EECS 16B Designing Information Devices and Systems II
Fall 2020 UC Berkeley

Note 18

1 Overview

In this note, we will be taking a look the **Singular Value Decomposition** or SVD. It is an extremely useful tool that is used in many fields such as Statistics, Image Processing, Machine Learning, and even in Control Systems and is hence referred to as the "Swiss Army Knife" of Linear Algebra.

The SVD lets us write out a matrix A as a weighted sum of rank 1 matrices sometimes referred to as **features.** The weight on each feature signifies its importance and when writing out the SVD, we order them from largest to smallest. Intuitively features with larger weights are more important and the SVD can be used to approximate a matrix A by truncating "less important" features.

In order to build the SVD, we will use the results from the Spectral Theorem and use the eigenvectors of the symmetric matrix A^TA. In addition, the basis vectors that come out of the SVD will be orthonormal by construction. From a geometric perspective, these orthonormal transformations can be thought of as rotations since an orthonormal matrix U does not change the norm of a vector $\vec{x}$.

2 Singular Value Decomposition

There are multiple ways to define the Singular Value Decomposition of a matrix. We can write out the SVD in its **compact** form as a sum of rank-1 matrices or we can write out the **full SVD** as a series of matrix multiplications.

The **compact SVD** of an $m \times n$ matrix A of rank k is

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_k \vec{u}_k \vec{v}_k^T = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T \tag{1}$$

The vectors $\vec{u}_i$ are orthonormal and are called the **left singular vectors**. The vectors $\vec{v}_i$ are also orthonormal and are the **right singular vectors**. The scalars σ_i are the **singular values** of A. For values of $i > k$, $\sigma_i = 0$.

Alternatively, the **full SVD** of an $m \times n$ matrix A with rank k is

$$A = U\Sigma V^T \tag{2}$$

$$U = \begin{bmatrix} | & & | \\ \vec{u}_1 & \ldots & \vec{u}_m \\ | & & | \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \\ & 0 & \end{bmatrix} \qquad V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \ldots & \vec{v}_n \\ | & & | \end{bmatrix} \tag{3}$$

Note that Σ is an $m \times n$ matrix whose shape will change based on the shape of A.

A is a tall matrix $m > n$

$$\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \\ & 0 & \end{bmatrix}$$

A is a wide matrix $n > m$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{bmatrix}$$

2.1 Understanding the SVD

The matrix A is a linear transformation that sends vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$. Therefore, the right singular vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ form a basis for $\mathbb{R}^n$ while the left singular vectors $\{\vec{u}_1, \ldots, \vec{u}_m\}$ form a basis for $\mathbb{R}^m$.

In addition to this, the choices of $\vec{v}_i$ and $\vec{u}_i$ are special in that

$$A\vec{v}_i = \sigma_i \vec{u}_i \tag{4}$$

We will now prove the existence of the SVD and its connection to the eigenvectors of the matrix $A^T A$.

2.1.1 Positive Definiteness

The Spectral Theorem tells us that the symmetric matrix $A^T A$ has a set of orthonormal eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$. We can show that the eigenvalues of $A^T A$ are all greater than or equal to 0.

$$\|A\vec{v}\|^2 = \langle A\vec{v}, A\vec{v}\rangle = \vec{v}^T A^T A\vec{v} = \lambda \vec{v}^T \vec{v} = \lambda \|\vec{v}\|^2 \implies \lambda = \frac{\|A\vec{v}\|^2}{\|\vec{v}\|^2} \geq 0 \tag{5}$$

The last inequality follows from the positive-definiteness of inner products and norms.

2.1.2 Eigenspaces

Now we show the relation between the eigenvectors of $A^T A$ and AA^T. If $\vec{v}$ is an eigenvector of $A^T A$ with nonzero eigenvalue, then the vector $\vec{w} = A\vec{v}$ must be an eigenvector of AA^T.

$$A^T A\vec{v} = \lambda \vec{v} \implies A(A^T A\vec{v}) = A(\lambda \vec{v}) \implies AA^T(A\vec{v}) = \lambda(A\vec{v}) \tag{6}$$

If $\vec{v}$ were an eigenvector of zero eigenvalue, then $\vec{v}$ is in the null-space of A since $\text{Nul}(A) = \text{Nul}(A^T A)$.

