Singular value decomposition

The *singular value decomposition* of a matrix is usually referred to as the *SVD*. This is the final and best factorization of a matrix:

$$A = U\Sigma V^T$$

where U is orthogonal, Σ is diagonal, and V is orthogonal.

In the decomoposition $A = U\Sigma V^T$, A can be *any* matrix. We know that if A is symmetric positive definite its eigenvectors are orthogonal and we can write $A = Q\Lambda Q^T$. This is a special case of a SVD, with $U = V = Q$. For more general A, the SVD requires two different matrices U and V.

We've also learned how to write $A = S\Lambda S^{-1}$, where S is the matrix of n distinct eigenvectors of A. However, S may not be orthogonal; the matrices U and V in the SVD will be.

How it works

We can think of A as a linear transformation taking a vector $\mathbf{v}_1$ in its row space to a vector $\mathbf{u}_1 = A\mathbf{v}_1$ in its column space. The SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space: $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

It's not hard to find an orthogonal basis for the row space – the Gram-Schmidt process gives us one right away. But in general, there's no reason to expect A to transform that basis to another orthogonal basis.

You may be wondering about the vectors in the nullspaces of A and A^T. These are no problem – zeros on the diagonal of Σ will take care of them.

Matrix language

The heart of the problem is to find an orthonormal basis $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_r$ for the row space of A for which

$$\begin{aligned} A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} &= \begin{bmatrix} \sigma_1 \mathbf{u}_1 & \sigma_2 \mathbf{u}_2 & \cdots & \sigma_r \mathbf{u}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}, \end{aligned}$$

with $\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_r$ an orthonormal basis for the column space of A. Once we add in the nullspaces, this equation will become $AV = U\Sigma$. (We can complete the orthonormal bases $\mathbf{v}_1, ...\mathbf{v}_r$ and $\mathbf{u}_1, ...\mathbf{u}_r$ to orthonormal bases for the entire space any way we want. Since $\mathbf{v}_{r+1}, ...\mathbf{v}_n$ will be in the nullspace of A, the diagonal entries $\sigma_{r+1}, ...\sigma_n$ will be 0.)

The columns of U and V are bases for the row and column spaces, respectively. Usually $U \neq V$, but if A is positive definite we can use the *same* basis for its row and column space!

Calculation

Suppose A is the invertible matrix $\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We want to find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the row space $\mathbb{R}^2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ in the column space $\mathbb{R}^2$, and positive numbers σ_1 and σ_2 so that the $\mathbf{v}_i$ are orthonormal, the $\mathbf{u}_i$ are orthonormal, and the σ_i are the scaling factors for which $A\mathbf{v}_i = \sigma_i u_i$.

This is a big step toward finding orthonormal matrices V and U and a diagonal matrix Σ for which:

$$AV = U\Sigma.$$

Since V is orthogonal, we can multiply both sides by $V^{-1} = V^T$ to get:

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

Rather than solving for U, V and Σ simultaneously, we multiply both sides by $A^T = V\Sigma^T U^T$ to get:

$$\begin{aligned} A^T A &= V\Sigma U^{-1} U\Sigma V^T \\ &= V\Sigma^2 V^T \\ &= V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^T. \end{aligned}$$

This is in the form $Q\Lambda Q^T$; we can now find V by diagonalizing the symmetric positive definite (or semidefinite) matrix $A^T A$. The columns of V are eigenvectors of $A^T A$ and the eigenvalues of $A^T A$ are the values σ_i^2. (We choose σ_i to be the positive square root of λ_i.)

To find U, we do the same thing with AA^T.

SVD example

We return to our matrix $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We start by computing

$$\begin{aligned} A^T A &= \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}. \end{aligned}$$

The eigenvectors of this matrix will give us the vectors $\mathbf{v}_i$, and the eigenvalues will gives us the numbers σ_i.

Two orthogonal eigenvectors of A^TA are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. To get an orthonormal basis, let $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. These have eigenvalues $\sigma_1^2 = 32$ and $\sigma_2^2 = 18$. We now have:

$$\underset{A}{\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}} = \underset{U}{\begin{bmatrix} \; & \; \\ \; & \; \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix}} \underset{V^T}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}.$$

We could solve this for U, but for practice we'll find U by finding orthonormal eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ for $AA^T = U\Sigma^2U^T$.

$$\begin{aligned} AA^T &= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}. \end{aligned}$$

Luckily, AA^T happens to be diagonal. It's tempting to let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, as Professor Strang did in the lecture, but because $A\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3\sqrt{2} \end{bmatrix}$ we instead have $\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Note that this also gives us a chance to double check our calculation of σ_1 and σ_2.

Thus, the SVD of A is:

$$\underset{A}{\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}} = \underset{U}{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix}} \underset{V^T}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}.$$

Example with a nullspace

Now let $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$. This has a one dimensional nullspace and one dimensional row and column spaces.

The row space of A consists of the multiples of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$. The column space of A is made up of multiples of $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$. The nullspace and left nullspace are perpendicular to the row and column spaces, respectively.

Unit basis vectors of the row and column spaces are $\mathbf{v}_1 = \begin{bmatrix} .8 \\ .6 \end{bmatrix}$ and $\mathbf{u}_1 =$

$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. To compute σ_1 we find the nonzero eigenvalue of A^TA.

$$\begin{aligned} A^TA &= \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}. \end{aligned}$$

Because this is a rank 1 matrix, one eigenvalue must be 0. The other must equal the trace, so $\sigma_1^2 = 125$. After finding unit vectors perpendicular to $\mathbf{u}_1$ and $\mathbf{v}_1$ (basis vectors for the left nullspace and nullspace, respectively) we see that the SVD of A is:

$$\underset{A}{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}} = \underset{U}{\frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}.$$

The singular value decomposition combines topics in linear algebra ranging from positive definite matrices to the four fundamental subspaces.

$\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$	is an orthonormal basis for the row space.
$\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r$	is an orthonormal basis for the column space.
$\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$	is an orthonormal basis for the nullspace.
$\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$	is an orthonormal basis for the left nullspace.

These are the "right" bases to use, because $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.