1. 奇异值分解(SVD)

Collected by Sizhe Zhou
2022. 4. 8.

1.1 奇异值分解的定义与性质

1.1.1 定义与定理

定义 1.1: $A \in \mathbb{R}^{m\times n}$， $A = U\Sigma V^T$

其中 U 为 m 阶 orthogonal matrix ($\mathbb{R}^{m\times m}$)

V 为 n 阶正交矩阵 ($\mathbb{R}^{n\times n}$)

Σ 是由降序排列的非负的 $m\times n$ 对角矩阵

s.t. $UU^T = \mathbb{1}$, $VV^T = \mathbb{1}$, $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_p)$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, $p = \min(m, n)$

σ_i 称为 A 的 singular value

U 的列向量称为 left singular vector

V 的列向量称为 right singular vector

(分解不唯一)

定理 1.1 (SVD基本定理): $A \in \mathbb{R}^{m\times n}$，如上的 SVD 存在

构造性证明：（假设 $m > n$）

(1) 确定 V 和 Σ：

A^TA 为 n 阶实对称 $\Rightarrow$ A^TA 的 eigenvalues 为非负R且存在 n 阶正交实矩阵 V，s.t. $V^T(A^TA)V = \Lambda$ where Λ 为 n 阶对角，其对角元素由 A^TA 的特征值组成

令 λ 是 A^TA 的一个特征值，x 是对应特征向量，则

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

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

可 assume V 的列排列使相应 eigen values 形成

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

then let $\sigma_j = \sqrt{\lambda_j}$, $j = 1, 2, \cdots, n$ (即 A 的奇异值)

设 $\text{rank}(A) = r$, $\Rightarrow$ $\text{rank}(A^TA) = r$. 由于 A^TA 为对称矩阵,

$\text{rank}(A) = \#$ positive λ_i.

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

对应: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$

令 $V_1 = [v_1 \ \ v_2 \ \cdots \ v_r]$ $\quad V_2 = [v_{r+1} \ \ v_{r+2} \ \cdots \ v_n]$

v_i 为 λ_i 对应的特征向量

$\Rightarrow V = [V_1 \ \ V_2]$ (此即 A 的 SVD 中的 V)

令 $\Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$

$m \times n$ 矩形对角阵 $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$ (即 A 的 SVD 中的 Σ)

$A^TA v_j = 0$, $j = r+1, \cdots, n$

于是 $\text{span}(v_{r+1}, \cdots, v_n) =$ null space of $A^TA =: N(A^TA)$

而 $N(A^TA) = N(A)$ $\Rightarrow$ $v_{r+1}, \cdots, v_n$ 构成 $N(A)$ 的标准正交基

$\Rightarrow AV_2 = 0$

由于 V 为正交矩阵,

$$\mathbb{1} = VV^T = V_1V_1^T + V_2V_2^T$$

$$A = A\mathbb{1} = AV_1V_1^T + AV_2V_2^T = AV_1V_1^T$$

(2) 确定 U

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

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

则有 $AV_1 = U_1 \Sigma_1$

U_1 的列向量构成了一组标准正交集，因为

$$u_i^T u_j = \left(\frac{1}{\sigma_i} v_i^T A^T\right)\left(\frac{1}{\sigma_j} A v_j\right)$$

$$= \frac{1}{\sigma_i \sigma_j} v_i^T (A^T A v_j)$$

$$= \frac{\sigma_j}{\sigma_i} v_i^T v_j = \delta_{ij} \ , \ i=1,2,\cdots,r; \ j=1,2,\cdots,r$$

由上可知，$u_1 \sim u_r$ 构成 $R(A)$ 的一组标准正交基

$\dim R(A) = r$， $\dim R(A)^\perp = m-r$。

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

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

$U = [U_1 \ U_2]$ 为 m 阶正交矩阵 (A 的 SVD 中的 U)

