SINGULAR VALUE DECOMPOSITION

In these notes we describe the singular value decomposition.

1. Reduced Singular Decomposition

Let $A \in \mathbb{R}^{n\times m}$ be a matrix. Consider the symmetric matrix $M = A^T A \in \mathbb{R}^{m\times m}$. According to the spectral decomposition theorem, that is, the Schur factorization specialized to symmetric real matrices, there exist a diagonal matrix $D \in \mathbb{R}^{m\times m}$ and an orthogonal matrix $V \in \mathbb{R}^{m\times m}$ such that

$$M = A^T A = VDV^T.$$

Let us denote the columns of V by $v^1, \dots, v^m \in \mathbb{R}^m$. Furthermore, let us define vectors

$$u^1 = Av^1, \quad u^2 = Av^2, \quad \dots, \quad u^n = Av^n.$$

We observe that the vectors $u^1, u^2, \dots, u^m$ are pairwise orthogonal and that are their lengths are the square roots of the corresponding entries in the diagonal matrix D. Indeed, for $1 \leq i, j \leq m$, we find that

$$\begin{aligned} u^i \cdot u^j = Av^i \cdot Av^j &= v^i \cdot A^T A v^j \\ &= v^i \cdot VDV^T v^j = V^T v^i \cdot DV^T v^j = e^i \cdot De^j = D_{ii}\delta^{ij}. \end{aligned}$$

Here, δ^{ij} denotes the Kronecker delta. We conclude that

$$\begin{aligned} u^i \cdot u^j &= 0, \quad i \neq j, \\ \|u^i\|^2 &= u^i \cdot u^i = D_{ii}. \end{aligned}$$

In particular, $\|u^i\| = \sqrt{D_{ii}}$.

In order to finish the construction of the (reduced) singular value decomposition, we define vectors $w^1, \dots, w^m \in \mathbb{R}^n$ and scalars $\sigma_1, \dots, \sigma_m \in \mathbb{R}$ as follows. For every $1 \leq i \leq m$ such that $u^i \neq 0$ is not a zero vector, we define

$$w^i = \frac{u^i}{\|u^i\|}, \quad \sigma_i = \|u^i\|.$$

The vectors w^i defined in that manner constitute a system of pairwise orthogonal unit vectors. Finally, for every $1 \leq i \leq m$ such that $u^i = 0$, we let $w^i = 0$ be the zero vector.

We then define $W \in \mathbb{R}^{n\times m}$ as the matrix whose columns are the vectors $w^1, \dots, w^m$. In particular, the non-zero columns of W are pairwise orthogonal unit vectors. Moreover, we define the matrix $\Sigma \in \mathbb{R}^{m\times m}$ as the diagonal matrix whose entries are the values $\sigma_1, \dots, \sigma_m$,

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{pmatrix}.$$

By construction, we have the matrix decomposition

$$A = W \Sigma V^T.$$

We call this the reduced singular value decomposition.

The reduced singular value decomposition already contains all the essential mathematical features of the full singular value decomposition, which we consider below among other technical issues.

Lemma 1
The number of non-zero singular values equals the rank of the matrix A.

Proof. Let $\sigma_1, \dots, \sigma_r$ be the non-zero diagonal entries of A. Those are precisely the non-zero eigenvalues of $A^T A$ with multiplicity. Their number equals the rank of $A^T A$, which equals the rank of A. □

Remark 1
Depending on the author, the term *singular value* is used with slight technical differences. Some reserve the term singular value only for the non-zero entries $\sigma_1, \dots, \sigma_r$ of Σ, or others use it only for the first entries $\sigma_1, \dots, \sigma_k$ of Σ with $k := \min(m, n)$.

2. Ordering of the Singular Values

It is customary to have the singular values numbered in descending order, though this may depend on the author. Indeed, given the diagonal matrix $\Sigma \in \mathbb{R}^{m \times m}$, there exists a permutation matrix $P \in \mathbb{R}^{m \times m}$ such that $P \Sigma P^T$ has the same set of diagonal entries as Σ with the same multiplicites but in descending order. We can then write

$$A = W \Sigma V = W P^T P \Sigma P^T P V^T$$

Obviously, the decomposition

$$A = (W P^T)(P \Sigma P^T)(P V^T)$$

takes on the same form as the original (reduced) singular value decomposition but now with a diagonal matrix whose singular values are ordered in descending manner.

By convention, we henceforth assume that the singular values are ordered in descending manner:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r.$$

This will simplify the exposition in the sequel.

3. Full Singular Value Decomposition

An aesthetical issue is that the matrix $W \in \mathbb{R}^{n\times m}$ has its non-zero columns constituting an orthonormal system, but in the form above the matrix is not an orthogonal matrix. In fact, it may not even be a square matrix.

We first consider the case $n \geq m$, that is, the matrix A has at least as many rows as columns. Since we assume that the singular values $\sigma_1, \dots, \sigma_m$ are ordered in descending manner, there exists a minimal $1 \leq r \leq m$ such that $\sigma_i = 0$ for all $r < i \leq m$.

We now define unit vectors $\hat{w}^1, \hat{w}^2, \dots, \hat{w}^n \in \mathbb{R}^n$ as follows. The vectors

$$\hat{w}^1 := w^1, \dots, \hat{w}^r := w^r.$$

are just the non-zero columns of W. We let the vectors $\hat{w}^{r+1}, \dots, \hat{w}^n$ be any choice of unit vectors such that $\hat{w}^1, \dots, \hat{w}^n$ form an orthonormal basis of $\mathbb{R}^n$.

