EE16B - Spring'20 - Lecture 11A Notes[1]

Murat Arcak

31 March 2020

[1] Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Singular Value Decomposition (SVD) Continued

Recall that SVD separates a rank-r matrix $A \in \mathbb{R}^{m\times n}$ into a sum of r rank-1 matrices:

$$\boxed{A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T} \tag{1}$$

where $\vec{u}_1, \ldots, \vec{u}_r \in \mathbb{R}^m$ are orthonormal, $\vec{v}_1, \ldots, \vec{v}_r \in \mathbb{R}^n$ are orthonormal, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

In most textbooks the SVD (1) is written as a product of three matrices. To derive this alternative form we first rewrite (1) as

$$A = U_1 S V_1^T \tag{2}$$

where $U_1 = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_r \end{bmatrix}$ is $m \times r$, $V_1 = \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix}$ is $n \times r$, and S is the $r \times r$ diagonal matrix with entries $\sigma_1, \ldots, \sigma_r$:

$$S = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}.$$

Recall that $\vec{u}_1, \cdots, \vec{u}_r$ correspond to eigenvectors of AA^T for non-zero eigenvalues, and similarly $\vec{v}_1, \cdots, \vec{v}_r$ are eigenvectors of $A^T A$ for non-zero eigenvalues.

Next we form the $m \times m$ orthonormal matrix

$$U = [U_1 \; U_2]$$

where the columns of $U_2 = [\vec{u}_{r+1} \; \cdots \; \vec{u}_m]$ are eigenvectors of AA^T corresponding to *zero* eigenvalues. Likewise we define $V_2 = [\vec{v}_{r+1} \; \cdots \; \vec{v}_n]$ whose columns are orthonormal eigenvectors of $A^T A$ for zero eigenvalues, and obtain the $n \times n$ orthogonal matrix

$$V = [V_1 \; V_2].$$

Then we write

$$\boxed{A = U \underbrace{\begin{bmatrix} S & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}}_{=:\,\Sigma} V^T} \tag{3}$$

which is identical to (2) but exhibits square and orthonormal matrices U and V^T, that is $U^T U = I$ and $V^T V = I$. Having square and orthonormal matrices U and V will allow us to give a geometric interpretation of SVD in the next section.

It is important to understand the dimensions of the matrices in (3). Σ is $m \times n$, same as A. U and V, however, are square: U is $m \times m$ and V is $n \times n$. If A is square ($m = n$), then all three are square. If A is a wide matrix with full row rank ($r = m < n$), then

$$A = U \underbrace{\begin{bmatrix} S & 0_{m \times (n-m)} \end{bmatrix}}_{= \Sigma} V^T$$

If A is a tall matrix with full column rank ($m > n = r$), then

$$A = U \underbrace{\begin{bmatrix} S \\ 0_{(m-n) \times n} \end{bmatrix}}_{= \Sigma} V^T$$

Geometric Interpretation of SVD

Note that multiplying a vector by an orthonormal matrix does not change its length. This follows because $U^T U = I$, which implies

$$\|U\vec{x}\|^2 = (U\vec{x})^T (U\vec{x}) = \vec{x}^T U^T U \vec{x} = \vec{x}^T \vec{x} = \|\vec{x}\|^2.$$

Thus we can interpret multiplication by an orthonormal matrix as a combination of operations that don't change length, such as rotations, and reflections.

Since S is diagonal with entries $\sigma_1, \ldots, \sigma_r$, multiplying a vector by Σ defined in (3) stretches the first entry of the vector by σ_1, the second entry by σ_2, and so on.

Combining these observations we interpret $A\vec{x}$ as the composition of three operations:

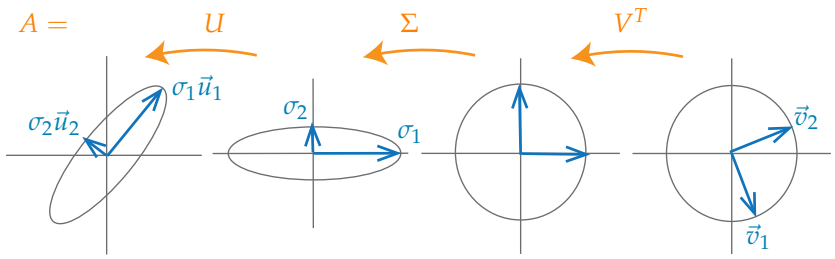
1) $V^T \vec{x}$ which reorients $\vec{x}$ without changing its length,

2) $\Sigma V^T \vec{x}$ which stretches the resulting vector along each axis with the corresponding singular value,

3) $U \Sigma V^T \vec{x}$ which again reorients the resulting vector without changing its length.

The figure below illustrates these three operations moving from the right to the left:

The geometric interpretation above reveals that σ_1 is the largest amplification factor a vector can experience upon multiplication by A:



if the length of $\vec{x}$ is $\|\vec{x}\| = 1$ then $\|A\vec{x}\| \leq \sigma_1$.

