理

若A为一$m \times n$实矩阵，则A的奇异值分解存在$A = U\Sigma V^T$，其中U是m**阶正交矩阵**，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。**

构造性证明：

1. 确定V和Σ：

   A^TA是n阶实对称矩阵，所以存在一个n阶正交矩阵V实现A^TA的对角化，即有$V^T(A^TA)V = \Lambda$。其中A^TA的特征值都是非负的。事实上，令λ是其一个特征值，x是对应的特征向量，则

   $$\|Ax\|^2 = (Ax)^T(Ax) = x^TA^TAx = \lambda x^Tx = \lambda\|x\|^2$$

   于是有$\lambda = \dfrac{\|Ax\|^2}{\|x\|^2} > 0$。

   假定正交矩阵V的列的排列使得对应的特征值形成降序排列$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_n \geqslant 0$，计算特征值的平方根$\sigma_j = \sqrt{\lambda_j}$。设矩阵$A$的秩是$r$，则矩阵$A^TA$的秩也是$r$。由于$A^TA$是对称矩阵，它的秩等于特征值的个数。所以有$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_r > 0, \lambda_{r+1} = \lambda_{r+2} = \cdots \lambda_n = 0$，进一步地$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \sigma_r > 0, \sigma_{r+1} = \sigma_{r+2} = \cdots \sigma_n = 0$，那么$V$可以表示为$V = [V_1\ V_2]$，其中$V_1 = [v_1, v_2, \cdots v_r]$是非0特征值对应的特征向量，$V_2 = [v_{r+1}, v_{r+2}, \cdots v_n]$为特征值0对应的特征向量。

   这就是矩阵A的奇异值分解中的n阶正交矩阵V。

   令

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

所以矩阵A的奇异值分解中的$m \times n$矩形对角矩阵Σ如下：

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

下面退出后面要用到的一个公式。由以上知，V_2的列向量就是A^TA对应于特征值为0的特征向量，因此$A^TA\nu_j = 0, \quad j = r+1, \cdots, n$。因此有$AV_2 = 0$，而
$I = VV^T = [V_1\ V_2][V_1\ V_2]^T = V_1V_1^T + V_2V_2^T$，所以得到
$A = AI = AV_1V_1^T + AV_2V_2^T = AV_1V_1^T$。

2. 确定U

   令

$$u_j = \frac{1}{\sigma_j}Av_j, j = 1, 2, \cdots, r$$
$$U_1 = [u_1, u_2 \cdots, u_r]$$

   有$AV_1 = U_1\Sigma_1$。

   U_1的列向量构成了一组标准的正交基，证明如下：

$$\begin{aligned} u_i^Tu_j &= (\frac{1}{\sigma_i}v_i^TA^T)(\frac{1}{\sigma_j}Av_j) \\ &= \frac{1}{\sigma_i\sigma_j}v_i^T(A^TAv_j) \\ &= \frac{\sigma_j}{\sigma_i}v_i^Tv_j \\ &= \delta_{ij} \end{aligned}$$

   令$U_2 = \{u_{r+1}, u_{r+2} \cdots u_m\}$为$N(A^T)$的一组标准正交基，且令$U = \{U_1\ U_2\}$

3. 证明以上的$U\Sigma V^T = A$

奇异值分解的计算

求矩阵

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

的奇异值分解。

1. 求矩阵A^TA的特征值和特征向量。

$$A = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$$

   特征值λ和特征向量x满足方程

$$(A^TA - \lambda I)x = 0$$

   即为齐次线性方程组

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

解得到$\lambda_1=10,\lambda_2=0$。分别代入原方程可得对应的单位特征向量为
$v_1=[\frac{1}{\sqrt{2}}\ \ \frac{1}{\sqrt{2}}]^T,v_2=[\frac{1}{\sqrt{2}}\ \ -\frac{1}{\sqrt{2}}]^T$。

