Example of Full Singular Value Decomposition

SVD is based on a theorem from linear algebra which says that a rectangular matrix A can be broken down into the product of three matrices - an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V. The theorem is usually presented something like this:

$$A_{mn} = U_{mm} S_{mn} V_{nn}^T$$

where $U^T U = I, V^T V = I$; the columns of U are orthonormal eigenvectors of AA^T, the columns of V are orthonormal eigenvectors of $A^T A$, and S is a diagonal matrix containing the square roots of eigenvalues from U or V in descending order.

The following example merely applies this definition to a small matrix in order to compute its SVD. In the next section, I attempt to interpret the application of SVD to document classification.

Start with the matrix

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

In order to find U, we have to start with AA^T. The transpose of A is

$$A^T = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

so

$$AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

Next, we have to find the eigenvalues and corresponding eigenvectors of AA^T. We know that eigenvectors are defined by the equation $A\vec{v} = \lambda\vec{v}$, and applying this to AA^T gives us

$$\begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

We rewrite this as the set of equations

$$11x_1 + x_2 = \lambda x_1$$

$$x_1 + 11x_2 = \lambda x_2$$

and rearrange to get

$$(11 - \lambda)x_1 + x_2 = 0$$

$$x_1 + (11 - \lambda)x_2 = 0$$

Solve for λ by setting the determinant of the coefficient matrix to zero,

$$\begin{vmatrix} (11-\lambda) & 1 \\ 1 & (11-\lambda) \end{vmatrix} = 0$$

which works out as

$$(11-\lambda)(11-\lambda) - 1 \cdot 1 = 0$$

$$(\lambda - 10)(\lambda - 12) = 0$$

$$\lambda = 10, \lambda = 12$$

to give us our two eigenvalues $\lambda = 10, \lambda = 12$. Plugging λ back in to the original equations gives us our eigenvectors. For $\lambda = 10$ we get

$$(11-10)x_1 + x_2 = 0$$

$$x_1 = -x_2$$

which is true for lots of values, so we'll pick $x_1 = 1$ and $x_2 = -1$ since those are small and easier to work with. Thus, we have the eigenvector $[1, -1]$ corresponding to the eigenvalue $\lambda = 10$. For $\lambda = 12$ we have

$$(11-12)x_1 + x_2 = 0$$

$$x_1 = x_2$$

and for the same reason as before we'll take $x_1 = 1$ and $x_2 = 1$. Now, for $\lambda = 12$ we have the eigenvector $[1, 1]$. These eigenvectors become column vectors in a matrix ordered by the size of the corresponding eigenvalue. In other words, the eigenvector of the largest eigenvalue is column one, the eigenvector of the next largest eigenvalue is column two, and so forth and so on until we have the eigenvector of the smallest eigenvalue as the last column of our matrix. In the matrix below, the eigenvector for $\lambda = 12$ is column one, and the eigenvector for $\lambda = 10$ is column two.

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Finally, we have to convert this matrix into an orthogonal matrix which we do by applying the Gram-Schmidt orthonormalization process to the column vectors. Begin by normalizing $\vec{v_1}$.

$$\vec{u_1} = \frac{\vec{v_1}}{|\vec{v_1}|} = \frac{[1,1]}{\sqrt{1^2+1^2}} = \frac{[1,1]}{\sqrt{2}} = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$$

Compute

$$\vec{w_2} = \vec{v_2} - \vec{u_1} \cdot \vec{v_2} * \vec{u_1} =$$

$$[1,-1] - [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \cdot [1,-1] * [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] =$$

$$[1,-1] - 0 * [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] = [1,-1] - [0,0] = [1,-1]$$

and normalize

$$\vec{u_2} = \frac{\vec{w_2}}{|\vec{w_2}|} = [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]$$

to give

$$U = \left[\begin{array}{cc} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{array} \right]$$

The calculation of V is similar. V is based on A^TA, so we have

$$A^TA = \left[\begin{array}{cc} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{array} \right] \left[\begin{array}{ccc} 3 & 1 & 1 \\ -1 & 3 & 1 \end{array} \right] = \left[\begin{array}{ccc} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{array} \right]$$

