SVD

Deva Ramanan

January 27, 2016

Let us represent a linear transformation as follows:

$$y = Ax, \quad A \in R^{n \times m} \tag{1}$$

where A is a matrix with n columns and m rows. This document uses the singular value decomposition (SVD) to decompose A into a series of geometric transformations, focusing intuition rather than a precise formulation. For simplicity, let $n = 2$ and $m = 3$, such that A transforms points in 2D to 3D.

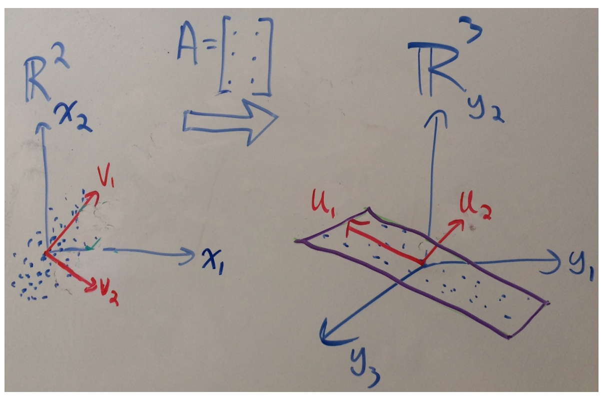

Figure 1: Visualizing a matrix $A \in R^{2\times 3}$ as a transformation of points from R^2 to R^3.

Orthonormal basis: First, let us recall that the projection of a vector $x \in R^n$ along a unit vector v (e.g., $v^T v = 1$) can be written as $v^T x$. Let us construct a set of n unit vectors and write them as a matrix

$$V = \begin{bmatrix} v_1 & v_2, \dots v_n \end{bmatrix}.$$

We can then compute the projection or *coordinates* of vector x along the unit vectors with a matrix multiplication $p = V^T x$. If all the unit vectors are orthogonal to each other ($v_i^T v_j = 0$ for $i \neq j$), then $V^T V = I$. This implies that V can be thought off as a rotation matrix (whos inverse is V^T), making it easy to undo the projection. The set of vectors in V form an *orthogonal basis* for R^n. Let us similar construct an orthonormal basis for the output space

$$U = \begin{bmatrix} u_1 & u_2 \dots \end{bmatrix}$$

.

SVD: An SVD allows us to characterize any linear operation $y = Ax$ for $A \in R^{n \times m}$ as follows:

1. *Project* x into an orthonormal basis $p = V^T x$ for the input space.
2. *Scale* the coordinates by values σ_1, σ_2, which can be written as $c = \Sigma p$, where $\Sigma = R^{n \times m}$ is a diagonal matrix:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \ldots \\ 0 & \sigma_2 & 0 \ldots \\ 0 & 0 & \sigma_3 \ldots \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{2}$$

3. *Reconstruct* a point in output space by taking a linear combination of the output basis vectors, $y = c_1 u_1 + c_2 u_2 \ldots = Uc$.

This allows us to write $y = Ax = U\Sigma V^T$. The SVD of A produces the three matrices U, Σ, V such that $U^T U = I, V^T V = I, \Sigma =$ diagonal. We typically use the terms *left-singular vectors, singular values, and right-singular vectors* to describe the three matrices due to the fact that $Av_i = \sigma_i u_i$ and $u_i^T A = \sigma_i v_i^T$, which somewhat resemble the definition of an eigenvector (more on this below). If $m > n$, than we only use n (out of the m) output basis vectors during reconstruction. If $n > m$, then we only project into m (out of the n) input basis vectors during the projection.

Proof (sketch): One can prove this by forming either of the two square matrices $B = A^T A$ and $C = AA^T$. It is straightforward to show that B and C are symmetric and positive semi-definite (PSD). One can then use properties of PSD matrices - namely, eigenvalues must be positive and the eigenvectors form an orthonormal basis. Because $B = V\Sigma^T U^T U \Sigma V^T = V\Sigma^2 V^T$, $Bv_i = \sigma_i^2 v_i$, implying the eigenvectors and eigenvalues of B are given by the square singular values and right singular vectors of A. An analogous argument based on the left singular vector holds for C.

Corollary 0: Any symmetric PSD matrix B can be decomposed as $B = V\Sigma^2 V^T$, where v_i are eigenvectors with non-negative eigenvalues σ^2. The proof follows by the property that any PSD matrix can be written as the product of two matrices $B = A^T A$ for some matrix A. As an aside, this is how I intuitively think of a PSD matrix B: there is an underlying transformation A that it corresponds to. This decomposition is sometimes called a *spectral eigendecomposition*, and can be geometrically viewed as a *projection* onto a rotated basis, *scaling* along that basis, and a *reconstruction* using the same basis vectors.

Corollary 1: The best k-rank approximation of A is given by constructing k largest singular vectors. In matlab notation:

$$\min_{A':rank(A')<=k} ||A - A'||_F = U(:, 1:k)\Sigma(1:k, 1:k)V(:, 1:k)^T, \quad \text{where} \quad ||A||_F = A(:)^T A(:) \tag{3}$$

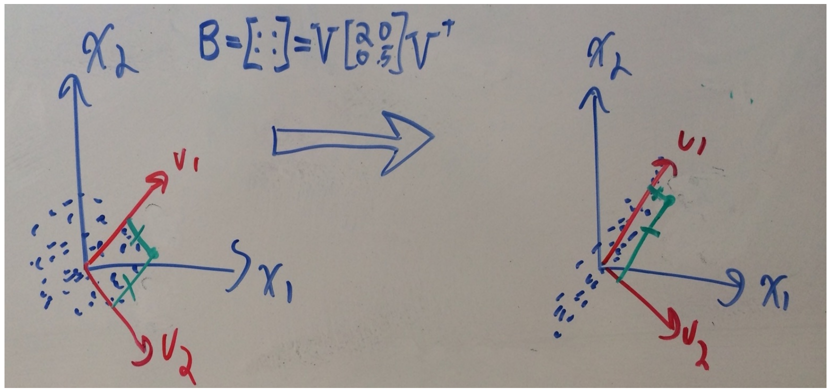

Figure 2: Visualizing the spectral eigendecomposition of a symmetric PSD matrix.

This makes intuitive sense geometrically; taking the k largest singular values and vectors produces a transformation A' that uses as much of the output space as possible. The sketch of the proof relies on the fact that U and V act as rotations and so do not effect the rank of A. The best k-rank approximation of A is then given by the best k-rank approximation of the (diagonal) matrix Σ.

Corollary 2: The solution of a homogenous least squares problem is given by smallest right singular value:

$$\min_{h:h^T h=1} ||Ah||^2 = V(:, end)$$

The proof sketch follows by the fact that any input v must project to one of the right singular vectors (because they form a basis). A closely related result is that for any PSD matrix $B = A^T A$, $\min_{h:h^T h=1} h^T Bh = V(:, end)$, where $V(:, end)$ the eigenvector with the smallest eigenvalue.

Corollary 3: The pseudoinverse of A is given by

$$A^+ = \underset{A^+}{\operatorname{argmin}} ||A^+ A - I||_F = V \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 \dots \\ 0 & \frac{1}{\sigma_2} & 0 \dots \\ 0 & 0 & \frac{1}{\sigma_3} \dots \\ \vdots & \vdots & \vdots \end{bmatrix}^T U^T$$

which could also be obtained by mimimizing $||AA^+ - I||_F$ (without proof).