2. 求正交矩阵V和对角矩阵Σ。

   由上一步得到

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

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

   注意要注意Σ的形状，必要时加上0行向量使其能与U,V进行矩阵乘法。

3. 求正交矩阵U。

   基于矩阵A的正奇异值计算得到列向量u_1:

$$u_1=\frac{1}{\sigma_1}Av_1=\frac{1}{\sqrt{10}}\begin{bmatrix}1 & 1\\2 & 2\\0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{5}}\\ \frac{1}{\sqrt{5}}\\0\end{bmatrix}$$

   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。即求解以下方程组

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

   即是解以下方程

$$x_1+2x_2=0$$
$$x_1+2x_2=0$$

   分别取(x_2,x_3)为$(1,0)$和$(0,1)$，得到$N(A^T)$的基

$$(-2,1,0)^T,(0,0,1)^T$$

   标准化之后得到

$$u_2=(-\frac{2}{\sqrt{5}},\frac{1}{\sqrt{5}},0),u_3=(0,0,1)$$

   正交矩阵U如下

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

4. 综上所述

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

紧奇异值分解与截断奇异值分解

给定一个5×4的矩阵A与它的**完全奇异值分解**:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} = U\Sigma V^T = \begin{bmatrix} 0 & 0 & \sqrt{0.2} & 0 & \sqrt{0.8} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \sqrt{0.8} & 0 & \sqrt{0.2} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- **紧奇异值分解:**

  设有$m \times n$实矩阵A，其秩为$rank(A) = r, r \leqslant min(m, n)$，则称$U_r \Sigma_r V_r^T$为$A$的紧奇异值分解。$A = U\Sigma V^T$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} = U_r\Sigma_r V_r^T = \begin{bmatrix} 0 & 0 & \sqrt{0.2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{0.8} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \sqrt{5} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

- **截断奇异值分解:**

  设有$m \times n$实矩阵A，其秩为$rank(A) = r, r \leqslant min(m, n), 0 < k < r$，则称$U_k \Sigma_k V_k^T$为$A$的截断奇异值分解。$A \approx U\Sigma V^T$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} \approx U_k\Sigma_k V_k^T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

可以看出紧奇异值分解对应着无损压缩，截断奇异值分解对应着有损压缩。

奇异值分解与矩阵近似

弗罗贝尼乌斯范数

矩阵的弗罗贝尼乌斯范数是向量的L_2范数的直接推广，对应着机器学习中的平方损失函数，定义如下:

设矩阵$A \in R^{m\times n}, A = [a_{ij}]_{m\times n}$，定义矩阵$A$的**弗罗贝尼乌斯范数**为$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij})^2\right)^{\frac{1}{2}}$。

设矩阵$A \in R^{m\times n}$，奇异值分解为$U\Sigma V^T$，其中$\Sigma = diag(\sigma_1, \sigma_2 \cdots, \sigma_n)$，则

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

以上定理的证明仅仅用到正交变换不改变向量的长度，所以$\|A\|_F = \|U\Sigma V^T\|_F = \|\Sigma\|_F$。

矩阵的最优近似

在秩不超过k的$m \times n$的矩阵的集合中，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。而$A' = U\Sigma' V^T$是达到最优值的一个矩阵。

截断奇异值分解得到的矩阵的秩为k，通常小于原始矩阵的秩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$U\Sigma V^T$也可以由外积形式表示。事实上，我们可以将A的奇异值分解看成矩阵$U\Sigma$和V^T的乘积，将$U\Sigma$按列分块，将V^T按行向量分块如下：

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

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

则$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T = \sum\limits_{k=1}^{n} \sigma_k u_k v_k^T$。这就是$A$的外积展开式，其中$u_k v_k^T$为$m \times n$矩阵，是列向量$u_k$和行向量$v_k^T$的外积：

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

则我们可以定义秩为$n-1$的A_{n-1}如下：

$$A_{n-1} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_{n-1} u_{n-1} v_{n-1}^T = \sum_{k=1}^{n-1} \sigma_k u_k v_k^T$$

类似地我们可以定义秩为k的A_k如下：

$$A_{n-1} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$$

并且该矩阵就是在弗罗贝尼乌斯范数意义下的A的最优近似矩阵，也就是截断奇异值分解。