Recitation 10

October 10, 2024

1 Recap

1.1 The (compact) SVD

For an $m \times n$ matrix A of rank r, the "compact" SVD of A is the decomposition:

$$A = \underbrace{\begin{pmatrix} u_1 & \cdots & u_r \end{pmatrix}}_{\hat{U}} \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}}_{\hat{\Sigma}} \underbrace{\begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix}^T}_{\hat{V}^T} = \sum_{i=1}^{r} \sigma_i u_i v_i^T ,$$

where the u_i are an orthonormal basis for $C(A)$ called the **left singular vectors**, the $\sigma_i > 0$ are scale factors called the **singular values** (usually sorted $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$), and the v_i are an orthonormal basis for $C(A^T)$ called the **right singular vectors**.

We can't prove why this factorization exists (yet) or explain how to compute it, but we will focus for now on how to *use* the SVD. It is one of most powerful tools of applied linear algebra.

1.2 The Truncated SVD and low-rank approximation

If keep only the first $k \leq r$ biggest singular values, we obtain a **rank-k approximation** for A called a **truncated SVD**:

$$A \approx \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

The Eckart–Young theorem tells us that this is the *best* possible rank-k approximation for A, in that it minimizes the distance $\|A - B\|_F$ over all possible rank-k matrices B. Here, $\|M\|_F$ is the **Frobenius norm** of a matrix M:

$$\|M\|_F = \sqrt{\sum i, j M_{i,j}^2} ,$$

the most obvious (but not the only!) norm of a matrix: we just treat the entries of the matrix as entries of a vector and take the usual Euclidean norm.

1.3 New: The "full" SVD

The matrices $\hat{U}$ and $\hat{V}$ above have orthonormal columns, but in general they are not square (they are $m \times r$ and $n \times r$, respectively), so they are not invertible (not "orthogonal" / "unitary" matrices). This can be inconvenient.

There is an easy fix. The columns of $\hat{U}$ are an orthonormal basis for $C(A)$. What's missing is a basis for the orthogonal complement $N(A^T)$. Let $u_{r+1}, \ldots, u_m$ be *any* orthonormal basis for $N(A^T)$. These vectors are necessarily orthogonal to $u_1, \ldots, u_r \in C(A)$, so we can put them together into a *single square, invertible, orthogonal* $m \times m$ matrix:

$$U = \begin{pmatrix} u_1 & \cdots & u_m \end{pmatrix} .$$

Similarly let $v_{r+1}, \ldots, v_n$ be *any* orthonormal basis for $N(A)$. These vectors are necessarily orthogonal to $v_1, \ldots, v_r \in C(A^T)$, so we can put them together into a *single square, invertible, orthogonal* $n \times n$ matrix:

$$V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} .$$

To put those extra rows/columns back into the SVD, we simply need to expand $\hat{\Sigma}$ with extra rows/columns to multiply the nullspace vectors by *zero*. This yields the "full" SVD:

$$A = \underbrace{\begin{pmatrix} u_1 & \cdots & u_m \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & 0 & \cdots \\ & & 0 & 0 & \cdots \\ & & \vdots & \vdots & \ddots \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^T}_{V^T},$$

where Σ is $m \times n$ ($n - r$ extra columns and $m - r$ extra rows of zeros compared to $\hat{\Sigma}$).

2 Exercises

1. How do the singular values and vectors of A^T relate to those of A?
2. If A is a square, invertible matrix, directly show that $A^{-1} = A^+$, or equivalently that $AA^+ = A^+A = I$.
3. Suppose Q is an $m \times n$ matrix with orthonormal columns ($m \geq n$). What is the compact SVD of Q?
4. Suppose that $A = xy^T$ is a rank-1 $m \times n$matrix for nonzero vectors $x \in \mathbb{R}^m$ and $y =\in \mathbb{R}^n$. What is the compact SVD of A?

3 Solutions

1. $A^T = \hat{V}\hat{\Sigma}^T\hat{U}^T = \hat{V}\hat{\Sigma}\hat{U}^T$ (since $\hat{\Sigma}^T = \hat{\Sigma}$: it is a diagonal matrix). Hence the singular values are the same, and the singular vectors swap: the *left* singular vectors of A^T are the *right* singular vectors $\hat{V}$ of A, and vice versa!

2. $m = n = r$, so both $\hat{U}$ and $\hat{V}$ are square, orthogonal (inverse = transpose) matrices. In fact, the compact SVD is equivalent to the full SVD in this case, so we can drop the "hats" over the letters. Hence, $A^{-1} = (U\Sigma V^T)^{-1} = V\Sigma^{-1}U^T = A^+$ (since $(V^T)^{-1} = V$, $U^{-1} = U^T$, and $\Sigma = \hat{\Sigma}$ is diagonal and hence trivially invertible). Or, equivalently, $AA^+ = U\Sigma V^T V\Sigma^{-1}U^T = U\Sigma\Sigma^{-1}U^T = UU^T = I$, and similarly for A^+A. Note that this all crucially relies on *all* the matrices being square!

3. Trivially, $Q = QI_nI_n$, where I_n is the $n \times n$ identity matrix. But this is an SVD! Just let $\hat{U} = Q$, $\hat{\Sigma} = I_n$, and $\hat{V} = I_n = I_n^T$. So, the left singular vectors are the columns of Q, the singular values are all **equal to one**, and the right singular vectors are the columns of I_n!

4. We need an orthonormal basis for the row and columns spaces, which we can get just by normalizing x and y. That is, let $u_1 = x/\|x\|$, let $v_1 = y/\|y\|$, and then we immediately get that $A = u_1\sigma_1 v_1^T$ where $\sigma_1 = \|x\|\,\|y\|$, which is the compact SVD for $r = 1$.