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

Murat Arcak

27 October 2016

Geometric Interpretation of SVD

To develop a geometric interpretation of SVD, we first rewrite

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

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 $\sigma_1, \ldots, \sigma_r$.

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

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

which is identical to (2) but exhibits square and orthonormal matrices U and V^T that enable the geometric interpretation below.

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.

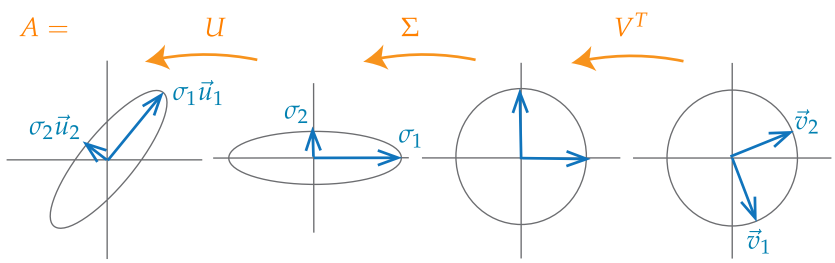
Combining these observations we interpret $A\vec{x}$ as the composition of three operations:

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

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

Now multiply (4) from the left by $\bar{x}^T$ and (6) from the right by x:

$$\begin{aligned} \bar{x}^T Qx &= \lambda \bar{x}^T x \\ \bar{x}^T Qx &= \bar{\lambda}\bar{x}^T x \end{aligned} \tag{7}$$

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:

$$\begin{aligned} x_2^T Qx_1 &= \lambda_1 x_2^T x_1 \\ x_1^T Qx_2 &= \lambda_2 x_1^T x_2. \end{aligned} \tag{8}$$

Note that $x_2^T Qx_1$ is a scalar, therefore its transpose is equal to itself: $x_1^T Qx_2 = x_2^T Qx_1$. 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 $\lambda_i \geq 0$. □

Principal Component Analysis (PCA)

PCA is an application of SVD in statistics that aims to find the most informative directions in a data set.

Suppose the $m \times n$ matrix A contains m measurements from n samples, for example m test scores for n students. If we subtract from each measurement the average over all samples, then each row of A is an n-vector with zero mean, and the $m \times m$ matrix

$$\frac{1}{n-1} A A^T$$

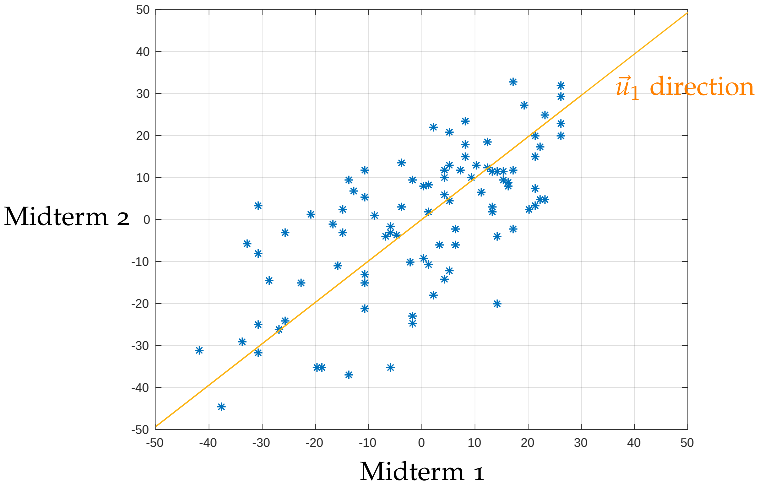
constitutes what is called the "covariance matrix" in statistics. Recall that the eigenvalues of this matrix are the singular values of A except for the scaling factor $n-1$, and its orthonormal eigenvectors correspond to $\vec{u}_1, \ldots, \vec{u}_m$ in the SVD of A.

The vectors $\vec{u}_1, \vec{u}_2, \ldots$ corresponding to large singular values are called principal components and identify dominant directions in the data set along which the samples are clustered. The most significant direction is $\vec{u}_1$ corresponding to σ_1.

As an illustration, the scatter plot below shows $m = 2$ midterm scores in a class of $n = 94$ students that I taught in the past. The data points are centered around zero because the class average is subtracted from the test scores. Each data point corresponds to a student and those in the first quadrant (both midterms ≥ 0) are those students who scored above average in each midterm. You can see that there were students who scored below average in the first and above average in the second, and vice versa.

For this data set the covariance matrix is:

$$\frac{1}{93} A A^T = \begin{bmatrix} 297.69 & 202.53 \\ 202.53 & 292.07 \end{bmatrix}$$



where the diagonal entries correspond to the squares of the standard deviations 17.25 and 17.09 for Midterms 1 and 2, respectively. The positive sign of the $(1,2)$ entry implies a positive correlation between the two midterm scores as one would expect.

The eigenvalues of AA^T, that is the singular values of A are $\sigma_1 =$ 215.08 , $\sigma_2 = 92.66$, and the corresponding eigenvectors of AA^T are:

$$\vec{u}_1 = \begin{bmatrix} 0.7120 \\ 0.7022 \end{bmatrix} \quad \vec{u}_2 = \begin{bmatrix} -0.7022 \\ 0.7120 \end{bmatrix}.$$

The principal component $\vec{u}_1$ is superimposed on the scatter plot and we see that the data is indeed clustered around this line. Note that it makes an angle of $\tan^{-1}(0.7022/0.7120) \approx 44.6^\circ$ which is skewed slightly towards the Midterm 1 axis because the standard deviation in Midterm 1 was slightly higher than in Midterm 2. We may interpret the points above this line as students who performed better in Midterm 2 than in Midterm 1, as measured by their scores relative to the class average that are then compared against the factor $\tan(44.6^\circ)$ to account for the difference in standard deviations.

The $\vec{u}_2$ direction, which is perpendicular to $\vec{u}_1$, exhibits less variation than the $\vec{u}_1$ direction ($\sigma_2 = 92.66$ vs. $\sigma_1 = 215.08$), but enough to convince you that you can do better on the final!