机器学习导论
第八章

王小航

奇异值分解

- 奇异值分解（Singular Value Decomposition, SVD）
- 从生物信息学到金融学等在内的很多应用中，SVD都是提取信息的强大工具
- 利用SVD实现，我们能用小得多的数据集来表示原始数据集，去除噪声和冗余信息，抽取相关特征

SVD的应用

- 奇异值分解是线性代数中重要的分解模型之一，在机器学习中随处可见，它的发展已经历经了上百年
- 但是最近几十年随着计算机的使用，我们发现了其更多的使用价值。
- 最早的SVD应用之一就是信息检索：隐形语义索引或隐性语义分析
- SVD 不仅仅应用在 PCA 、图像压缩、数字水印、 推荐系统和文章分类、 LSA (隐性语义分析)、特征压缩(或数据降维)中，在信号分解、信号重构、信号降噪、数据融合、同标识别、目标跟踪、故障检测和神经网络等方面也有很好的应用，是很多机器学习算法的基石

推荐系统

- 考虑图中给出的矩阵，它是由餐馆的菜和品菜师对这些菜的意见构成的。品菜师可以采用1到5之间的任意一个整数来对菜评级。如果品菜师没有尝过某道菜，则评级为0。

	鳗鱼饭	日式炸鸡排	寿司饭	烤牛肉	手撕猪肉
Ed	0	0	0	2	2
Peter	0	0	0	3	3
Tracy	0	0	0	1	1
Fan	1	1	1	0	0
Ming	2	2	2	0	0
Pachi	5	5	5	0	0
Jocelyn	1	1	1	0	0

矩阵补全

- 推荐系统可以看成为矩阵补全（matrix completion）问题，顾名思义就是将一个含有缺失值的矩阵通过一定的方法将其恢复为一个完全的矩阵

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

- $U\Sigma V^T$： 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$U\Sigma V^T$给出

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

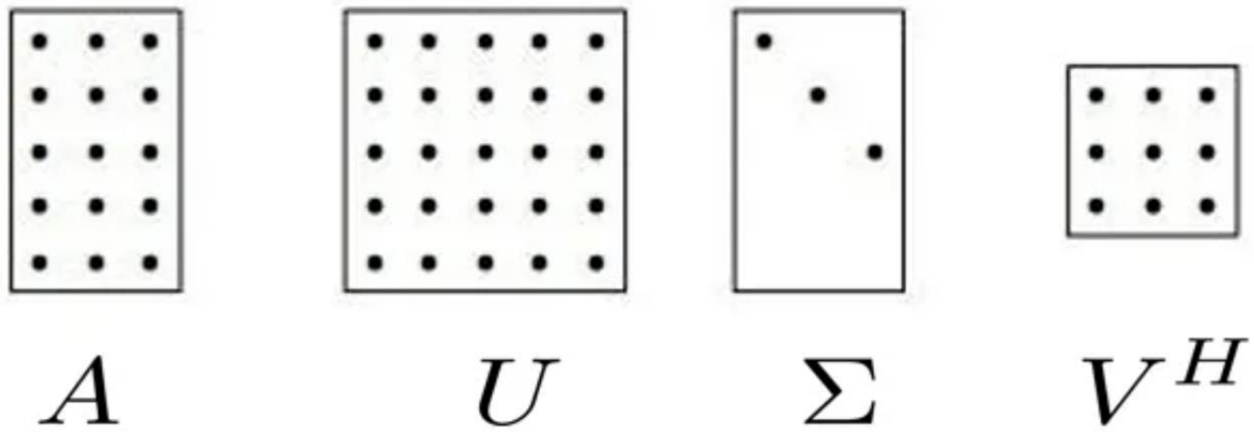
例

- 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



m行n列，m>n

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$U\Sigma V^T$也是A的一个奇异值分解

紧奇异值分解与截断奇异值分解

- $A = U\Sigma V^T$ 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。

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 $U_r\Sigma_r V_r^T$ 为A的紧奇异值分解（compact singular value decomposition)，即

$$A = U_r\Sigma_r V_r^T$$

- 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

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

例

- A的紧奇异值分解是$A = U_r \Sigma_r V_r^T$

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

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 $\mathrm{rank}(A) = r$，且 $0 < k < r$，则称 $U_k \Sigma_k V_k^{\mathrm{T}}$ 为矩阵 A 的截断奇异值分解（truncated singular value decomposition）

