Singular Value Decomposition

Theorem I (Singular Value Decomposition)

Let **A** be a p by n matrix of real elements (not all zeroes) with $p \geq n$. Then there is a p by p orthogonal matrix **U**, an n by n orthogonal matrix **V**, and a p by n matrix Λ such that

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}' \quad \text{and} \quad \mathbf{U}'\mathbf{A}\mathbf{V} = \Lambda$$

where

$$\Lambda = \begin{bmatrix} \Lambda_n \\ \hline \mathbf{0} \end{bmatrix}$$

and $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_p$, $\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_n$, where $\mathbf{I}_p$ and $\mathbf{I}_n$ are p by p and n by n identity matrices respectively. Λ_n is an n by n diagonal matrix and **0** is a p-n by n matrix of zeroes. The diagonal entries of Λ_n are non-negative with exactly s entries strictly positive ($s \leq n$).

Theorem II – the famous Eckart-Young Theorem – solves the general least squares problem of approximating one matrix by another of lower rank. Geometrically, suppose the matrix is a set of p points in an n-dimensional space and we wish to find the best two-dimensional plane through the p points such that the distances from the points to the surface of the plane are minimized. Technically, let A be a p by n matrix of rank 15 and let B be a p by n matrix of rank 2. Given A, the problem is to find the matrix B such that $\sum_{i=1}^{p}\sum_{j=1}^{n}\left(a_{ij} - b_{ij}\right)^2$ is minimized.

Theorem II was never explicitly stated by Eckart and Young. Rather, they use two theorems from linear algebra (Theorem I was the first) and a very clever argument to show the truth of their result. Later, Keller (1962) independently rediscovered the Eckart-Young result (Theorem II).

Theorem II (Eckart and Young)

Given a p by n matrix **A** of rank $r \le n \le p$, and its singular value decomposition, $\mathbf{U\Lambda V'}$, with the singular values arranged in decreasing sequence

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \ldots \lambda_n \ge 0$$

then there exists a p by n matrix **B** of rank s, $s \le r$, which minimizes the sum of the squared error between the elements of **A** and the corresponding elements of **B** when

$$\mathbf{B} = \mathbf{U\Lambda_s V'}$$

where the diagonal elements of $\mathbf{\Lambda_s}$ are

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \ldots \lambda_s > \lambda_{s+1} = \lambda_{s+2} = \ldots = \lambda_n = 0$$

Theorem I states that every real matrix can be written as the product of two orthogonal matrices and one diagonal matrix. Theorem II states that the least squares approximation in s dimensions of a matrix **A** can be found by replacing the smallest n-s roots of Λ with zeroes and remultiplying $\mathbf{U\Lambda V'}$.

Because the lower p-n rows of Λ are all zeros, it is convenient to discard them and work only with the n by n diagonal matrix Λ_n . In addition, the p-n eigenvectors in **U** corresponding to the p-n lower rows of Λ may also be discarded. With these deletions of redundant rows and columns, **U** is a p by n matrix, Λ is an n by n diagonal matrix, and **V** is an n by n matrix. Hence $\mathbf{U'U = V'V = VV' = I_n}$. A decomposition according to Theorem I will be assumed to be in this form.

Example

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 3 & 2 & 1 & 3 \\ 4 & 3 & 1 & 4 \\ 2 & 1 & 3 & 1 \\ 1 & 5 & 2 & 2 \\ 1 & 2 & 2 & 3 \end{bmatrix} = \mathbf{U}\Lambda\mathbf{V}' =$$

$$\begin{bmatrix} -.380 & .120 & -.439 & .565 \\ -.404 & .345 & .057 & -.215 \\ -.545 & .429 & -.051 & -.432 \\ -.265 & -.068 & .884 & .215 \\ -.446 & -.817 & -.142 & -.321 \\ -.355 & -.102 & .004 & .546 \end{bmatrix} \begin{bmatrix} 11.485 & 0 & 0 & 0 \\ 0 & 3.270 & 0 & 0 \\ 0 & 0 & 2.653 & 0 \\ 0 & 0 & 0 & 2.089 \end{bmatrix} \begin{bmatrix} -.444 & -.558 & -.324 & -.621 \\ .556 & -.654 & -.351 & .374 \\ .435 & -.277 & .732 & -.444 \\ -.512 & -.428 & .485 & .526 \end{bmatrix}$$