(3) 证明 $U\Sigma V^T = A$

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

$$= U_1 \Sigma_1 V_1^T = A V_1 V_1^T = A$$

$\Rightarrow$ A 存在 SVD

1.1.2 紧奇异值分解与截断奇异值

1. 紧奇异值分解 $\rightarrow$ 无损压缩

定义1.2： $A \in \mathbb{R}^{m\times n}$，$rank(A)=r$，$[r < \min(m,n)]$

称 $U_r\Sigma_r V_r^T$ 为A的 compact SVD，$A=U_r\Sigma_r V_r^T$

U_r 为 $m\times r$，V_r 为 $n\times r$，Σ_r 为 r 阶对角

(前 r 列 of U)　(前 r 列 of V)　(前 r 列 of Σ)

$rank(\Sigma_r)=r$ obviously

2. 截断奇异值分解 → 有损压缩

定义1.3： $A \in \mathbb{R}^{m\times n}$，$rank(A)=r$，$0<k<r$，

$U_k\Sigma_k V_k^T$ 为 A的 truncated SVD

$A \approx U_k\Sigma_k V_k^T$

U_k 为 $m\times k$，V_k 为 $n\times k$，Σ_k 为 k 阶对角

(前 k 列)　(前 k 列)　(前 k 列)

$rank(U_k\Sigma_k V_k^T) < r$

1.1.3 几何解释

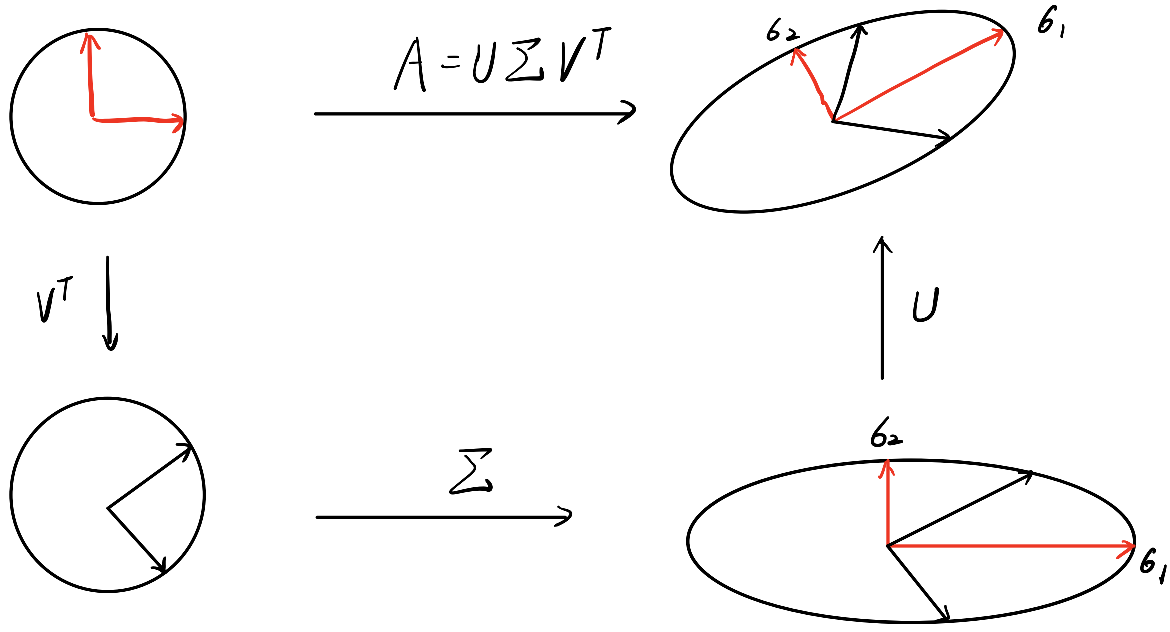
· $A \in \mathbb{R}^{m\times n}$ 表示从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换

