Lecture 5: SVD

Michael S. Floater

September 19, 2018

These notes are based on Section 2.10 of the book.

1 SVD for a square, non-singular matrix

To understand the main ideas of the SVD factorization we start first with a non-singular $n \times n$ matrix A.

The SVD (singular value decomposition) is based on the eigenvalues and eigenvectors of the two matrices $A^T A$ and AA^T, both of which are symmetric and positive definite. They have the same eigenvalues but different eigenvectors. These common eigenvalues are positive and we can take their square roots, and these are called the singular values of A.

We consider first $A^T A$. It has real, positive eigenvalues and a set of orthonormal eigenvectors,

$$A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j, \quad j = 1, \ldots, n.$$

We order the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$, and let D be the diagonal matrix with entries $D_{i,i} = \lambda_i$, $i = 1, \ldots, n$.

We let V be the orthogonal matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Since

$$\mathbf{v}_i^T A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = \lambda_j \delta_{ij}$$

for all $i, j = 1, \ldots, n$, we find

$$V^T A^T A V = D.$$

Since D is diagonal and has positive elements we can form $D^{1/2}$ and its inverse $D^{-1/2}$.

We now define the matrix

$$U = AVD^{-1/2}.$$

It follows that

$$UD^{1/2}V^T = AVD^{-1/2}D^{1/2}V^T = AVV^T = A,$$

i.e., we have constructed the factorization of A,

$$A = UD^{1/2}V^T. \tag{1}$$

Moreover, we find that the columns of U are orthonormal, since

$$U^TU = (AVD^{-1/2})^T AVD^{-1/2} = D^{-1/2}V^TA^TAVD^{-1/2} = D^{-1/2}DD^{-1/2} = I.$$

We also find that

$$\begin{aligned} U^TAA^TU &= (AVD^{-1/2})^T AA^T(AVD^{-1/2}) = D^{-1/2}(V^TA^TA)(A^TAV)D^{-1/2} \\ &= D^{-1/2}(DV^T)(VD)D^{-1/2} = D^{-1/2}DDD^{-1/2} = D. \end{aligned}$$

So the columns of U are eigenvectors of AA^T:

$$AA^T\mathbf{u}_j = \lambda_j\mathbf{u}_j, \quad j = 1, \ldots, n.$$

We further let $\sigma = \sqrt{\lambda_i}$, $i = 1, \ldots, n$, and we let S be the diagonal matrix with diagonal $S_{i,i} = \sigma_i$, $i = 1, \ldots, n$, so $S = D^{1/2}$. The factorization

$$A = USV^T \tag{2}$$

is called the *singular value decomposition* of A. The numbers σ_i, $i = 1, \ldots, n$, are the *singular values* of A.

Notice that we can alternatively write the decomposition as

$$A = \sum_{j=1}^{n} \sigma_j \mathbf{u}_j \mathbf{v}_j^T,$$

and each matrix $\mathbf{u}_j\mathbf{v}_j^T$ has rank 1. This leads to the idea of 'low rank approximation': we could approximate A by removing the terms in the sum with small singular values.

Here is an example in which $n = 3$:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & -2 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Its SVD is $A = USV^T$, where

$$U = \begin{bmatrix} -0.1200 & -0.8097 & 0.5744 \\ 0.9018 & 0.1531 & 0.4042 \\ -0.4153 & 0.5665 & 0.7118 \end{bmatrix},$$

$$S = \begin{bmatrix} 2.4605 & 0 & 0 \\ 0 & 1.6996 & 0 \\ 0 & 0 & 0.2391 \end{bmatrix},$$

$$V = \begin{bmatrix} -0.4153 & -0.5665 & 0.7118 \\ -0.9018 & 0.1531 & -0.4042 \\ 0.1200 & -0.8097 & -0.5744 \end{bmatrix}.$$

2 Full SVD

There are two generalizations of the above factorization: (1) there could be some zero eigenvalues, and (2) we can allow a rectangular matrix A.

We now allow A to be an arbitrary $n \times m$ (real) matrix. We allow $n > m$ and $n = m$ and $n < m$.

As before, both $A^T A$ and AA^T are square, positive semi-definite matrices, but their dimensions are now $m \times m$ and $n \times n$ respectively.

$A^T A$ has dimension $m \times m$ and has real, non-negative eigenvalues and orthonormal eigenvectors

$$A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j, \quad j = 1, \ldots, m,$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ and $\lambda_{k+1} = \cdots = \lambda_m = 0$ for some k, where k is the rank of A, $k \leq \min\{n, m\}$.

Let D be the diagonal matrix with entries $D_{i,i} = \lambda_i$, $i = 1, \ldots, m$ and let V be the orthogonal matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_m$. Then

$$V^T A^T A V = D = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix},$$

where D_1 is a $k \times k$ matrix containing the non-zero portion of D. Since D_1 is diagonal and has positive elements we can form $D_1^{1/2}$ and its inverse $D_1^{-1/2}$.

Next we partition V into blocks $V = [V_1 \;\; V_2]$, where V_1 consists of the first k columns of V, and V_2 the remaining $m - k$. Then

$$\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} A^T A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} V_1^T A^T A V_1 & V_1^T A^T A V_2 \\ V_2^T A^T A V_1 & V_2^T A^T A V_2 \end{bmatrix},$$

and so

$$V_1^T A^T A V_1 = D_1, \qquad V_2^T A^T A V_2 = 0.$$

We now define the $n \times k$ matrix

$$U_1 = A V_1 D_1^{-1/2}.$$

It follows that

$$A = U_1 S_1 V_1^T. \tag{3}$$

where $S_1 = D_1^{1/2}$. Moreover, we find that the columns of U_1 are orthonormal, i.e.,

$$U_1^T U_1 = I,$$

and also that

$$U_1^T A A^T U_1 = D_1,$$

and so the columns of U_1 are eigenvectors of AA^T:

$$AA^T \mathbf{u}_j = \lambda_j \mathbf{u}_j, \quad j = 1, \ldots, k.$$

In this factorization U_1 is $n \times k$, S_1 is $k \times k$, and V_1 is $k \times m$.