Note that we can write Λ as the sum:

$$\begin{bmatrix} 11.485 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 3.270 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2.653 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.089 \end{bmatrix}$$

Which in symbols we can write as:

$$\Lambda = \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4$$

Hence,

$$\mathrm{A} = \mathbf{U}\Lambda\mathbf{V}' = \mathbf{U}\left[\Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4\right]\mathbf{V}' = \mathbf{U}\Lambda_1\mathbf{V}' + \mathbf{U}\Lambda_2\mathbf{V}' + \mathbf{U}\Lambda_3\mathbf{V}' + \mathbf{U}\Lambda_4\mathbf{V}'$$

Now, observe that

$$\mathbf{U}\Lambda_1\mathbf{V}' = \begin{bmatrix} -.380 \\ -.404 \\ -.545 \\ -.265 \\ -.446 \\ -.355 \end{bmatrix} (11.485) \begin{bmatrix} -.444 & -.558 & -.324 & -.621 \end{bmatrix}$$

Because of the columns of zeroes in Λ_1

To see this, note that

$$\begin{bmatrix} -.380 & .120 & -.439 & .565 \\ -.404 & .345 & .057 & -.215 \\ -.545 & .429 & -.051 & -.432 \\ -.265 & -.068 & .884 & .215 \\ -.446 & -.817 & -.142 & -.321 \\ -.355 & -.102 & .004 & .546 \end{bmatrix} \begin{bmatrix} 11.485 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = (11.485) \begin{bmatrix} -.380 & 0 & 0 & 0 \\ -.404 & 0 & 0 & 0 \\ -.545 & 0 & 0 & 0 \\ -.265 & 0 & 0 & 0 \\ -.446 & 0 & 0 & 0 \\ -.355 & 0 & 0 & 0 \end{bmatrix} = \mathbf{U\Lambda_1}$$

because the columns of zeroes cancel. When $\mathbf{U\Lambda_1}$ is multiplied through $\mathbf{V'}$ the corresponding rows of $\mathbf{V'}$ are multiplied by zero so they disappear as well. This fact allows us to write $\mathbf{U\Lambda_1}$ V' as the sum:

$$\mathbf{A} = \mathbf{U\Lambda V'} = \mathbf{u}_1\lambda_1\mathbf{v}_1' + \mathbf{u}_2\lambda_2\mathbf{v}_2' + \mathbf{u}_3\lambda_3\mathbf{v}_3' + \mathbf{u}_4\lambda_4\mathbf{v}_4'$$

If you want a matrix B of rank 3 that is the best least squares approximation to A, then it is

$$\mathbf{B} = \mathbf{u}_1\lambda_1\mathbf{v}_1' + \mathbf{u}_2\lambda_2\mathbf{v}_2' + \mathbf{u}_3\lambda_3\mathbf{v}_3'$$

The residual matrix is

$$\mathbf{E} = \mathbf{A} - \mathbf{B} = \mathbf{u}_4\lambda_4\mathbf{v}_4'$$

And the sum of the squared residuals is λ_4^2 (recall that the sum of squares of all the elements in A is $\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2$. In this example,

$$(1^2 + 2^2 + 1^2 + 4^2 + 3^2 + 2^2 + 1^2 + 3^2 + 4^2 + 3^2 + 1^2 + 4^2 + 2^2 + 1^2 + 3^2 + 1^2 + 1^2 + 5^2 + 2^2$$
$$+ 2^2 + 1^2 + 2^2 + 2^2 + 3^2) = 154 = (11.485^2 + 3.270^2 + 2.653^2 + 2.089^2)$$