§3 奇异值分解

3.1 引言

- 对角阵是我们最喜爱的一类矩阵.
- n阶矩阵A相似于对角阵 $\Leftrightarrow$A存在n个线性无关的特征向量.
- 实对称矩阵正交相似于对角阵

3.1 引言

问题:如何“对角化”$m \times n$矩阵?

3.2 奇异值分解(Singular Value Decomposition)

设A是一个$m \times n$矩阵, 则存在m阶正交矩阵U和n阶正交矩阵V, 满足

$$A = U \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}_{m\times n} V^T =: U\Sigma V^T,$$

其中r = rankA. 习惯上,设 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. 称$\sigma_1, \cdots, \sigma_r$为奇异值(singular value). 称U和V的前r列向量为奇异向量(singular vector).

这个分解为奇异值分解, 简称**SVD**.

3.2 奇异值分解(Singular Value Decomposition)

令 $U = (\mathbf{u}_1 \cdots \mathbf{u}_r \mathbf{u}_{r+1} \cdots \mathbf{u}_m), V = (\mathbf{v}_1 \cdots \mathbf{v}_r \mathbf{v}_{r+1} \cdots \mathbf{v}_n).$

$$\text{SVD} \Rightarrow \begin{array}{l} A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \\ A^T\mathbf{u}_i = \sigma_i \mathbf{v}_i, \end{array} (i = 1, \cdots, r) \quad \begin{array}{l} A\mathbf{v}_j = \mathbf{0}(j = r+1, \cdots, n), \\ A^T\mathbf{u}_k = \mathbf{0}(k = r+1, \cdots, m). \end{array}$$

$$\Rightarrow \quad A^T A\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i, \quad AA^T\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i \quad (1 \le i \le r).$$

3.2 奇异值分解(Singular Value Decomposition)

设A是秩为r的$m \times n$ 实矩阵,则AA^T为m阶实对称矩阵, A^TA 为n阶实对称矩阵.

(1) AA^T 与A^TA的特征值为非负数.

证明: 设 $A^TA\mathbf{x} = \lambda\mathbf{x}(\mathbf{x} \neq \mathbf{0})$, 则 $\mathbf{x}^TA^TA\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x}$, 即

$$||A\mathbf{x}||^2 = \lambda||\mathbf{x}||^2,$$

故 $\lambda \geq 0$. 同理 AA^T 的特征值也全是非负数.

3.2 奇异值分解(Singular Value Decomposition)

(2) AA^T与A^TA的非零特征值集合相同.

证明: 因为$r(AA^T) = r(A^TA) = r(A) = r$, 故

$$\sharp\{A^TA\text{的非零特征值}\} = r = \sharp\{AA^T\text{的非零特征值}\}$$

设λ是A^TA的非零特征值. 即$\exists\mathbf{x}$, 使得$A^TA\mathbf{x} = \lambda\mathbf{x}$.
则有$AA^TA\mathbf{x} = \lambda A\mathbf{x}$. 故$\lambda$也是$AA^T$的非零特征值. 反之亦然.
因此,AA^T 与A^TA具有相同的非零特征值.

3.2 奇异值分解(Singular Value Decomposition)

不妨设AA^T和A^TA的这r个非零特征值为$\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0$,其中$\sigma_i > 0$.

(3) 设 $\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{R}^n$ 为n阶实对称方阵A^TA的单位正交特征向量,

$$A^TA(\mathbf{v}_1 \cdots \mathbf{v}_n) = (\mathbf{v}_1 \cdots \mathbf{v}_n)\begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \end{pmatrix}.$$

记$V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$, 则$V^TV = I_n$.

3.2 奇异值分解(Singular Value Decomposition)

注意到 $A^TA\mathbf{v}_i = \sigma_i^2\mathbf{v}_i, \quad (1 \le i \le r)$.
故 $\mathbf{v}_i^TA^TA\mathbf{v}_i = \sigma_i^2\mathbf{v}_i^T\mathbf{v}_i$, 即 $||A\mathbf{v}_i||^2 = \sigma_i^2$.