Now recalling our results from the last section, let's compute the norm of the vector $\vec{w} = A\vec{v}$

$$\|\vec{w}\| = \|A\vec{v}\| = \sqrt{\lambda}\|\vec{v}\| \tag{7}$$

This means if we were to normalize $\vec{w}$ as a unit vector $\vec{u}$, it would follow that

$$\vec{u} = \frac{\vec{w}}{\|\vec{w}\|} = \frac{A\vec{v}}{\sqrt{\lambda}} \implies A\vec{v} = \sqrt{\lambda}\vec{u} \tag{8}$$

As a result, we define $\sigma = \sqrt{\lambda}$ and call a singular value of A.

2.1.3 Summary

If A is an $m \times n$ matrix of rank k, then $A^T A$ will have k orthonormal eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_k\}$. For each of these k eigenvectors, there exist orthonormal eiegenvectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of AA^T.

As a result, for $i = 1, \ldots, k$ we can say that

$$A\vec{v}_i = \sigma_i \vec{u}_i \tag{9}$$

the remaining vectors $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ form the $\text{Nul}(A)$ and correspond to singular values $\sigma_i = 0$. The vectors $\{\vec{u}_{k+1}, \ldots, \vec{u}_m\}$ form the $\text{Nul}(A^T)$ can also be picked through Gram-Schmidt process.

2.2 SVD From the Other Side

While we have defined the left-singular vectors $\vec{u}$ from the right-singular vectors $\vec{v}$, we could have done the entire process from the other side. This perspective will be very useful when computing the SVD.

Starting with an eigenvector $\vec{u}$ of AA^T, we can show that $\vec{w} = A^T \vec{u}$ is an eigenvector of $A^T A$.

$$AA^T \vec{u} = \lambda \vec{u} \implies A^T(AA^T \vec{u}) = A^T(\lambda \vec{u}) \implies A^T A(A^T \vec{u}) = \lambda (A^T \vec{u}) \tag{10}$$

The norm of $\|\vec{w}\| = \sqrt{\lambda} = \sigma$ and we can again define a relation between $\vec{u}$ and $\vec{v}$.

$$\vec{v} = \frac{\vec{w}}{\|\vec{w}\|} = \frac{A^T \vec{u}}{\sigma} \implies A^T \vec{u} = \sigma \vec{v} \tag{11}$$

3 Computing the SVD

We will show two examples of computing the SVD for tall and wide matrices but note that in practice, we will always compute the SVD of large matrices using numerical tools.

SVD of a Tall Matrix

Let's first look at a 3×2 matrix A and compute its SVD

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

Step 1: Compute the symmetric matrix $A^T A$

$$A^T A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Step 2: Find orthonormal eigenpairs $(\lambda_i, \vec{v}_i)$ of $A^T A$ for $i = 1, \ldots, k$ and order them from largest to smallest

$$\lambda^2 - 4\lambda + 3 = 0 \implies \lambda_1 = 3, \lambda_2 = 1$$

$$\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 3: Compute the singular values $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{3} \qquad \sigma_2 = \sqrt{1}$$

Step 4: Compute the right singular vectors $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$ for $i = 1, \ldots, k$.

$$\vec{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \sqrt{6}/3 \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \qquad \vec{u}_2 = \frac{1}{1} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 5: Complete the bases U and V through Gram-Schmidt or by computing the appropriate null spaces.

Since $\text{Rank}(A) = 2$ we don't need to add any move vectors to V. However, we will need to add a third vector to U. We can find $\vec{u}_3$ by finding a basis for $\text{Nul}(A^T)$.

$$\vec{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

To summarize, the SVD of the matrix A can be written as

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

SVD of a Wide Matrix

Now let us look at a 2×3 matrix A and compute its SVD

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

Step 1: Compute the symmetric matrix AA^T [1]

$$AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Step 2: Find orthonormal eigenpairs $(\lambda_i, \vec{u}_i)$ of AA^T for $i = 1, \ldots, k$ and order them from largest to smallest

$$\lambda^2 - 4\lambda + 3 = 0 \implies \lambda_1 = 3, \lambda_2 = 1$$

$$\vec{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \vec{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

[1] We compute AA^T instead of A^TA since it is a smaller, 2×2 matrix. In general, it will be easier to diagonalize a smaller matrix so we pick A^TA for tall matrices and AA^T for wide matrices