We let $\hat{W} \in \mathbb{R}^{n\times n}$ be the matrix with columns $\hat{w}^1, \dots, \hat{w}^n$. Furthermore, we obtain the matrix $\hat{\Sigma} \in \mathbb{R}^{n\times m}$ by appending $n - m$ zero rows to the diagonal matrix Σ. It is evident from the construction that

$$\hat{W} \cdot \hat{\Sigma} = W \cdot \Sigma$$

give the same $n \times m$ matrix. The identity

$$A = \hat{W} \cdot \hat{\Sigma} \cdot V^T$$

is an immediate consequence.

We now consider the case $n \leq m$, that is, the matrix A has at least as many columns as rows. Since we assume that the singular values $\sigma_1, \dots, \sigma_m$ are ordered in descending manner, there exists again a minimal $1 \leq r \leq m$ such that $\sigma_i = 0$ for all $r < i \leq m$. By construction, we have $r \leq n$.

Whereas above we have appended columns to W and appended rows to Σ, we know delete columns of W and delete rows of Σ.

We let now define unit vectors $\hat{w}^1, \hat{w}^2, \dots, \hat{w}^n \in \mathbb{R}^n$ as follows. The vectors

$$\hat{w}^1 := w^1, \dots, \hat{w}^r := w^r.$$

are just the non-zero columns of W. We let the vectors $\hat{w}^{r+1}, \dots, \hat{w}^n$ be any choice of unit vectors such that $\hat{w}^1, \dots, \hat{w}^n$ form an orthonormal basis of $\mathbb{R}^n$.

We let $\hat{W} \in \mathbb{R}^{n\times n}$ be the matrix with columns $\hat{w}^1, \dots, \hat{w}^n$. Furthermore, we obtain the matrix $\hat{\Sigma} \in \mathbb{R}^{n\times m}$ by deleting $n - m$ zero columns of the diagonal matrix Σ. It is evident from the construction that

$$\hat{W} \cdot \hat{\Sigma} = W \cdot \Sigma$$

give the same $n \times m$ matrix. The identity

$$A = \hat{W} \cdot \hat{\Sigma} \cdot V^T$$

is an immediate consequence.

Renaming variables, we have the following result, which gives what is known as the (full) singular value decomposition.

Theorem 1
Let $A \in \mathbb{R}^{n\times m}$. Then there exist orthogonal matrices $W \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ and a diagonal rectangular matrix $\Sigma \in \mathbb{R}^{n\times m}$ such that

$$A = W \cdot \Sigma \cdot V^T.$$

We can assume that the diagonal entries of Σ are ordered in an descending manner.

Lemma 2
Let $A \in \mathbb{R}^{n\times m}$. Then the non-zero singular values of A are unique.

Proof. The non-zero singular values are precisely the non-negative square roots of the square matrix $A^TA \in \mathbb{R}^{m\times m}$. Since the eigenvalues of any matrix and their multiplicites are unique, the claim follows. □

4. Pseudoinverse via Singular Value Decomposition

The singular value decomposition can be used in the computation of the pseudoinverse. Let $A \in \mathbb{R}^{n\times m}$ be a matrix with singular value decomposition

$$A = W \cdot \Sigma \cdot V^T,$$

where $W \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with descending diagonal entries.

The pseudoinverse of the matrix Σ is easily found. The first r diagonal entries of Σ are its non-zero entries $\sigma_1, \dots, \sigma_r$. We let $\Sigma^\dagger \in \mathbb{R}^{m\times n}$ be the $m \times n$ matrix whose non-zero entries are precisely the non-zero entries $\sigma_1^{-1}, \dots, \sigma_r^{-1}$.

Example 1

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^\dagger = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}^\dagger = \begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} \pi & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}^\dagger = \begin{pmatrix} \frac{1}{\pi} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \end{pmatrix},$$

Theorem 2
The pseudoinverse of A is the matrix

$$A^\dagger = V \cdot \Sigma^\dagger \cdot W^T.$$

Proof. We show that for any $b \in \mathbb{R}^n$ the solution of the least-squares problem $Ax = b$ is given by $x = V\Sigma^\dagger W^T b$. It then follows that $V\Sigma^\dagger W^T = A^\dagger$ is the pseudoinverse of A.

We first determine the kernel and range of A in terms of the kernel and range of Σ. Since W and V are invertible, we have

$$\operatorname{ran}(A) = \operatorname{ran}(W\Sigma V^T) = W \cdot \operatorname{ran}(\Sigma V^T) = W \cdot \operatorname{ran}(\Sigma),$$
$$\ker(A) = \ker(W\Sigma V^T) = V \cdot \ker(W\Sigma) = V \cdot \ker(\Sigma).$$

Now let $x \in \mathbb{R}^m$. We see

$$x \perp \ker(A) \iff x \perp \ker(W\Sigma V^T) \iff x \perp \ker(\Sigma V^T) \iff Vx \perp \ker(\Sigma)$$

and

$$\begin{aligned} Ax - b \perp \operatorname{ran}(A) &\iff W\Sigma V^T x - b \perp \operatorname{ran}(W\Sigma V^T) \\ &\iff \Sigma V^T x - W^T b \perp \operatorname{ran}(\Sigma V^T) \iff \Sigma V^T x - W^T b \perp \operatorname{ran}(\Sigma). \end{aligned}$$

We conclude that x is the minimum norm solution of the generalized least-squares problem $Ax = b$ if and only if $z = V^T x$ is the minimum norm solution of the generalized least-squares problem $\Sigma z = W^T b$. The latter is the case if and only if $z = V^T x$ and $z = \Sigma^\dagger W^T b$. In conclusion, $x = V\Sigma^\dagger W^T b$. □