Lecture 5

SVD II

5.1 Computing the SVD (review)

To compute the SVD $U\Sigma V^*$ of a matrix $A \in \mathbb{C}^{m\times n}$ with $\operatorname{rank} A = r$:

1. Form the product A^*A.
2. Identify the r positive eigenvalues of A^*A. Call them λ_i.
3. Identify r orthonormal eigenvectors v_i of A^*A such that $Av_i = \lambda_i$.
4. Define $\sigma_i = \sqrt{\lambda_i}$.
5. Define $u_i = \sigma_i^{-1} A v_i$.
6. Thus far U is an $m \times r$ matrix, Σ is an $r \times r$ matrix, and V is an $n \times r$ matrix. The factorization $A = U\Sigma V^*$ is sometimes reported at this stage, and is called the **truncated SVD**.
7. If the **full SVD** is desired, then complete the columns of U and V to orthonormal bases, and pad Σ with zeros so that it is $m \times n$.

Computation example

We will compute the SVD of the matrix A.

$$A = \begin{pmatrix} 1 & -j \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.1}$$

$$A^*A = \begin{pmatrix} 1 & 0 & 1 \\ j & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -j \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.2}$$

$$= \begin{pmatrix} 2 & -j \\ j & 2 \end{pmatrix} \tag{5.3}$$

$$\chi(s) = s^2 - 4s + 3 = (s-3)(s-1) \tag{5.4}$$

First eigenvalue and eigenvector:

$$\lambda_1 = 3 \tag{5.5}$$

$$(A^*A - 3I)\, v_1 = \begin{pmatrix} -1 & -j \\ j & -1 \end{pmatrix} v_1 = 0 \tag{5.6}$$

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix} \tag{5.7}$$

Second eigenvalue and eigenvector:

$$\lambda_2 = 1 \tag{5.8}$$

$$(A^*A - I)\, v_2 = \begin{pmatrix} 1 & -j \\ j & 1 \end{pmatrix} v_2 = 0 \tag{5.9}$$

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} j \\ 1 \end{pmatrix} \tag{5.10}$$

Singular values:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \tag{5.11}$$

$$\sigma_2 = \sqrt{\lambda_2} = 1 \tag{5.12}$$

Left singular vectors:

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ j \\ 1 \end{pmatrix} \tag{5.13}$$

$$u_2 = \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ j \end{pmatrix} \tag{5.14}$$

Truncated SVD in the factorization style:

$$A = U_t \Sigma_t V_t^* = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \\ \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.15}$$

and in the dyad style:

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^* = \sqrt{3} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{j}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \end{pmatrix} + 1 \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{j}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.16}$$

For a full SVD, make U square by orthogonalizing and normalizing the columns of $\begin{pmatrix} U_t & I \end{pmatrix}$ from left to right, dropping zero columns.

$$\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & 1 & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{3} & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{3} & 0 & 1 \end{pmatrix} \quad \text{(orthogonalize col. 3)} \tag{5.17}$$

$$\rightsquigarrow \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{\sqrt{3}} & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & 0 & 1 \end{pmatrix} \quad \text{(normalize col. 3)} \tag{5.18}$$

As there are three orthonormal columns, we are done. The following is the full SVD of A:

$$A = U_f \Sigma_f V_f^* = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \\ \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.19}$$

Note that U_f and V_f are both invertible, but Σ_f has rank 2, which is the rank of A.

5.2 SVD of a wide matrix

To compute the SVD as we proved its existence, you need to form the product A^*A. If A has more columns than rows, this matrix is pretty big. The SVD of A can be computed more efficiently by computing the SVD of A^*, then using the following identity.

$$A^* = (U\Sigma V^*)^* = V\Sigma^* U^* \tag{5.20}$$

5.3 Application of SVD: PCA

Recall that the SVD can be used for dimensionality reduction as follows, for a matrix $A \in \mathbb{C}^{m\times n}$ of rank r.

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^* \tag{5.21}$$

$$\approx \sum_{i=1}^{r'} \sigma_i u_i v_i^* \tag{5.22}$$

By halting the sum early, at $r' < r$, we retain the r' biggest summands in a decomposition of A into rank 1 matrices.

Let $X \in \mathbb{C}^{n\times p}$ be a matrix of n points in p-space, collected as rows.[1] Each point can represent an observation, and each column a feature.

$$X = \begin{pmatrix} x_1{}^\top \\ x_2{}^\top \\ \vdots \\ x_m{}^\top \end{pmatrix} \tag{5.23}$$

Represent each point as a displacement from the sample average $\overline{x}$, and call this matrix of displacements $\widetilde{X}$.

$$\widetilde{X} = \begin{pmatrix} x_1{}^\top - \overline{x} \\ x_2{}^\top - \overline{x} \\ \vdots \\ x_m{}^\top - \overline{x} \end{pmatrix} \tag{5.24}$$

Use the SVD to write $\widetilde{X}$ as a sum of rank 1 matrices, from most important to least important. (Often $n \gg p$ and $\operatorname{rank} A = p$.)

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^* \tag{5.25}$$

Two data science interpretations of this SVD are the following:

- The vectors v_i are projections onto orthogonal directions of maximum variance, from greatest to least.
- Often times we observe that the singular values fall off a cliff or become negligible. In that even a low-rank approximation to is appropriate.

We will explore both of these in the next lecture.

[1] Usually x_i means column i, but using x_i to represent row i is more common for data science, possibly because the way we write math means that it's easier to picture a lot of rows than a lot of columns.