Step 3: Compute the singular values $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{3} \qquad \sigma_2 = \sqrt{1}$$

Step 4: Compute the left singular vectors $\vec{v}_i = \frac{A^T\vec{u}_i}{\sigma_i}$ for $i = 1, \dots, k$.

$$\vec{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}1 & 0\\1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{6}}\\\frac{\sqrt{6}}{3}\\\frac{1}{\sqrt{6}}\end{bmatrix} \qquad \vec{v}_2 = \frac{1}{1}\begin{bmatrix}1 & 0\\1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}-\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix}$$

Step 5: Complete the bases U and V through Gram-Schmidt or by computing the appropriate null spaces. Since $\text{Rank}(A) = 2$ we don't need to add any move vectors to U. However, we will need to add a third vector to V. We can find $\vec{v}_3$ by finding a basis for $\text{Nul}(A)$.

$$\vec{v}_3 = \begin{bmatrix}\frac{1}{\sqrt{3}}\\-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}$$

To summarize, the SVD of the matrix A can be written as

$$A = \begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{3} & 0 & 0\\0 & 1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}}\\-\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{bmatrix}$$

4 Fundamental Theorem of Linear Algebra

The results from the SVD can be summarized by the **Fundamental Theorem of Linear Algebra** which states that for an $m \times n$ matrix A

$$\text{Col}(A) \perp \text{Nul}(A^T) \tag{12}$$

$$\text{Nul}(A) \perp \text{Col}(A^T) \tag{13}$$

In otherwords, the $\text{Col}(A)$ is orthogonal to the $\text{Nul}(A^T)$ and the $\text{Nul}(A)$ is orthogonal to the $\text{Col}(A^T)$.

4.1 Proof

4.1.1 Basis for Col(A^T)

If A is of rank k, then the first k right-singular vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ form a basis for the $\text{Col}(A^T)$. To see this, recall that the right-singular vectors are eigenvectors of A^TA.

$$A^T(A\vec{v}_i) = \lambda\vec{v}_i \implies \vec{v}_i \in \text{Col}(A^T) \tag{14}$$

Since $\text{Rank}(A) = \text{Rank}(A^T) = k$ and $\{\vec{v}_1, \dots, \vec{v}_k\}$ are all in $\text{Col}(A^T)$, they must form a basis.

4.1.2 Dimension of Nul(A)

By the Rank-Nullity Theorem,

$$\text{Rank}(A) + \dim \text{Nul}(A) = n \tag{15}$$

Since Rank$(A) = k$, it follows that dim Nul$(A) = n - k$.

4.1.3 Basis for Nul(A)

The last $n-k$ right-singular vectors $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ are all eigenvectors of eigenvalue 0. Hence, they form a basis for Nul(A^TA). Since, Nul$(A^TA) =$ Nul(A), we can say that $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ forms a basis for Nul(A).

4.1.4 Orthogonality

From the Spectral Theorem, we can pick an orthonormal set of eigenvectors for the matrix A^TA. Therefore, since all of the vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are orthonormal, the individual bases for Col(A^T) and Nul(A) must also be orthogonal. Since the bases for two vector spaces are orthogonal, we conclude by saying every vector in Col(A^T) must be orthogonal to Nul(A).

We can use a similar argument using the eigenvectors of AA^T to show that Col$(A) \perp$ Nul(A^T).

5 Conclusion

In this note, we developed the Singular Value Decomposition of a matrix A through the orthonormal eigenspaces of the matrices A^TA and AA^T. The two different forms we define for the SVD were:

- **Compact SVD:**

$$A = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_k \vec{u}_k \vec{v}_k^T$$

- **Full SVD:**

$$A = \sum_{i=1}^{n} \sigma_i \vec{u}_i \vec{v}_i^T = \begin{bmatrix} | & & | \\ \vec{u}_1 & \ldots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \\ & 0 & \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v}_1 & \ldots & \vec{v}_n \\ | & & | \end{bmatrix}^T$$

The left and right singular vectors of the U and V matrices are orthonormal and the singular values σ_i are ordered from largest to smallest. The first k left-singular vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ span the Col(A) while the last $n-k$ right-singular vectors $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ span the Nul(A).

In the next note, we will focus on applications of the Singular Value Decomposition and take a look at the geometric interpretation by viewing the U and V matrices as orthonormal rotations.

Contributors:

- Taejin Hwang.

