Preview of Tensors:
Revisiting the SVD

If $A = U \Sigma V^T = \left[\vec{u}_1 \middle| \cdots \middle| \vec{u}_m \right] \Sigma \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix}$

then $\operatorname{rank}(A) = r = \#$ of nonzero singular values $\sigma_1, \dots, \sigma_n$

Pf: <u>Lemma</u>: If B invertible then

$$\operatorname{rank}(AB) = \operatorname{rank}(A) = \operatorname{rank}(BA)$$

<u>Pf</u>:

$\operatorname{rank}(AB) \le \operatorname{rank}(A)$ by HW

but: $\operatorname{rank}(A) = \operatorname{rank}((AB)B^{-1}) \le \operatorname{rank}(AB)$

↪ $\operatorname{rank}(A) = \operatorname{rank}(AB)$ ✓

↪ likewise for BA.

So: $\operatorname{rank}(A) = \operatorname{rank}(U \Sigma V^T)$

$= \operatorname{rank}(\Sigma)$ since U, V invertible

$= r$ (why?)

Recall: can write

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T = \sum_{i=1}^{r} \sigma_i \vec{u}_r \vec{v}_r^T$$

This shows: rank(A) = min # of rank 1 matrices it takes to sum to A

Why: ↳ ofc, can't write A as less than $r = \text{rank}(A)$ rank 1 matrices

Since by HW: $\text{rank}(A_1 + \cdots + A_k)$
$\leq \text{rank}(A_1) + \cdots + \text{rank}(A_k)$
$\leq k$ if all $\text{rank}(A_i) = 1$.

↳ above shows: we can achieve A as sum of ~~...~~ r rank 1 matrices

↳ this observation will allow us to generalize concept of rank to tensors.

↳ shows: if we defined rank(A) as least # of rank 1 matrices needed to sum to A — then this would be equiv to orig. def'n (# of indep. columns)

If we take initial segments:

$$A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^{\,T} \qquad (\text{for } k \le r)$$

Then: ① A_k is rank k (why? — we knew SVD of A_k! — why?)

② A_k is "best rank k approx'n to A" in a sense we will now make precise:

<u>Define</u>: L^2 norm of A is defined by:

$$\|A\|_2 = \min_{\|\vec{x}\|=1} \|A\vec{x}\| = \min_{\vec{x}\in\mathbb{R}^n} \frac{\|A\vec{x}\|}{\|\vec{x}\|}$$

$= \sigma_1$, first singular value of A.

↪ sometimes write $\|A\|$ for $\|A\|_2$.

↪ can show: $\|\cdot\|_2$ is norm on the space $\mathbb{R}^{m\times n}$:

$\|A\| \ge 0$, $\|cA\| = |c|\,\|A\|$, $\|A+B\| \le \|A\|+\|B\|$, $\|AB\| \le \|A\|\,\|B\|$

$A, B \in \mathbb{R}^{m \times n}$

Thm (Eckart-Young): if B is rank k $(k \leq r = \text{rank}(A))$ then:

$$\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}$$

why?

i.e. A is closest (in L^2 norm) to A_k among all rank k matrices.

Pf: - Sps not, i.e. sps there is B w/ $\text{rank}(B) = k$ s.t. $\|A - B\|_2 < \|A - A_k\|$
$= \sigma_{k+1}$

- By def'n of $\|\cdot\|_2$, for any $\vec{w}$ we have

$$\frac{\|(A-B)\vec{w}\|_2}{\|w\|} < \sigma_{k+1} \Rightarrow \|(A-B)\| < \sigma_{k+1}\|\vec{w}\|$$

- By rank-nullity: nullspace $N(B)$ of B is rank $n-k$.

- For any $\vec{w} \in N(B)$ we have $(A-B)\vec{w} = A\vec{w}$, so:

$$\|(A-B)\vec{w}\| = \|A\vec{w}\| < \sigma_{k+1}\|\vec{w}\|$$

- So $N(B)$ is an $(n-k)$-dimensional subspace of $\mathbb{R}^n$ s.t. $\forall \vec{w} \in N(B)$

$$\|A\vec{w}\| < \sigma_{k+1} \|\vec{w}\|$$

- Now consider the subspace V spanned by $\vec{v}_1, \dots, \vec{v}_{k+1}$ → the first $k+1$ singular vectors of A.

- This space is $(k+1)$-dim'l and why? for every $\vec{w} \in V$ we have

$$\|A\vec{w}\| \geq \sigma_{k+1} \|\vec{w}\|$$

Why? because we can write every such $\vec{w}$ as a linear combo of $\vec{v}_1, \dots, \vec{v}_{k+1}$ and these vectors are all stretched at least by a factor of σ_{k+1}! (since $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{k+1}$)

- Now: since $\dim(N(B)) + \overbrace{\dim(V)}^{=n+1} > n$ there must be some vector $\vec{u}$ in their intersection

- But then

$\|A\vec{w}\| < \sigma_{k+1}\|\vec{w}\|$ since $\vec{w} \in N(B)$

and $\|A\vec{w}\| \geq \sigma_{k+1}\|\vec{w}\|$ since $\vec{w} \in V$

a contradiction ✓

<u>Summary</u> SVD gives a way of writing A as sum of r ($=\operatorname{rank}(A)$) many rank 1 matrices:

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$$

Given $k \leq r$ the initial sum $A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T$ is rank matrix that best approximates A (wrt L^2 norm) among all rank k matrices

↳ what about wrt <u>other</u> norms?

Define the Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$ defined by:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

e.g. $\left\| \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \right\|_F = \sqrt{1^2 + 0^2 + (-1)^2 + 2^2}$
$= \sqrt{6}$

F norm treats A "like a $m \times n$ length vector"

Can check $\|\cdot\|_F$ also satisfies norm properties:

$\|A\|_F \geq 0$, $\|cA\|_F = |c| \, \|A\|_F$,
$\|A+B\|_F \leq \|A\|_F + \|B\|_F$, $\|AB\|_F \leq \|A\|_F \|B\|_F$

- Notice: if Q orthogonal, then since $\|Q\vec{x}\|_2 = \|\vec{x}\|$ for all vectors, and $\|A\|_F$ sums the norms of columns of A and QA multiplies each column by Q we must have $\|QA\|_F = \|A\|_F$

This also gives us a way to write $\|A\|_F$ in terms of A's singular values: if $A = U\Sigma V^T$

$$\|A\|_F = \|U\Sigma V^T\|_F = \|\Sigma\|_F$$

orthog

$$= \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

Amazingly, $A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T$ is also nearest rank k matrix to A <u>in the Frobenius norm</u>:

<u>Theorem</u> (Eckart - Young for $\|\cdot\|_F$)

If B is rank k ($k \leq r = \mathrm{rank}(A)$) then

$$\|A - B\|_F \geq \|A - A_k\|_F$$

$$= \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

<u>Pf: omitted.</u>