Mathematics for AI

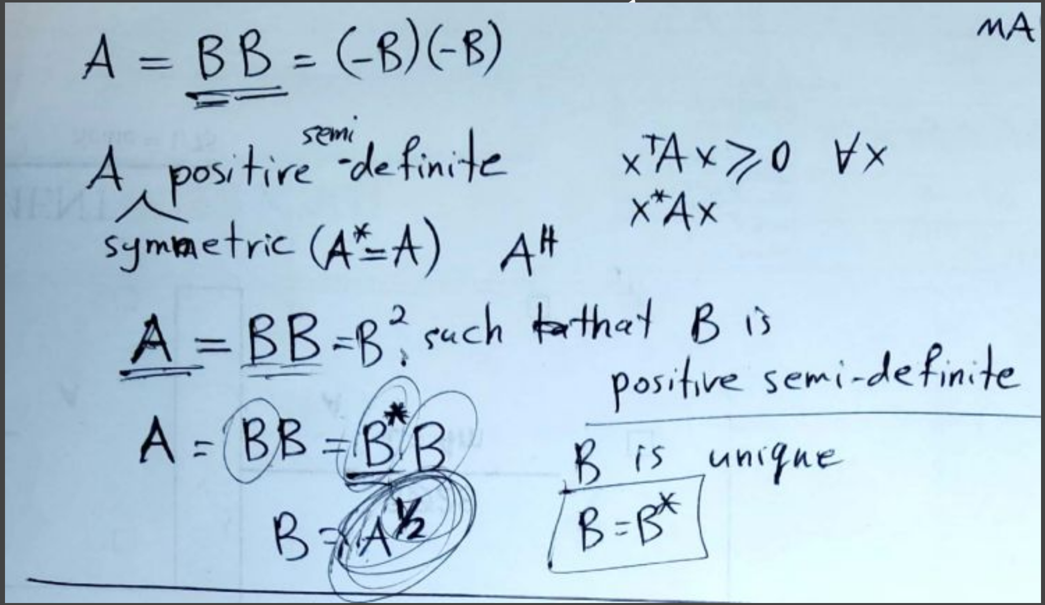
Lecture 13
SVD, Matrix Norm, Low-rank approximation

Remember Square root of positive semi-definite matrices

K. N. Toosi
University of Technology

For a (symmetric) positive semi-definite matrix A there is a unique positive semi-definite matrix P such that A = P P (= P^H P). P is

called the square root of



Remember: SVD

K. N. Toosi
University of Technology

$$\underset{m\times n}{A} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T}$$

$$U^T U = U U^T = I$$

$$V^T V = V V^T = I$$

Σ diagonal

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \end{bmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$$

$$U = [u_1\ u_2 - u_n] \quad V = [v_1\ v_2 \cdots v_n] \quad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots\ \sigma_i \end{bmatrix}$$

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i \boxed{\underbrace{\vec{u}_i \cdot \vec{v}_i^{\,T}}_{m\times n}}$$

SVD and Eigen-decomposition

MA13

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

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

$$= V \Sigma^T \Sigma V^T = V \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$$

$$A^T A = V \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} V^T$$

$$V^T = V^{-1} \implies (A^T A) V = V \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{bmatrix}$$

$$(A^T A) \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 v_1 & \sigma_2^2 v_2 & \sigma_3^3 v_3 \end{bmatrix}$$

$$(A^T A) v_i = \sigma_i^2 v_i$$

(σ_i, v_i) are eigenpairs of $A^T A$

Remember: SVD, rank, column-, row-, and null spaces

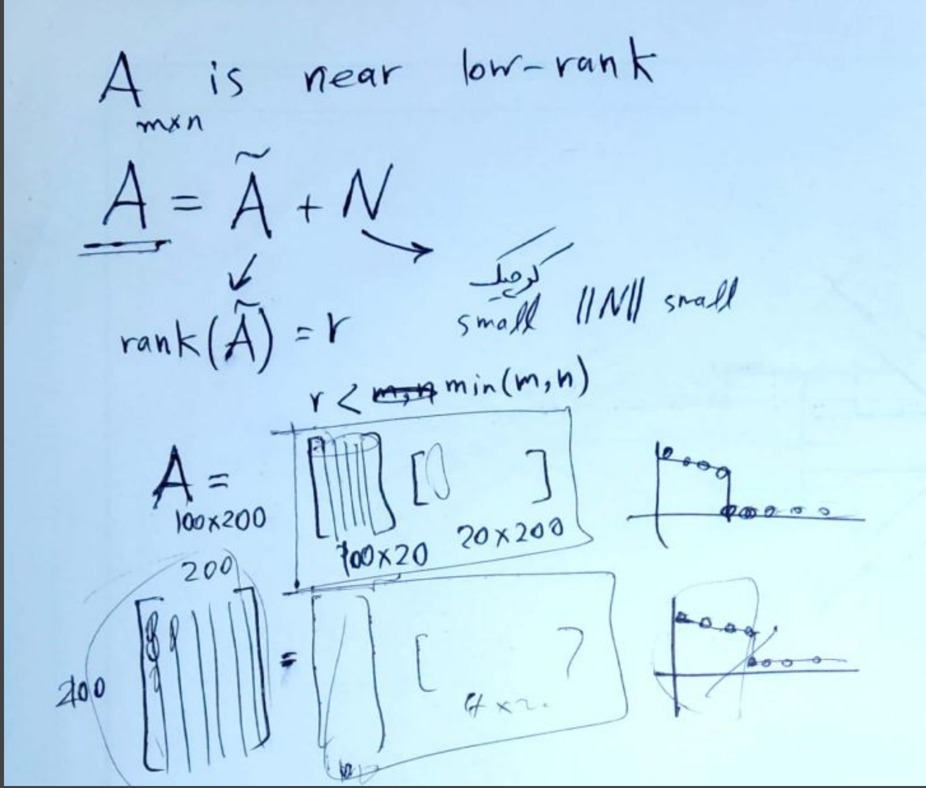
$$\underset{m\times n}{A} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T} = \Big[\underbrace{u_1\, u_2 \cdots u_r}_{\substack{\text{orthonormal}\\ \text{basis}\\ C(A)}}\ \underbrace{u_{r+1} \cdots u_m}_{N(A^T)}\Big] \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$\mathrm{rank}(A) = r \leq \min(m, n)$$