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

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

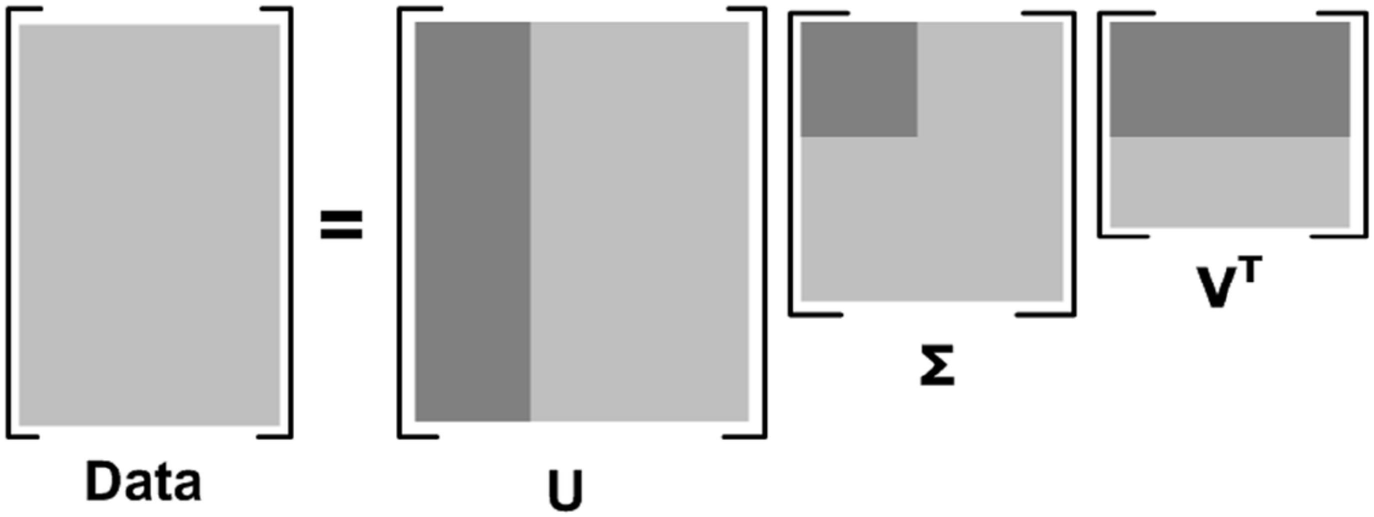
- 矩阵A的秩为3

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

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

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

例



主要性质

- （1）设矩阵A的奇异值分解为 $A = U\Sigma V^T$ ，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$A^{\mathrm{T}}A$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$A^{\mathrm{T}}A$的特征向量
- U的列向量是AA^{T}的特征向量
- Σ的奇异值是$A^{\mathrm{T}}A$和AA^{T}的特征值的平方根。

主要性质

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

- （3）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）。

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$A^T A$的特征值和特征向量得到。
- $A^T A$的特征向量构成正交矩阵V的列
- $A^T A$的特征值λ_j的平方根为奇异值σ_i，即

$$\sigma_j = \sqrt{\lambda_j}, \qquad j = 1,2, \dots, n$$

- 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$A = U\Sigma V^T$

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$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$
- 将特征值$\lambda_i(i=1,2,\ldots,n)$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:

$$V=(v_1,v_2,\ldots,v_n)$$

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

$$\sigma_i = \sqrt{\lambda_i}, \qquad i = 1,2,\ldots,n$$

- 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

$$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$$

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$u_j = \frac{1}{\sigma_j} A v_j$， $j = 1,2,\dots,r$
- 得到$U_1 = (u_1, u_2, \dots, u_r)$
- 求A^T的零空间的一组标准正交基$\{u_{r+1}, u_{r+2}, \dots, u_m\}$，令
$$U_2 = (u_{r+1}, u_{r+2}, \dots, u_m)$$
- 并令$U = (U_1, U_2)$
- （5）得到奇异值分解$A = U\Sigma V^T$

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$^{\mathrm{T}}$A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

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

例

- 该方程有非零解的充要条件是

$$\begin{vmatrix} 5-\lambda & 5 \\ 5 & 5-\lambda \end{vmatrix} = 0 \qquad \lambda^2 - 10\lambda = 0$$

- 解此方程，得矩阵A$^{\mathrm{T}}$A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

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

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

- N(A^{T})的一组标准正交基是$u_2 = \left(-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right)^T, u_3 = (0,0,1)^T$

- 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

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

用python实 现SVD

```
>>> from numpy import *
>>> U,Sigma,VT=linalg.svd([[1, 1],[7, 7]])

>>> U
array([[-0.14142136, -0.98994949],
[-0.98994949,  0.14142136]])
>>> Sigma
array([ 10.,    0.])
>>> VT
array([[-0.70710678, -0.70710678],
[-0.70710678,  0.70710678]])
```

SVD在图像压缩中的应用

```
def img_compress(img,k):
    U,s,VT=np.linalg.svd(img)
    Sigma = np.diag(s)
    D = U[:,:k]@Sigma[:k,:k]@VT[:k,:]
    D[D<0] = 0
    D[D>255] = 255
    return np.rint(D).astype('uint8')
```

参考网址

- https://blog.csdn.net/qq_36523839/article/details/82347332
- https://zhuanlan.zhihu.com/p/480389473

拓展：研究领域

- 矩阵补全和奇异值软阈值算法
- https://zhuanlan.zhihu.com/p/93400890