· 线性变换可分解为 $\begin{cases} \text{一个坐标系的旋转或反射变换} \\ \text{一个坐标轴的缩放变换} \\ \text{另一个坐标系的旋转或反射变换} \end{cases}$

· $v_1 \sim v_n$ 构成 $\mathbb{R}^n$ 的标准正交基，表示 $\mathbb{R}^n$ 中正交坐标系的旋转或反射变换；

$u_1 \sim u_m$ 同理；

Σ 中 $\sigma_1,\sigma_2,\cdots,\sigma_n$ 表示 $\mathbb{R}^n$ 中原始正交坐标系坐标轴的 $\sigma_1,\sigma_2,\cdots,\sigma_n$ 倍缩放变换.



1.1.4 主要性质

(1) $A=U\Sigma V^T$，则有

$$A^TA = V(\Sigma^T\Sigma)V^T$$

$$AA^T = U(\Sigma\Sigma^T)U^T$$

即 A^TA，AA^T 的特征分解存在，且可由 A 的 SVD 表示

V 的列向量为 A^TA 的特征向量，U 的列向量为 AA^T 的特征向量

Σ 的奇异值是 A^TA 和 AA^T 的特征值的平方根

(2) $AV=U\Sigma$，比较第 j 列，

$\Rightarrow\ Av_j=\sigma_j u_j$ ，$j=1,2,\cdots,n$

类似地，$A^TU=V\Sigma^T$

$\Rightarrow\ A^Tu_j=\sigma_j v_j$，$j=1,2,\cdots,n$

$A^Tu_j=0$，$j=n+1,n+2,\cdots,m$

(3) 奇异值 $\sigma_1,\sigma_2,\cdots,\sigma_n$ 唯一，而 U, V 不唯一

(4) $\text{rank}(A)=\text{rank}(\Sigma)=\#\ \text{positive}\ \sigma_i$ (包含重复)

(5) A 的 r 个右奇异向量 $v_1,v_2,\cdots,v_r$ 构成 $R(A^T)$ 的一组标准

正交基。

A的 $n-r$ 个右奇异向量 $v_{r+1}, v_{r+2}, \cdots, v_n$ 构成 $N(A)$ 的一组标准正交基

A的 r 个左奇异向量 $u_1, u_2, \cdots, u_r$ 构成 $R(A)$ 的一组标准正交基。

A的 $m-r$ 个左奇异向量 $u_{r+1}, u_{r+2}, \cdots, u_m$ 构成 $N(A^T)$ 的一组标准正交基

1.2 奇异值分解的计算

(1) 首先求 A^TA 的特征值与特征向量

$W = A^TA$

求解 $(W - \lambda I)x = 0$

将得到的特征值由大到小排列:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

将 λ_i 代入特征方程求得对应特征向量

(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 $V = [v_1 \ v_2 \ \cdots \ v_n]$

(3) 求 $m \times n$ 对角矩阵 Σ

计算A的奇异值 $\sigma_i = \sqrt{\lambda_i}$, $i = 1, 2, \cdots, n$

$\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$ 其余元素为 0

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

求 A^T 的零空间的一组标准正交基 $\{u_{r+1}, u_{r+2}, \cdots u_m\}$

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

令 $U = [U_1 \ U_2]$

(5) $A = U\Sigma V^T$

·实际应用中 SVD 是通过求 A^TA 特征值进行的，但不直接计算 A^TA.

1.3 实际应用的奇异值分解与矩阵近似

1.3.1 Frobenius norm

定义 1.4： $A \in \mathbb{R}^{m\times n}$,

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

引理 1.1： $A = U\Sigma V^T$, $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$

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

证明：若 Q 为 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 为 n 阶正交矩阵，则有

$$\|AP^T\|_F = \|A\|_F$$

故 $\|A\|_F = \|U\Sigma V^T\|_F = \|\Sigma\|_F$

1.3.2 矩阵的最优近似

定理 1.2： $A \in \mathbb{R}^{m\times n}$, $rank(A) = r$, 设 M 为 $\mathbb{R}^{m\times n}$ 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 $X \in M$，使得