basis $R(A) = C(A^T)$: $v_1^T, v_2^T, \ldots, v_r^T$

$N(A)$: $v_{r+1}^T, \ldots, v_n^T$

Near-low-rank matrices



K. N. Toosi
University of Technology

Remember: vector norm

~~mat~~ # vector norm $\quad \|\cdot\| \quad V \longrightarrow \mathbb{R}^+$

$[0, \infty)$

$\vec{v} \in V$

$$\|\vec{v}\| \geq 0$$

$$\|\vec{v}\| = \vec{0} \iff \vec{v} = \vec{0}$$

$$\|\alpha \vec{v}\| = |\alpha| \, \|\vec{v}\|$$

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|v\| \quad \text{triangle inequality}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

$$\|x\| = \|x\|_2 \quad L^2\text{-norm}$$

$$= \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$$

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \dots + |x_n|^p\right)^{1/p} \quad \underline{\text{p-norm}}$$

$$\|x\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$$

K. N. Toosi
University of Technology

Matrix Norm

K. N. Toosi
University of Technology

$A \in \mathbb{R}^{m\times n}$ $\quad \|\cdot\| \quad \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^{+}$ MA13 V

matrix norm

$$\|A\| \geq 0$$

$$\|A\| = 0 \iff A = 0_{m\times n} \in \mathbb{R}^{m\times n}$$

$$\|\alpha A\| = |\alpha| \, \|A\|$$

$$\|A+B\| \leq \|A\| + \|B\|$$

$$\|AB\| \leq \|A\| \, \|B\| \longrightarrow$$ sub-multiplicative norm

Elementwise norms, Forbenius norm

K. N. Toosi
University of Technology

element-wise norms

$$\|A\|_F = \sqrt{\overbrace{\sum_{i=1}^{m}\sum_{j=1}^{n} (A_{ij})^2}^{\text{vect}(A)}}$$

$$= \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2}$$

$\sigma_i \rightarrow$ singular values

Frobenius Norm

np.linalg.norm(A,'fro')

Operator norm

K. N. Toosi
University of Technology

operator norm

$$\max_{x} \frac{\|Ax\|_{norm1}}{\|x\|_{norm2}} = \max \{ \|Ax\| \mid \|x\| = 1 \}$$

Spectral Norm

K. N. Toosi
University of Technology

$$\max_x \frac{\|Ax\|_2}{\|x\|_2}$$

spectral norm $= \sigma_{max}$

largest singular value

$\|A\|$

np.linalg.norm(A)

$L_{p,q}$ Norms

$$\|A\|_{p,q} = \left\| \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \right\|_{p,q} = \left\| \begin{bmatrix} \|a_1\|_p \\ \|a_2\|_p \\ \vdots \\ \|a_n\|_p \end{bmatrix} \right\|_q$$

Nuclear norm (trace norm)

$$\|A\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i = \left\| \begin{bmatrix}\sigma_1\\ \sigma_2\\ \vdots\\ \sigma\end{bmatrix} \right\|_1$$

nuclear norm

trace norm

$$= \mathrm{trace}\left[(A^T A)^{\frac{1}{2}}\right]$$

Matrix norm

Look at the wikipedia article

https://en.wikipedia.org/wiki/Matrix_norm

Low-rank approximation

K. N. Toosi
University of Technology

find closest matrix to $A \in \mathbb{R}^{m\times n}$ whose rank is r.

$$\hat{A} = \underset{B\in\mathbb{R}^{m\times n}}{\operatorname{argmin}} \|A - \cancel{A}B\| \quad \text{subject to} \quad \boxed{\operatorname{rank}(B)=r}$$

low-rank estimation

Eckart–Young–Mirsky theorem

Eckart Young Mirsky

$\|\ \|_F \qquad \|\ \|_2$ spect-

$$A = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{\min(m,n)} \end{bmatrix} V^T$$

$$\hat{A} = \underset{m\times m}{U} \underset{m\times n}{\begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & 0 & \\ & & & & & & 0 \end{bmatrix}} \underset{n\times n}{V^T}$$

SVD-thresholding