5374 Fall '22

Numerical Linear Algebra Lecture 24

Today * More on the SVD

(low rank approximation, compression, dimensionality reduction)

Remarks on the SVD (continued)

4. The sup norm of A and the SVD

Let A be $m\times n$, recall

$$\|A\| := \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

Claim: If $A = U\Sigma V^t$, then

$$\|A\| = \sigma_1$$

Proof. Observation: If

$$\Sigma = \begin{array}{c} \quad n \\ \left(\begin{array}{cccccc} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & & & & \\ \vdots & & \ddots & & & \\ & & \cdots & \sigma_r & 0 \cdots & \\ 0 & & & & & \end{array} \right) \end{array} \Bigg| \, m$$

$(r \leq \min(m,n))$

Then given $x \in \mathbb{R}^n$,

$$\|\Sigma x\|_2 = \left\| \begin{pmatrix} \sigma_1 x_1 \\ \sigma_2 x_2 \\ \vdots \\ \sigma_n x_n \end{pmatrix} \right\|_2$$

$$= \sqrt{(\sigma_1 x_1)^2 + \cdots + (\sigma_n x_n)^2}$$

$$= \sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_n^2 x_n^2}, \quad \text{since } \sigma_1^2 \geq \sigma_k^2 \ \forall k$$

$$\leq \sqrt{\sigma_1^2 (x_1^2 + \cdots x_n^2)}$$

$$= \sigma_1 \|x\|_2$$

$$\Rightarrow \quad \frac{\|\Sigma x\|_2}{\|x\|_2} \leq \sigma_1 \qquad \forall \ x \neq 0$$

Moreover, $\|\Sigma e_1\|_2 = \left\| \begin{pmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2 = \sigma_1$

Thus $$\max_{x \neq 0} \frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_1$$

The general claim follows from here as follows (we are going to use that if U, V are orthogonal matrices then $\|Uy\|_2 = \|y\|_2$ and $\|Vx\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$)

Well,

$$A v_1 = (U \Sigma V^t) v_1, \qquad V^t v_1 = e_1$$

$$= U \Sigma e_1 = \sigma_1 U e_1 = \sigma_1 u_1$$

$$\Longrightarrow \quad \|Av_1\|_2 = \|\sigma_1 u_1\|_2 = \sigma_1 \|u_1\|_2 = \sigma_1$$

$$\Longrightarrow (\|v_1\|_2 = 1) \qquad \frac{\|Av_1\|_2}{\|v_1\|_2} = \sigma_1$$

On the other hand, take $x \in \mathbb{R}^n$ $(x \neq 0)$

$$\frac{\|Ax\|_2}{\|x\|_2} = \frac{\|(U\Sigma)(V^t x)\|_2}{\|x\|_2}$$

$$\leq \frac{\|U\Sigma\| \; \|V^t x\|_2}{\|x\|_2} \qquad (\|V^t x\|_2 = \|x\|_2$$

$$\leq \|U\Sigma\|$$

Next, let again $x \in \mathbb{R}^n$ $(x \neq 0)$

applied it um

$(y = \Sigma x)$

$$\frac{\|U(\Sigma x)\|_2}{\|x\|_2} = \frac{\|\Sigma x\|_2}{\|x\|_2} \qquad (\text{since } \|Uy\|_2 = \|y\|_2$$

$$\leq \sigma_1$$

$$\Longrightarrow \quad \|U\Sigma\| \leq \sigma_1$$

Thus $\quad \|A\| = \sigma_1$

5. Observe that if

$$A = U \Sigma V^t \qquad (\ A \text{ is } n \times n, \text{ invertible}\ )$$

Then

$$\begin{aligned} A^{-1} &= (U \Sigma V^t)^{-1} \\ &= (V^t)^{-1} \Sigma^{-1} U^{-1} \\ &= V \Sigma^{-1} U^t \end{aligned}$$

$$\Longrightarrow \quad \|A^{-1}\| = \frac{1}{\sigma_n}$$

In particular, if the singular values of A are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, then

$$\operatorname{cond}(A) = \frac{\sigma_1}{\sigma_n}$$

6. If A is symmetric, then the singular values of A are the absolute values of the eigenvalues of A.

7. Other things that follow from the SVD is the determinant (up to a sign) because (A is $n \times n$)

$$\det A = \underbrace{(\det U)}_{(\pm 1)} (\det \Sigma) \underbrace{(\det V^t)}_{(\pm 1)}$$

$$\Rightarrow \quad \det A = \pm \prod_{k=1}^{n} \sigma_k$$

$$|\det A| = \prod_{k=1}^{n} \sigma_k$$

8. Let A have exactly r non-zero singular values, i.e. the singular values of A are s.t.

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_r > 0$$

and either $r = \min(m, n)$ or $\sigma_{r+1} = 0$

The number r is exactly the rank of A.

Low Rank Approximation

An alternative way of writing the SVD is, if r is the rank of A, then

$(\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0)$

$$A = \sigma_1 u_1 \otimes v_1 + \sigma_2 u_2 \otimes v_2 + \dots + \sigma_r u_r \otimes v_r$$

(Recall : $(x \otimes y) z = (z, y) x$)

where $u_1, \dots, u_m$ and $v_1, \dots, v_n$ represent the column vectors of U and V.

Observe: $\mathrm{Im}(A) = \mathrm{span}\{u_1, \dots, u_r\}$

$\mathrm{Ker}(A) = \mathrm{span}\{v_{r+1}, \dots, v_n\}$

Every matrix of rank-1 can be written in the form

$$x \otimes y$$

where x, y are non-zero vectors. If $y \in \mathbb{R}^n$, $x \in \mathbb{R}^m$ the matrix will be $m \times n$.

Another way of stating the SVD is that any matrix of rank r is equal to the sum of exactly r rank-1 matrices generated by orthogonal vectors:

$$A = \sigma_1 u_1 \otimes v_1 + \cdots + \sigma_r u_r \otimes v_r$$

Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, the first terms tend to have more influence than the latter ones

Picture a matrix A of the form

$$A = 1000\, u_1 \otimes v_1 + 1000\, u_2 \otimes v_2 + 100\, u_3 \otimes v_3$$
$$+ (0.0001)\, u_4 \otimes v_4 + \cdots + (0.0001)\, u_{1000} \otimes v_{1000}$$

This would be a matrix of rank 1000 where the smallest 997 singular values are very small, and thus, in certain contexts, it might make sense to approximate A with

$$A_3 := 1000\, u_1 \otimes v_1 + 1000\, u_2 \otimes v_2 + 1000\, u_3 \otimes v_3$$

which is a rank-3 matrix.

There is a Theorem justifying how this provides the best possible low rank approximation

Theorem: Let A be a $m \times n$ matrix with SVD given by:

$$A = \sigma_1 u_1 \otimes v_1 + \cdots + \sigma_r u_r \otimes v_r \quad (\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0)$$

For every $\ell = 1, 2, \ldots, r$ define

$$A_\ell = \sigma_1 u_1 \otimes v_1 + \cdots + \sigma_\ell u_\ell \otimes v_\ell$$

Then A_ℓ is the best ℓ-rank approximation to A as measured in the $\| \; \|_{Fr}$ and $\| \; \|_{op}$ norm of A; that is

$$\left(\|A\|_{op} = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \right)$$

$$\|A - A_\ell\|_{Fr} = \min \{ \|A - B\|_{Fr} \mid \operatorname{rank}(B) = \ell \}$$

$$\|A - A_\ell\|_{op} = \min \{ \|A - B\|_{op} \mid \operatorname{rank}(B) = \ell \}.$$