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

称矩阵 X 为 A 在 Frobenius norm 下的最优近似

定理1.3：若秩为k的 $X\in M$，满足

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

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

特别地，若 $A'=U\Sigma'V^T$，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 $\|A-A'\|_F = (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}} = \min_{S\in M}\|A-S\|_F$

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

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

设 $X=Q\Omega P^T$，其中

$$\Omega = \begin{bmatrix} w_1 & & & & \\ & \ddots & & 0 & \\ & & w_k & & \\ & & & 0 & \\ & 0 & & & \ddots \\ & & & & 0 \end{bmatrix} = \begin{bmatrix} \Omega_k & 0 \\ 0 & 0 \end{bmatrix}$$

若令 $B=Q^TAP$，则 $A=QBP^T$，

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

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} 为 $k\times k$ 子矩阵

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

用反证法证 $B_{12}, B_{21}=0$:

若 $B_{12}\neq 0$，令 $Y = Q\begin{bmatrix} B_{11} & B_{12} \\ 0 & 0 \end{bmatrix} P^T$

则 $Y\in M$，且 $\|A-Y\|_F^2 = \|B_{21}\|_F^2 + \|B_{22}\|_F^2 < \|A-X\|_F^2$

$\Rightarrow$ 矛盾 $\Rightarrow B_{12}=0$，同理 $B_{21}=0$

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

再证 $B_{11}=\Omega_k$. 为此令

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

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

$$\Rightarrow \|B_{11}-\Omega_k\|_F^2 = 0 \Rightarrow B_{11}=\Omega_k$$

最后看 B_{22}. 若 $(m-k)\times(n-k)$ 的 $B_{22}=U_1\Lambda V_1^T$, 则

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

证明 Λ 的对角元素为 A 的奇异值。为此，令

$$U_2 = \begin{bmatrix} \mathbb{1}_k & 0 \\ 0 & U_1 \end{bmatrix}, \quad V_2 = \begin{bmatrix} \mathbb{1}_k & 0 \\ 0 & V_1 \end{bmatrix}$$

U_2，V_2 的分块与 B 的分块一致。注意 B 及 B_{22} 的 SVD

$$U_2^T Q^T A P V_2 = \begin{bmatrix} \Omega_k & 0 \\ 0 & \Lambda \end{bmatrix}$$

$$A = (QU_2)\begin{bmatrix} \Omega_k & 0 \\ 0 & \Lambda \end{bmatrix}(PV_2)^T$$

由此知 Λ 的对角元素为 A 的奇异值，故有

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

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

以上定理表明，在秩不超过 k 的 $m\times n$ 的矩阵集合中，存在矩阵 A 的 Frobenius norm 意义下的最优近似矩阵 X。$A'=U\Sigma'V^T$ 是达到最优值的一个矩阵。

1.3.3 矩阵的外积展开式

- $U\Sigma = [\sigma_1 u_1 \quad \sigma_2 u_2 \quad \cdots \quad \sigma_n u_n]$

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

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

称为 A 的外积展开式，其中 $u_k v_k^T$ 为 $m\times n$ 矩阵，为列向量 u_k 和行向量 v_k^T 的外积

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

- A 的外积展开式也可写成

$$A = \sum_{k=1}^{n} A_k = \sum_{k=1}^{n} \sigma_k u_k v_k^T$$

其中 $A_k = \sigma_k u_k v_k^T$ 是 $m\times n$ 矩阵

由以上知，若 A 的秩为 n，则

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T$$

一般地，设矩阵 $A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$

则 A_k 的秩为 k，且 A_k 是秩为 k 的矩阵中在 Frobenius norm 意义下 A 的最优近似矩阵。A_k 为 A 的截断 SVD

· 通常 σ_i 递减很快，所以 k 取值很小时，A_k 对 A 也有很好的近似