Numerical Linear Algebra

Lecture Notes 4:Singular Value Decomposition

Professor: Zhihua Zhang

1 Motivation

Square matrix has eigenvalues, but rectangular matrix has not. Here is why singular value decomposition (SVD) comes. Consider $\mathbf{A} \in \mathbb{R}^{m\times n}(m \geq n)$, $\text{rank}(\mathbf{A}) = k(k \leq n)$.

$$\mathbf{A}\mathbf{A}^T\mathbf{Y} = \mathbf{Y}\mathbf{\Lambda},$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_k)$.

Let $\mathbf{Z} \triangleq \mathbf{A}^T\mathbf{Y}\mathbf{\Sigma}^{-1/2}$, then $\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_k$. From the definition of $\mathbf{Z}$, we can get $\text{span}(\mathbf{Z}) \subset \text{span}(\mathbf{A}^T)$. $\mathbf{A}\mathbf{Z} = \mathbf{Y}\mathbf{\Lambda}^{1/2}$. Since $\text{rank}(\mathbf{Z}) = \text{rank}(\mathbf{A})$, $\text{span}(\mathbf{Z}) = \text{span}(\mathbf{A}^T)$. Let $\mathbf{Z}_0$ in $\mathbf{Z}$'s complementary space, $\hat{\mathbf{Z}} = (\mathbf{Z} \quad \mathbf{Z}_0)$ and $\hat{\mathbf{Z}}^T\hat{\mathbf{Z}} = \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T = \mathbf{I}_n$.

$$\mathbf{A}\hat{\mathbf{Z}} = (\mathbf{A}\mathbf{Z} \quad \mathbf{A}\mathbf{Z}_0) \qquad \mathbf{A}\mathbf{Z}_0 = \mathbf{0}$$

$$\mathbf{A} = \mathbf{A}\hat{\mathbf{Z}}\hat{\mathbf{Z}}^T = \mathbf{Y}\mathbf{\Lambda}^{1/2}\mathbf{Z}^T$$

$\mathbf{Y}$ is a column-orthogonal matrix, $\mathbf{Z}$ is a column-orthogonal matrix and $\mathbf{\Lambda}^{1/2}$ is a diagonal matrix.

2 SVD theorem

Theorem 2.1 (SVD Theorem)**.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}(m \geq n)$ be a nonzero matrix, then there exist orthogonal matrices $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_m] \in \mathbb{R}^{m\times n}$ and $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_n] \in \mathbb{R}^{n\times n}$, such that*

$$\mathbf{U}^T\mathbf{A}\mathbf{V} = \mathbf{\Sigma} = diag(\sigma_1, ..., \sigma_p) \in \mathbb{R}^{m\times n},$$

where $p = \min\{m, n\}$ when $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p \geq 0$.

Proof via induction.

1. When $n = 1$, it is surely correct.

2. Suppose the theorem holds when $\mathbf{A} \in \mathbb{R}^{(m-1)\times(n-1)}$. Now consider $\mathbf{A} \in \mathbb{R}^{m\times n}$. If $\mathbf{A} = 0$, then it is correct. If $\mathbf{A} \neq 0$, then $\sigma_1 = |||\mathbf{A}|||_2 = \max_{||\mathbf{v}||_2=1} ||\mathbf{A}\mathbf{v}||_2 \neq 0$. Let $\mathbf{v}_1 = \arg\max_{||\mathbf{v}||_2=1} ||\mathbf{A}\mathbf{v}||_2$, $\mathbf{u}_1 = \frac{\mathbf{A}\mathbf{v}_1}{\sigma_1}$. Construct

$$\hat{\mathbf{V}} = \begin{bmatrix}\mathbf{v}_1, \mathbf{V}_1\end{bmatrix} \quad \text{and} \quad \hat{\mathbf{U}} = \begin{bmatrix}\mathbf{u}_1, \mathbf{U}_1\end{bmatrix},$$

where $\mathbf{U}_1\mathbf{U}_1^T = \mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}_m$ and $\mathbf{V}_1\mathbf{V}_1^T = \mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}_n$. Then we have

$$\hat{\mathbf{U}}^T\mathbf{A}\hat{\mathbf{V}} \quad = \quad \begin{bmatrix}\mathbf{u}_1^T\mathbf{A}\mathbf{v}_1 & \mathbf{u}_1^T\mathbf{A}\mathbf{V}_1 \\ \mathbf{U}_1^T\mathbf{A}\mathbf{v}_1 & \mathbf{U}_1^T\mathbf{A}\mathbf{V}_1\end{bmatrix}.$$

