A Short Note on Singular Value Decomposition and Least-Squares

Shuo Han

January 27, 2011

The singular value decomposition is one of the most useful matrix decompositions in numerical linear algebra and many fields of science and engineering. Unfortunately, it not usually covered in an introductory linear algebra course, although it is a straightforward (and powerful) extension of eigenvalue decomposition. The SVD of a real $m \times n$ rectangular matrix $A \in \mathbb{R}^{m\times n}$ is given by[1]

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

where $U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$. The matrix Σ is of the following form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_m & \cdots & 0 \end{bmatrix} \quad (m < n),$$

or

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \vdots \\ & & & 0 \end{bmatrix} \quad (m > n).$$

Also, both U and V are orthonormal. A matrix U is orthonormal if its columns (denoted as u_i) satisfy

$$u_i^T u_j = \delta_{ij}, \quad \forall i, j,$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$.

You may wonder why it is V^T but not V in the decomposition, and the following should give you some hints. Without loss of generality, suppose $m > n$, let us see what will happen if we apply the matrix A to any column v_i of V:

$$Av_i = U\Sigma V^T v_i = U\Sigma \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} v_i = U\Sigma \begin{bmatrix} v_1^T v_i \\ v_2^T v_i \\ \vdots \\ v_n^T v_i \end{bmatrix} = U\Sigma e_i = \sigma_i u_i,$$

[1] The SVD can be extended to complex matrices with little effort.

where $e_i \in \mathbb{R}^n$ is a vector whose entries are all zeros except that the i-th one is 1. You verify the above steps and make sure every step makes sense. On the other hand, by left multiplying u_i^T, you should be able to get v_i^T:

$$u_i^T A = \sigma_i v_i^T.$$

If we take the 2-norm of the above result on Av_i, we obtain

$$\|Av_i\|_2 = \|\sigma_i u_i\|_2 = \sqrt{(\sigma_i u_i)^T(\sigma_i u_i)} = \sigma_i \sqrt{u_i^T u_i} = \sigma_i.$$

If A is not full rank, then there exists some v_i is in the (right) kernel space of A, i.e. $Av_i = 0$ and $\sigma_i = 0$. However, in practice, because of noise and other factors, the corresponding singular value is not exactly zero (albeit it should be close to zero). In this case, to find a vector x that "best solves" $Ax = 0$, you can still make $x = v_i$, which is the column vector that corresponds to the smallest singular value $\sigma_i = \sigma_{\min}$, so that the norm (think of is as the "size") of Ax will be the smallest. More formally, we call this the best solution in the least-squares (i.e., 2-norm) sense (think about why we need the constraint $\|x\|_2 = 1$ below):

$$x_{\text{LS}}^* = \arg\min_x \|Ax\|_2 \quad \text{subject to} \quad \|x\|_2 = 1.$$