For $\vec{x} = \vec{v}_1$ we get $\|A\vec{x}\| = \sigma_1$ with *equality* because $V^T\vec{v}_1$ is the first unit vector which, when multiplied by Σ, gets stretched by σ_1.

Symmetric Matrices

We say that a square matrix Q is *symmetric* if

$$Q = Q^T.$$

Note that the matrices A^TA and AA^T we used to compute a SVD for A are automatically symmetric: using the identities $(AB)^T = B^TA^T$ and $(A^T)^T = A$ you can verify $(A^TA)^T = A^TA$ and $(AA^T)^T = AA^T$.

Below we derive important properties of symmetric matrices that we used without proof in our SVD procedures.

A symmetric matrix has real eigenvalues and eigenvectors.

Let Q be symmetric and let

$$Qx = \lambda x, \tag{4}$$

that is λ is an eigenvalue and x is an eigenvector. Let $\lambda = a + jb$ and define the conjugate $\bar{\lambda} = a - jb$. To show that $b = 0$, that is λ is real, we take conjugates of both sides of $Qx = \lambda x$ to obtain

$$Q\bar{x} = \bar{\lambda}\bar{x} \tag{5}$$

where we used the fact that Q is real. The transpose of (5) is

$$\bar{x}^TQ^T = \bar{\lambda}\bar{x}^T \tag{6}$$

and, since $Q = Q^T$, we write

$$\bar{x}^TQ = \bar{\lambda}\bar{x}^T. \tag{7}$$

Now multiply (4) from the left by $\bar{x}^T$ and (7) from the right by x:

$$\begin{aligned} \bar{x}^TQx &= \lambda\bar{x}^Tx \\ \bar{x}^TQx &= \bar{\lambda}\bar{x}^Tx \end{aligned} \tag{8}$$

Since the left hand sides are the same we have $\lambda \bar{x}^T x = \bar{\lambda} \bar{x}^T x$, and since $\bar{x}^T x \neq 0$, we conclude $\lambda = \bar{\lambda}$. This means $a + jb = a - jb$ which proves that $b = 0$.

Now that we know the eigenvalues are real we can conclude the eigenvectors are also real, because they are obtained from the equation $(Q - \lambda I)x = 0$ where $Q - \lambda I$ is real. □

The eigenvectors can be chosen to be orthonormal.

We will prove this for the case where the eigenvalues are distinct although the statement is true also without this restriction[2]. Orthonormality of the eigenvectors means they are orthogonal and each has unit length. Since we can easily normalize the length to one, we need only to show that the eigenvectors are orthogonal.

[2] A further fact is that a symmetric matrix admits a complete set of eigenvectors even in the case of repeated eigenvalues and is thus diagonalizable.

Pick two eigenvalue-eigenvector pairs: $Qx_1 = \lambda_1 x_1$, $Qx_2 = \lambda_2 x_2$, $\lambda_1 \neq \lambda_2$. Multiply $Qx_1 = \lambda_1 x_1$ from the left by x_2^T, and $Qx_2 = \lambda_2 x_2$ by x_1^T:

$$\begin{aligned} x_2^T Q x_1 &= \lambda_1 x_2^T x_1 \\ x_1^T Q x_2 &= \lambda_2 x_1^T x_2. \end{aligned} \tag{9}$$

Note that $x_2^T Q x_1$ is a scalar, therefore its transpose is equal to itself: $x_1^T Q x_2 = (x_1^T Q x_2)^T = x_2^T Q^T x_1 = x_2^T Q x_1$. This means that the left hand sides of the two equations above are identical, hence

$$\lambda_1 x_2^T x_1 = \lambda_2 x_1^T x_2.$$

Note that $x_1^T x_2 = x_2^T x_1$ is the inner product of x_1 and x_2. Since $\lambda_1 \neq \lambda_2$, the equality above implies that this inner product is zero, that is x_1 and x_2 are orthogonal. □

The final property below proves our earlier assertion that AA^T and $A^T A$ have nonnegative eigenvalues. (Substitute $R = A^T$ below for the former, and $R = A$ for the latter.)

If Q can be written as $Q = R^T R$ for some matrix R, then the eigenvalues of Q are nonnegative.

To show this let x_i be an eigenvector of Q corresponding λ_i, so that

$$R^T R x_i = \lambda_i x_i.$$

Next multiply both sides from the left by x_i^T:

$$x_i^T R^T R x_i = \lambda_i x_i^T x_i = \lambda_i \|x_i\|^2.$$

If we define $y = Rx_i$ we see that the left hand side is $y^T y = \|y\|^2$, which is nonnegative. Thus, $\lambda_i \|x_i\|^2 \geq 0$. Since the eigenvector is nonzero, we have $\|x_i\| \neq 0$ which implies $\lambda_i \geq 0$. □