To make a factorization of A with square matrices U and V, we choose any $n \times (n - k)$ matrix U_2 such that

$$U = [U_1 \quad U_2]$$

is orthogonal. We further let $\sigma = \sqrt{\lambda_i}$, $i = 1, \ldots, k$, and we let S be the diagonal $n \times m$ matrix with diagonal $S_{i,i} = \sigma_i$, $i = 1, \ldots, k$. We now obtain the factorization

$$A = USV^T, \tag{4}$$

which is called the *singular value decomposition* of A. The numbers σ_i, $i = 1, \ldots, k$, are the *singular values* of A. The last diagonal entries of S may be zero. The dimensions of U, S, V are $n \times n$, $n \times m$ and $m \times m$.

Here is an example in which $n = 4$ and $m = 2$, and A has (full) rank $k = 2$:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}.$$

Its SVD is $A = USV^T$, where

$$U = \begin{bmatrix} -0.1525 & -0.8226 & -0.3945 & -0.3800 \\ -0.3499 & -0.4214 & 0.2428 & 0.8007 \\ -0.5474 & -0.0201 & 0.6979 & -0.4614 \\ -0.7448 & 0.3812 & -0.5462 & 0.0407 \end{bmatrix},$$

$$S = \begin{bmatrix} 14.2691 & 0 \\ 0 & 0.6268 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} -0.6414 & 0.7672 \\ -0.7672 & -0.6414 \end{bmatrix}.$$

The reduced form is $A = U_1 S_1 V_1^T$, where

$$U_1 = \begin{bmatrix} -0.1525 & -0.8226 \\ -0.3499 & -0.4214 \\ -0.5474 & -0.0201 \\ -0.7448 & 0.3812 \end{bmatrix},$$

$$S_1 = \begin{bmatrix} 14.2691 & 0 \\ 0 & 0.6268 \end{bmatrix},$$

$$V_1 = \begin{bmatrix} -0.6414 & 0.7672 \\ -0.7672 & -0.6414 \end{bmatrix}.$$

3 Least squares again

Let us now return to the overdetermined system of equations

$$A\mathbf{x} - \mathbf{b} = 0,$$

where $n > m$, and look again at the least squares problem of minimizing $\|A\mathbf{x} - \mathbf{b}\|$. By the orthogonality of U and V, we find that

$$\begin{aligned}\|A\mathbf{x} - \mathbf{b}\|^2 &= \|U^T(AVV^T\mathbf{x} - \mathbf{b})\|^2 = \|SV^T\mathbf{x} - U^T\mathbf{b}\|^2 \\ &= \sum_{i=1}^{k}(\sigma_i(V^T\mathbf{x})_i - \mathbf{u}_i^T\mathbf{b})^2 + \sum_{i=k+1}^{n}(\mathbf{u}_i^T\mathbf{b})^2,\end{aligned}$$

where $\mathbf{u}_i$ denotes the i-th column of U. Thus we achieve the minimum by forcing the first sum to be zero:

$$(V^T\mathbf{x})_i = \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}, \quad i = 1, \ldots, k,$$

Multiplying both sides by $\mathbf{v}_j$, and since

$$\mathbf{v}_j(V^T\mathbf{x})_i = (\mathbf{v}_j\mathbf{v}_i^T)\mathbf{x} = \begin{cases} 0 & j \neq i, \\ x_i & j = i, \end{cases},$$

we deduce that the solution is

$$\mathbf{x} = \sum_{i=1}^{k}\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i.$$

This is an explicit formula for $\mathbf{x}$ but it is not suitable in general since it requires finding the eigenvalues and eigenvectors of A^TA.

Yet another approach to solving the least squares problem is to solve the so-called normal equations.

Theorem 1 $\mathbf{x} \in \mathbb{R}^m$ *is a solution to the least squares problem if and only if*

$$A^T(A\mathbf{x} - \mathbf{b}) = 0.$$

Proof. If $\mathbf{x}$ is a solution then it minimizes the function

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2 = \mathbf{x}^TA^TA\mathbf{x} - 2\mathbf{x}^TA^T\mathbf{b} + \mathbf{b}^T\mathbf{b}.$$

Hence the gradient $\nabla f(\mathbf{x}) = 2A^TA\mathbf{x} - 2A^T\mathbf{b}$ is zero, and so $A^T(A\mathbf{x} - \mathbf{b}) = 0$.

Conversely, suppose that $A^T(A\mathbf{x} - \mathbf{b}) = 0$. Let $\mathbf{z}$ be any vector in $\mathbb{R}^m$ and let $\mathbf{y} = \mathbf{z} - \mathbf{x}$. Then

$$\begin{aligned}\|A\mathbf{z} - \mathbf{b}\|^2 &= \|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|^2 \\ &= \|A\mathbf{x} - \mathbf{b}\|^2 + 2\mathbf{y}^TA^T(A\mathbf{x} - \mathbf{b}) + \|A\mathbf{y}\|^2 \\ &= \|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 \geq \|A\mathbf{x} - \mathbf{b}\|^2.\end{aligned}$$

$\square$

If $A^T A$ is non-singular this offers a simple way of solving the least squares problem. However, it should be used with caution since building the matrix $A^T A$ may lead to loss of accuracy. The QR method is in general more stable.

4 Exercises

Exercise 1 *Compute the SVD of*

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

Exercise 2 *Compute the SVD of*

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