Matrix Decompositions

April 17, 2023

1 Singular Value Decomposition

This method is often referred as the "fundamental theorem of linear algebra", because it applies to all matrices, not only square matrices.

Theorem - SVD Theorem
Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be a rectangular matrix of rank $r \in [0, min(m,n)]$. The SVD of $\boldsymbol{A}$ is a decomposition of the form

$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^T_{n\times n} \tag{1}$$

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times n}$ with column vectors $\boldsymbol{u}_i$, $i = 1,\ldots,m$, and an orthogonal matrix $\boldsymbol{V}^{n\times n}$ with column vectors $\boldsymbol{v}_j$, $j = 1,\ldots,n$. Moreover, $\boldsymbol{\Sigma}$ is an $m\times n$ matrix with $\Sigma_{ii} = \sigma_i \geqslant 0$ and $\Sigma_{ij} = 0, i \neq j$.

The diagonal elements $\sigma_i, i = 1,\ldots,r$, of $\boldsymbol{\Sigma}$ are called *singular values*, $\boldsymbol{u}_i$ are called the *left-singular vectors*, and $\boldsymbol{v}_j$ are called the *right-singular vectors*. By convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant r \geqslant 0$.

The *singular value matrix* $\boldsymbol{\Sigma}$ is unique, but needs some attention. It has the same dimension as $\boldsymbol{A}$, which means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding. Specifically, if $m > n$, then the matrix $\boldsymbol{\Sigma}$ has diagonal structure up to row n and then consists of $\boldsymbol{0}^T$ row vectors from $n+1$ to m so that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix} \tag{2}$$

If $m < n$, the matrix $\boldsymbol{\Sigma}$ has a diagonal structure up to column m and columns tha consist of $\boldsymbol{0}$ from $m+1$ to n:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \sigma_m & 0 & \ldots & 0 \end{bmatrix} \tag{3}$$

The SVD exists for every matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$.

2 Construction of the SVD

The SVD shares some similarities with the eigendecomposition of a square matrix.
Let's compare the eigendecomposition of an SPD matrix

$$\boldsymbol{S} = \boldsymbol{S}^T = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^T \tag{4}$$

with the corresponding SVD

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{5}$$

if we set

$$\boldsymbol{U} = \boldsymbol{P} = \boldsymbol{V}, \quad \boldsymbol{D} = \boldsymbol{\Sigma} \tag{6}$$

we see that the SVD of SPD matrices is their eigendecomposition.
Computing the SVD of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is equivalent to finding two sets of orthonormal bases $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases we will construct the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$.

We first start with the orthonormal set of right-singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. Then we continue with constructing the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Then we link those two sets and and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be singular values.

Let us start with the right-singular set of vectors. The spectral theorem tells us that the eigenvectors of a symmetric matrix form an ONB, which also means it can be diagonalized. Moreover, we can always construct a symmetric, positive semidefinite matrix $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n \times n}$ from a rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Thus, we can always diagonlize $\boldsymbol{A}^T\boldsymbol{A}$ and obtain

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^T = \boldsymbol{P} \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} \boldsymbol{P}^T \tag{7}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$. Let us assume the SVD of $\boldsymbol{A}$ exists and inject Eq. 1 and Eq. 7. This yieds

$$\boldsymbol{A}^T\boldsymbol{A} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T) = \boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{8}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{V} \begin{bmatrix} \sigma_i^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_n^2 \end{bmatrix} \boldsymbol{V}^T \tag{9}$$

Conclusively, we can compare

$$\boldsymbol{V}^T = \boldsymbol{P}^T \tag{10}$$

$$\sigma_i^2 = \lambda_i \tag{11}$$

Therefore, the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ that compose $\boldsymbol{P}$ are the right-singular vectors $\boldsymbol{V}$ of $\boldsymbol{A}$. The eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ are the squared singular values of $\boldsymbol{\Sigma}$.

Moving on with the left-singular vectors $\boldsymbol{U}$, we follow a similar procedure, which yields

$$\boldsymbol{A}\boldsymbol{A}^T = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T = \boldsymbol{U} \begin{bmatrix} \sigma_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_m^2 \end{bmatrix} \boldsymbol{U}^T \tag{12}$$

The spectral theorem tells us that $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^T$ can be diagonalized and we can find an ONB of eigenvectors $\boldsymbol{A}\boldsymbol{A}^T$, which are collected in $\boldsymbol{S}$. The orthonormal eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ are the left-singular vectors $\boldsymbol{U}$ and form an orthonormal basis in the codomain of the SVD.

The only remaining mystery is the structure of the matrix $\mathbf{\Sigma}$. Since $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{A}^T\boldsymbol{A}$ have the same non-zero eigenvalues, the non-zero entries of $\mathbf{\Sigma}$ in the SVD for both cases have to be the same.

Finally, we get that for any two orthogonal eigenvectors $\boldsymbol{v}_i, \boldsymbol{v}_j, i \neq j$, it holds that

$$(\boldsymbol{A}\boldsymbol{v}_i)^T(\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_i^T(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{v}_j = \boldsymbol{v}_i^T(\lambda_j\boldsymbol{v}_j) = \lambda_j(\boldsymbol{v}_i^T\boldsymbol{v}_j) = 0 \tag{13}$$

For the case that $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r-dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need the left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{14}$$

where the last equality shows us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$ are such that $\sigma_i^2 = \lambda_i$.
With a little bit rearranging we obtain the *singular value equation*

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i, \quad i = 1, \ldots, r \tag{15}$$

Concatenating the $\boldsymbol{v}_i$ as the columns of $\boldsymbol{V}$ and the $\boldsymbol{u}_i$ as the columns of $\boldsymbol{U}$ yields

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\mathbf{\Sigma} \tag{16}$$

where $\mathbf{\Sigma}$ has the same dimensions as $\boldsymbol{A}$ and the diagonal structure for rows $1, \ldots, r$. Hence, right-multiplying with $\boldsymbol{V}^T$ yields $\boldsymbol{A} = \boldsymbol{U}\mathbf{\Sigma}\boldsymbol{V}^T$, which is the SVD of $\boldsymbol{A}$.

The SVD is used in variety of ML applications like least-squares problems in curve fitting to solving systems of linear equations. These applications harness various important properties of the SVD, its relation to the rank of a matrix, and its ability to approximate matrices of a given rank with lower-rank matrices. Substituting a matrix with its SVD has often the advantage of making a calculation more robust to numerical rounding errors.