EE16B - Fall'16 - Lecture 9A Notes[1]

Murat Arcak

25 October 2016

Singular Value Decomposition (SVD)

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 $\sigma_1, \ldots, \sigma_r$ are real, positive numbers called *singular values*. By convention, we order them from the largest to smallest:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

Finding a SVD

To find a SVD of the form (1) we use either the $n \times n$ matrix $A^T A$ or the $m \times m$ matrix AA^T. We will see later that these matrices have only *real eigenvalues*, r of which are positive and the remaining zero, and a complete set of *orthonormal eigenvectors*. For now we take this as a fact and propose the following procedures to find a SVD for A:

SVD procedure using $A^T A$:

1. Find the eigenvalues λ_i of $A^T A$ and order them from the largest to smallest, so that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$.

2. Find orthonormal eigenvectors $\vec{v}_i$, so that

$$A^T A\, \vec{v}_i = \lambda_i \vec{v}_i \quad i = 1, \ldots, r. \tag{2}$$

3. Let $\sigma_i = \sqrt{\lambda_i}$ and obtain $\vec{u}_i$ from

$$A\vec{v}_i = \sigma_i \vec{u}_i \quad i = 1, \ldots, r. \tag{3}$$

Justification: To see that $\vec{u}_i$, $i = 1, \ldots, r$, obtained from (3) are orthonormal, multiply (3) from the left by $(A\vec{v}_j)^T = \sigma_j \vec{u}_j^T$:

$$(A\vec{v}_j)^T A\vec{v}_i = \sigma_j \sigma_i \vec{u}_j^T \vec{u}_i. \tag{4}$$

The left hand side is $\vec{v}_j^T A^T A \vec{v}_i = \vec{v}_j^T \lambda_i \vec{v}_i = \sigma_i^2 \vec{v}_j^T \vec{v}_i$, therefore

$$\sigma_i^2 \vec{v}_j^T \vec{v}_i = \sigma_j \sigma_i \vec{u}_j^T \vec{u}_i. \tag{5}$$

The vectors $\vec{v}_i$, $i = 1, \ldots, r$, are orthonormal by construction, which means $\vec{v}_j^T \vec{v}_i = 1$ if $i = j$, and 0 if $i \neq j$. Thus, (5) becomes

$$\sigma_j \sigma_i \vec{u}_j^T \vec{u}_i = \begin{cases} \sigma_i^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{6}$$

and $\sigma_j \sigma_i$ cancels with σ_i^2 when $i = j$, proving orthonormality of $\vec{u}_i$, $i = 1, \ldots, r$.

To see why σ_i, $\vec{u}_i$, $\vec{v}_i$ resulting from the procedure above satisfy (1), rewrite (3) in matrix form as:

$$A \underbrace{\begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix}}_{\triangleq V_1} = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}.$$

Next, multiply both sides from the right by V_1^T:

$$AV_1V_1^T = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \underbrace{\begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_r^T \end{bmatrix}}_{V_1^T} \tag{7}$$

and note that the right hand side is indeed the decomposition in (1). We conclude by showing that the left hand side is equal to A. To this end define $V_2 = \begin{bmatrix} \vec{v}_{r+1} & \cdots & \vec{v}_n \end{bmatrix}$ whose columns are the remaining orthonormal eigenvectors for $\lambda_{r+1} = \cdots = \lambda_n = 0$. Then $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ is an orthonormal matrix and, thus,

$$VV^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = V_1V_1^T + V_2V_2^T = I.$$

Multiplying both sides from the left by A, we get

$$AV_1V_1^T + AV_2V_2^T = A. \tag{8}$$

Since the columns of V_2 are eigenvectors of A^TA for zero eigenvalues we have $A^TAV_2 = 0$, and multiplying this from the left by V_2^T we get $V_2^TA^TAV_2 = (AV_2)^T(AV_2) = 0$. This implies $AV_2 = 0$ and it follows from (8) that $AV_1V_1^T = A$. Thus, the left hand side of (7) is A, which proves that σ_i, $\vec{u}_i$, $\vec{v}_i$ proposed by the procedure above satisfy (1). □

An alternative approach is to use the $m \times m$ matrix AA^T which is preferable to using the $n \times n$ matrix A^TA when $m < n$. Below we summarize the procedure and leave its justification as an exercise.

SVD procedure using AA^T:

1. Find the eigenvalues λ_i of AA^T and order them from the largest to smallest, so that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_m = 0$.

2. Find orthonormal eigenvectors $\vec{u}_i$, so that

$$AA^T \vec{u}_i = \lambda_i \vec{u}_i \quad i = 1, \ldots, r. \tag{9}$$

3. Let $\sigma_i = \sqrt{\lambda_i}$ and obtain $\vec{v}_i$ from

$$A^T \vec{u}_i = \sigma_i \vec{v}_i \quad i = 1, \ldots, r. \tag{10}$$

Example: Let's follow this procedure to find a SVD for

$$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}.$$

We calculate

$$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$$

which happens to be diagonal, so the eigenvalues are $\lambda_1 = 32$, $\lambda_2 = 18$, and we can select the orthonormal eigenvectors:

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{11}$$

The singular values are $\sigma_1 = \sqrt{\lambda_1} = 4\sqrt{2}$, $\sigma_2 = \sqrt{\lambda_2} = 3\sqrt{2}$ and, from (10),

$$\vec{v}_1 = \frac{1}{\sigma_1} A^T \vec{u}_1 = \frac{1}{4\sqrt{2}} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
$$\vec{v}_2 = \frac{1}{\sigma_2} A^T \vec{u}_2 = \frac{1}{3\sqrt{2}} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

which are indeed orthonormal. We leave it as an exercise to derive a SVD using, instead, $A^T A$.

Note that we can change the signs of $\vec{u}_1$ and $\vec{u}_2$ in (11), and they still serve as orthonormal eigenvectors. This implies that SVD is not unique. However, changing the sign of $\vec{u}_i$ changes the sign of $\vec{v}_i$ in (10) accordingly, therefore the product $\vec{u}_i \vec{v}_i^T$ remains unchanged.

Another source of non-uniqueness arises when we have repeated singular values, as illustrated in the next example.

Example: To find a SVD for

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

note that AA^T is the identity matrix, which has repeated eigenvalues at $\lambda_1 = \lambda_2 = 1$ and admits any pair of orthonormal vectors as eigenvectors. We parameterize all such pairs as

$$\vec{u}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \tag{12}$$

where θ is a free parameter. Since $\sigma_1 = \sigma_2 = 1$, we obtain from (10):

$$\begin{aligned} \vec{v}_1 &= \frac{1}{\sigma_1} A^T \vec{u}_1 = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix} \\ \vec{v}_2 &= \frac{1}{\sigma_2} A^T \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ -\cos\theta \end{bmatrix}. \end{aligned} \tag{13}$$

Thus, (12)-(13) with $\sigma_1 = \sigma_2 = 1$ constitute a valid SVD for any choice of θ. You can indeed verify that

$$\begin{aligned} \vec{u}_1 \vec{v}_1^T + \vec{u}_2 \vec{v}_2^T = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \end{bmatrix} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \end{aligned} \tag{14}$$