令 $\mathbf{u}_i := \dfrac{A\mathbf{v}_i}{\sigma_i} \in \mathbb{R}^m (1 \le i \le r)$, 则 $AA^T\mathbf{u}_i = \sigma_i^2\mathbf{u}_i$, 并且

$$\mathbf{u}_i^T\mathbf{u}_j = \frac{(A\mathbf{v}_i)^T}{\sigma_i}\frac{A\mathbf{v}_j}{\sigma_j} = \frac{\mathbf{v}_i^T(A^TA\mathbf{v}_j)}{\sigma_i\sigma_j} = \frac{\sigma_j^2\mathbf{v}_i^T\mathbf{v}_j}{\sigma_i\sigma_j} = \frac{\sigma_j}{\sigma_i}\delta_{ij} = \delta_{ij}.$$

故 $\{\mathbf{u}_i \mid 1 \le i \le r\}$ 是 AA^T 的单位正交特征向量.

3.2 奇异值分解(Singular Value Decomposition)

又

$$\boxed{A\mathbf{v}_i = \sigma_i \mathbf{u}_i,} \qquad \boxed{A^T\mathbf{u}_i = \sigma_i \mathbf{v}_i,} \qquad 1 \le i \le r.$$

故

$\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 为 $C(A)$ 的一组单位正交基;

$\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ 为 $C(A^T)$ 的一组单位正交基.

3.2 奇异值分解(Singular Value Decomposition)

$$A(\mathbf{v}_1 \cdots \mathbf{v}_r) = (\mathbf{u}_1 \cdots \mathbf{u}_r) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

记为 $A_{m\times n}V_{n\times r} = U_{m\times r}\Sigma_{r\times r}$.

回顾 $\mathbb{R}^n = C(A^T) \oplus N(A), \mathbb{R}^m = C(A) \oplus N(A^T)$,
扩充$\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$为$\mathbb{R}^n$的一组基$\{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$,
扩充$\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$为$\mathbb{R}^m$的一组基$\{\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_m\}$.

3.2 奇异值分解(Singular Value Decomposition)

$$\Rightarrow \boxed{A(\mathbf{v}_1 \cdots \mathbf{v}_r \mathbf{v}_{r+1} \cdots \mathbf{v}_n) = (\mathbf{u}_1 \cdots \mathbf{u}_r \mathbf{u}_{r+1} \cdots \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}_{m\times n}}$$

即$\boxed{A_{m\times n}V_{n\times n} = U_{m\times m}\Sigma_{m\times n}}$，其中$V^TV = I_n, U^TU = I_m.$

$$\Rightarrow \boxed{\begin{aligned} &A = U\Sigma V^T, \qquad (SVD) \\ &A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T. \end{aligned}}$$

3.3 例题

例: 求矩阵 $A=\begin{pmatrix}1&0\\0&1\\1&-1\end{pmatrix}$ 的奇异值分解.

解:

因 $A^TA=\begin{pmatrix}1&0&1\\0&1&-1\end{pmatrix}\begin{pmatrix}1&0\\0&1\\1&-1\end{pmatrix}=\begin{pmatrix}2&-1\\-1&2\end{pmatrix}$,

故A^TA的特征多项式为

$$|\lambda I-A^TA|=\begin{vmatrix}\lambda-2&1\\1&\lambda-2\end{vmatrix}=(\lambda-1)(\lambda-3).$$

3.3 例题

因此A^TA的特征值为$\sigma_1^2=3,\sigma_2^2=1.$ 它们相应的单位特征向量分别为

$$\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix},\qquad \mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

AA^T 的三个特征值分别为$\sigma_1^2=3,\sigma_2^2=1,0,$ 且属于它们的单位特征向量分别为

$$\mathbf{u}_1=\frac{1}{\sigma_1}A\mathbf{v}_1=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\0\end{pmatrix},\quad \mathbf{u}_2=\frac{1}{\sigma_2}A\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\2\end{pmatrix},\quad \mathbf{u}_3=\frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\-1\end{pmatrix}.$$

3.3 例题

令 $V = (\mathbf{v}_1\ \mathbf{v}_2)$, $\quad U = (\mathbf{u}_1\ \mathbf{u}_2, \mathbf{u}_3)$, $\quad \Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$,

则$A = U\Sigma V^T = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$

3.3 例题

例 (奇异值分解的几何意义):

设 $A=\begin{pmatrix}3 & -4\\ 5 & 0\end{pmatrix}$, 则线性变换$\mathbf{x}\mapsto A\mathbf{x}$ 把$\mathbb{R}^2$ 中的单位圆周

$S^1=\{(x,y)\in\mathbb{R}^2 \mid x^2+y^2=1\}$ 映成$\mathbb{R}^2$ 中的一个椭圆.