It is easy to note that $\mathbf{U}_1^T\mathbf{A}\mathbf{v}_1 = \sigma_1\mathbf{U}_1^T\mathbf{u}_1 = 0$. Let $\mathbf{a} = \mathbf{u}_1^T\mathbf{A}\mathbf{V}_1$, $\mathbf{z} = (\sigma_1 \quad \mathbf{a}^T)$, $\mathbf{Y} = \hat{\mathbf{U}}^T\mathbf{A}\hat{\mathbf{V}}$. Then we have

$$\sigma_1\sqrt{\sigma_1^2 + \mathbf{a}^T\mathbf{a}} \geq ||\mathbf{Y}||_2||\mathbf{z}||_2 \geq ||\mathbf{Yz}||_2 = ||(\sigma_1^2 + ||\mathbf{a}_1||_2^2 \quad \mathbf{a}^T\mathbf{V}_1^T\mathbf{A}^T\mathbf{U}_1^T)^T||_2 \geq \sigma_1^2 + \mathbf{a}^T\mathbf{a}$$

Therefore, $\sigma_1\sqrt{\sigma_1^2 + \mathbf{a}^T\mathbf{a}} \geq \sigma_1^2 + \mathbf{a}^T\mathbf{a}$, we can get $\mathbf{a}^T\mathbf{a} = 0$, i.e. $\mathbf{a} = 0$. So

$$\hat{\mathbf{U}}^T\mathbf{A}\hat{\mathbf{V}} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \mathbf{U}_1^T\mathbf{A}\mathbf{V}_1 \end{bmatrix}$$

. Since $\mathbf{U}_1^T\mathbf{A}\mathbf{V}_1 \in \mathbb{R}^{(m-1)\times(n-1)}$, there exist orthogonal $\hat{\mathbf{U}}_1$, $\hat{\mathbf{V}}_1$ and diagonal $\hat{\Sigma}_1$, such that $\mathbf{U}_1^T\mathbf{A}\mathbf{V}_1 = \hat{\mathbf{U}}_1\hat{\Sigma}_1\hat{\mathbf{V}}_1$. Then one has

$$\mathbf{A} = \hat{\mathbf{U}} \begin{bmatrix} 1 & 0 \\ 0 & \hat{\mathbf{U}}_1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{\mathbf{\Sigma}}_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \hat{\mathbf{V}}_1 \end{bmatrix} \hat{\mathbf{V}}^T.$$

Both $\hat{\mathbf{U}} \begin{bmatrix} 1 & 0 \\ 0 & \hat{\mathbf{U}}_1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & \hat{\mathbf{V}}_1 \end{bmatrix} \hat{\mathbf{V}}^T$ are orthogonal matrices, so the theorem is correct.

□

Proposition 2.1 (Unique problem)**.** *If* $\sigma_1 > \sigma_2 > ... > \sigma_p$*, then there exists an unique SVD for* $\mathbf{A} \in \mathbb{R}^{m\times n}$*. Otherwise, suppose* $\sigma_1 = \sigma_2 = ...\sigma_{i_1} > \sigma_{i_1+1}... > \sigma_p > 0$*, i.e. the number of* σ_{i_1} *is* i_1*, the number of* σ_{i_2} *is* i_2*, ..., the number of* σ_{i_k} *is* i_k*. Then we have*

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \begin{bmatrix} \sigma_{i_1}\mathbf{I}_{i_1} & & \\ & \ddots & \\ & & \sigma_{i_k}\mathbf{I}_{i_k} \end{bmatrix} \mathbf{V}^T \\ &= \mathbf{U} \begin{bmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_k \end{bmatrix} \begin{bmatrix} \sigma_{i_1}\mathbf{I}_{i_1} & & \\ & \ddots & \\ & & \sigma_{i_k}\mathbf{I}_{i_k} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T & & \\ & \ddots & \\ & & \mathbf{Q}_k^T \end{bmatrix} \mathbf{V}^T, \end{aligned}$$

where Q_j *is an orthogonal matrix in* $\mathbb{R}^{i_j\times i_j}$*.*

Since $\mathbf{U} \begin{bmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_k \end{bmatrix}$ *and* $\begin{bmatrix} \mathbf{Q}_1^T & & \\ & \ddots & \\ & & \mathbf{Q}_k^T \end{bmatrix} \mathbf{V}^T$ *are still orthogonal matrices, the SVD for* $\mathbf{A}$ *is not unique.*