Find the eigenvalues of A^TA by

$$\left[\begin{array}{ccc} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{array} \right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right] = \lambda \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right]$$

which represents the system of equations

$$10x_1 + 2x_3 = \lambda x_1$$

$$10x_2 + 4x_3 = \lambda x_2$$

$$2x_1 + 4x_2 + 2x_3 = \lambda x_2$$

which rewrite as

$$(10-\lambda)x_1 + 2x_3 = 0$$

$$(10-\lambda)x_2 + 4x_3 = 0$$

$$2x_1 + 4x_2 + (2-\lambda)x_3 = 0$$

which are solved by setting

$$\left| \begin{array}{ccc} (10-\lambda) & 0 & 2 \\ 0 & (10-\lambda) & 4 \\ 2 & 4 & (2-\lambda) \end{array} \right| = 0$$

This works out as

$$(10-\lambda) \left| \begin{array}{cc} (10-\lambda) & 4 \\ 4 & (2-\lambda) \end{array} \right| + 2 \left| \begin{array}{cc} 0 & (10-\lambda) \\ 2 & 4 \end{array} \right| =$$

$$(10-\lambda)[(10-\lambda)(2-\lambda)-16]+2[0-(20-2\lambda)]=$$
$$\lambda(\lambda-10)(\lambda-12)=0,$$

so $\lambda=0, \lambda=10, \lambda=12$ are the eigenvalues for A^TA. Substituting λ back into the original equations to find corresponding eigenvectors yields for $\lambda=12$

$$(10-12)x_1+2x_3=-2x_1+2x_3=0$$

$$x_1=1, x_3=1$$
$$(10-12)x_2+4x_3=-2x_2+4x_3=0$$
$$x_2=2x_3$$
$$x_2=2$$

So for $\lambda=12$, $\vec{v_1}=[1,2,1]$. For $\lambda=10$ we have

$$(10-10)x_1+2x_3=2x_3=0$$

$$x_3=0$$
$$2x_1+4x_2=0$$
$$x_1=-2x_2$$
$$x_1=2, x_2=-1$$

which means for $\lambda=10$, $\vec{v_2}=[2,-1,0]$. For $\lambda=0$ we have

$$10x_1+2x_3=0$$

$$x_3=-5$$
$$10x_1-20=0$$
$$x_2=2$$
$$2x_1+8-10=0$$
$$x_1=1$$

which means for $\lambda=0$, $\vec{v_3}=[1,2,-5]$. Order $\vec{v_1}$, $\vec{v_2}$, and $\vec{v_3}$ as column vectors in a matrix according to the size of the eigenvalue to get

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & 5 \end{bmatrix}$$

and use the Gram-Schmidt orthonormalization process to convert that to an orthonormal matrix.

$$\vec{u_1} = \frac{\vec{v_1}}{|\vec{v_1}|} = [\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}]$$

$$\vec{w_2} = \vec{v_2} - \vec{u_1} \cdot \vec{v_2} * \vec{u_1} = [2, -1, 0]$$

$$\vec{u_2} = \frac{\vec{w_2}}{|\vec{w_2}|} = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0]$$

$$\vec{w_3} = \vec{v_3} - \vec{u_1} \cdot \vec{v_3} * \vec{u_1} - \vec{u_2} \cdot \vec{v_3} * \vec{u_2} = [\frac{-2}{3}, \frac{-4}{3}, \frac{10}{3}]$$

$$\vec{u_3} = \frac{\vec{w_3}}{|\vec{w_3}|} = [\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}}]$$

All this to give us

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

when we really want its transpose

$$V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

For S we take the square roots of the non-zero eigenvalues and populate the diagonal with them, putting the largest in s_{11}, the next largest in s_{22} and so on until the smallest value ends up in s_{mm}. The non-zero eigenvalues of U and V are always the same, so that's why it doesn't matter which one we take them from. Because we are doing full SVD, instead of reduced SVD (next section), we have to add a zero column vector to S so that it is of the proper dimensions to allow multiplication between U and V. The diagonal entries in S are the singular values of A, the columns in U are called left singular vectors, and the columns in V are called right singular vectors.

$$S = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix}$$

Now we have all the pieces of the puzzle

$$A_{mn} = U_{mm} S_{mn} V_{nn}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\sqrt{12}}{\sqrt{2}} & \frac{\sqrt{10}}{\sqrt{2}} & 0 \\ \frac{\sqrt{12}}{\sqrt{2}} & \frac{-\sqrt{10}}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$