事实上, 由于$A^TA=\begin{pmatrix}34 & -12\\ -12 & 16\end{pmatrix}$的特征值为$\sigma_1^2=40,\sigma_2^2=10,$

它们相应的单位特征向量分别为

3.3 例题

$$\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \qquad \mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

AA^T 的特征值也为$\sigma_1^2 = 40, \sigma_2^2 = 10$,且属于它们的单位特征向量分别为

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

令 $V = (\mathbf{v}_1 \ \mathbf{v}_2), \quad U = (\mathbf{u}_1 \ \mathbf{u}_2), \quad \Sigma = \begin{pmatrix} \sigma_1 & \\ & \sigma_2 \end{pmatrix} = \begin{pmatrix} 2\sqrt{10} & \\ & \sqrt{10} \end{pmatrix},$

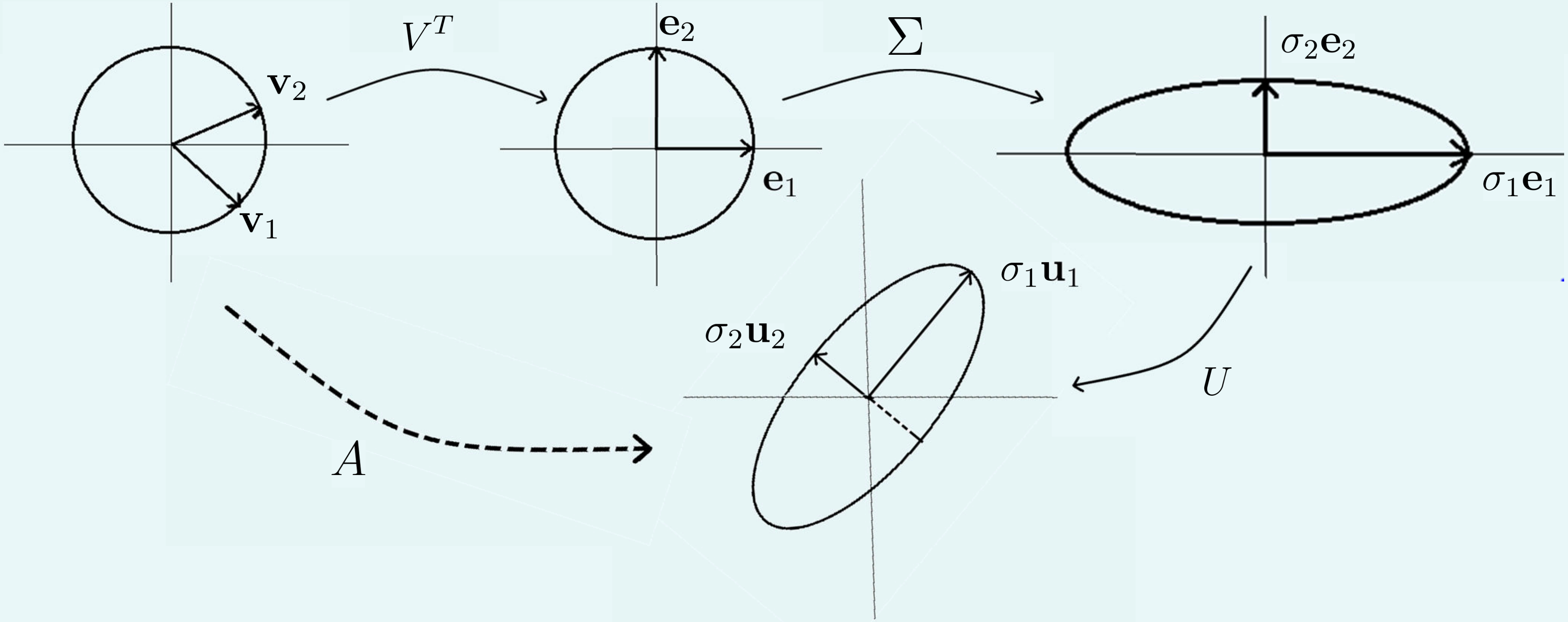
3.3 例题

则A的奇异值分解为

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

3.3 例题

矩阵A在单位圆周S^1上的作用可被分解成三步:



3.3 例题

注: 一般地, 设秩为r的$m \times n$矩阵A有SVD: $A_{m\times n} = U\Sigma V^T$, 则从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换$\mathbf{x} \mapsto A\mathbf{x}$可看成是以下三步的复合:

① $\mathbb{R}^n$中的旋转$\mathbf{x} \mapsto V^T\mathbf{x}$,

② $\mathbb{R}^n$中的向量$V^T\mathbf{x}$的前r个分量做伸缩,其余分量变为0:

$$V^T\mathbf{x} \mapsto \Sigma V^T\mathbf{x},$$

③ 再在$\mathbb{R}^m$中做旋转$\Sigma V^T\mathbf{x} \mapsto U\Sigma V^T\mathbf{x}$.

3.4 奇异值分解的应用

(1) SVD与矩阵的四个基本子空间:

设 $A = U\Sigma V^T$ 是$m \times n$实矩阵A的奇异值分解, r = r(A), 则

- 正交矩阵U的前r列是C(A)的一组标准正交基;
- 正交矩阵U的后m-r列是$N(A^T)$ 的一组标准正交基;
- 正交矩阵V的前r列是$C(A^T)$ 的一组标准正交基;
- 正交矩阵V的后n-r列是N(A)的一组标准正交基.

3.4 奇异值分解的应用

(2) SVD与图像压缩:

设秩 r 的 $m \times n$ 矩阵A的奇异值分解为

$$A = U\Sigma V^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T,$$

其中$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$

令 $A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (1 \leq k < r),$

称为A的k阶逼近.特别地, k = 1时, A_1是1阶逼近.

3.4 奇异值分解的应用

例: 一幅规格是$m \times n$像素的照片可用一个$m \times n$矩阵来储存.
利用矩阵的奇异值分解, 只需存储矩阵的奇异值σ_i, 奇异向量$\mathbf{u}_i, \mathbf{v}_i$的分量, 总计 $r \cdot (m+n+1)$ 个数据, 而不是原始的$m \cdot n$ 个数据.
通常$r \ll m, r \ll n$, 则$r \cdot (m+n+1) \ll m \cdot n$.

比值$\dfrac{m \cdot n}{r \cdot (m+n+1)}$称为图像的压缩比 (其倒数称为数据压缩率).

若 $\sigma_1, \cdots \sigma_k$ 远大于$\sigma_{k+1}, \cdots \sigma_r$, 则$A_k \approx A$, 图像不失真且压缩了存储量.

3.4 奇异值分解的应用

(3) SVD与特征值:

命题: 设$|\lambda|_{\max}$是矩阵A的特征值的模长的最大值, 则

$$\sigma_1 \geq |\lambda|_{\max}, \sigma_1 \geq |a_{ij}|, \quad \forall i, j.$$

证明: 设A有奇异值分解 $A = U\Sigma V^T$, 则对任意向量**x**, 有

$$||A\mathbf{x}|| = ||U\Sigma V^T\mathbf{x}|| = ||\Sigma V^T\mathbf{x}|| \leq \sigma_1||V^T\mathbf{x}|| = \sigma_1||\mathbf{x}||.$$

3.4 奇异值分解的应用

若$A\mathbf{x} = \lambda\mathbf{x}$, 则 $||A\mathbf{x}|| = |\lambda| \cdot ||\mathbf{x}||$, 故 $\sigma_1 \geq |\lambda|$, 特别有$\sigma_1 \geq |\lambda|_{\max}$.

又若取 $\mathbf{x} = (1, 0, \cdots, 0)$, 则A**x**表示A的第一列向量, 且
$||A\mathbf{x}|| \leq \sigma_1 ||\mathbf{x}|| = \sigma_1$, 而

$$|a_{i1}| \leq \sqrt{a_{11}^2 + \cdots + a_{n1}^2} \leq \sigma_1.$$

同理, $|a_{ij}| \leq \sigma_1$.

3.4 奇异值分解的应用

命题 (奇异值与奇异矩阵)**:** 矩阵A列满秩 $\Leftrightarrow$ A的奇异值均非零.
特别地, 方阵A非奇异 $\Leftrightarrow$ A的奇异值均非零.

注： 矩阵的奇异值较之其特征值的一个优点是非零奇异值的个数恰好是矩阵的秩，而矩阵的非零特征值的个数一般比其秩小（如幂零矩阵无非零特征值），故常以此来计算矩阵的秩.