Math 225 lecture 18 Oct 18^{th} 2023 ①

section 7.4

Goal: Introduce students to the ideas behind the singular value decomposition (SVD). define the the pieces for it and give a rough idea of its meaning.

Class Q we can interpret the σ_i's as the lengths of which vectors?

Recall: some matrices are orthogonally diagonalizable (the symmetric ones) i.e. $A = QDQ^T$, some are diagonalizable (geometric & algebraic mults of the e. vals match) i.e. $A = PDP^{-1}$. what do we do if A is not square, or simply not diagonalizable?

let A be an $m \times n$ matrix A^T is $n \times m$

A^TA is a symmetric $n \times n$ matrix that can be formed by any A.

↳ we call the square roots of the eigenvalues of A^TA the "singular values" of A

[aside: these eigenvalues are always positive real numbers so their square roots are valid. for real see our previous notes. for positive see pg 590, its based on examining $(A\vec{v})\cdot(A\vec{v})$ for $\vec{v}$ an eigen value of A]

we can take all of these square roots & place them in order

this is the last non-zero singular value $r = \text{rank}(A)$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0 \dots \geq 0 \qquad \sigma_i = \sqrt{\lambda_i}$$

for λ_i e. val of A^TA

Hilroy

Ex: $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ $A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

$\Rightarrow A^TA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix}$ evals of A^TA are $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = 1$

$\Rightarrow \sigma_1 = \sqrt{9} = 3 \quad \sigma_2 = \sigma_3 = \sqrt{1} = 1$

$3 \geq 1 \geq 1$

Ex: $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ $A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

$A^TA = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ now we find eigen values.

$$\det(A^TA - \lambda I) = \left| \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \right| = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 1)(\lambda - 3)$$

$\therefore$ evals of A^TA are $\lambda_1 = 3$ & $\lambda_2 = 1$

making the singular values of A $\sigma_1 = \sqrt{3}$ $\sigma_2 = \sqrt{1} = 1$

Question, how do we interpet these numbers? What is their significance?

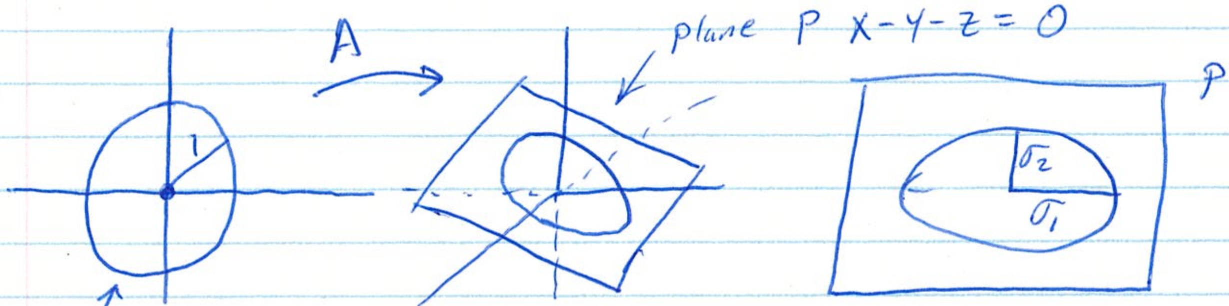
Answer (one potential one) geometry

$$A = \begin{bmatrix}1 & 1\\ 1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix} = \begin{bmatrix}a+b\\ a\\ b\end{bmatrix}$$

this is a transformation (where?)

$$A: \mathbb{R}^2 \to \mathbb{R}^3$$

better pictures on pg. 591



unit circle in $\mathbb{R}^2$

note $A\vec{v}$ always lies in the plane P as that that is $Col(A) \subset \mathbb{R}^3$.

This leads to the goal of today, the Singular value decomposition (SVD)

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

A is the original $m \times n$ matrix

U is an $m \times m$ orthogonal matrix

Σ is $m \times n$ and "diagonal"

V^T is $n \times n$ & orthogonal

lets look at the 3 pieces individually.

(1) Σ is built from singular values by listing the non-zero ones on the "diagonal"

i.e. $\sigma_1 = 3 \quad \sigma_2 = 1$

Σ 2×2 $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ 2×3 $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ 3×3 $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(note: the situation where you don't have the room does not arrize. since there are $r = \text{rank}(A)$ many non zero values & row rank(A) = rank(A^TA) = col rank(A) $\therefore r \leq m$ & $r \leq n$ & will always fit)

(2) V these are the "right singular vectors"

this is an $n \times n$ matrix but its just an orthogonal basis of e. vects for A^TA (i.e. essentially the Q for A^TA)

i.e. $V = [\vec{v}_1 \dots \vec{v}_n]$ with $A^TA\vec{v}_i = \lambda_i v_i$ & $\{V_i\}$ orthonormal.

(3) U these are called the "left singular vectors" & have some complications

first note that $A\vec{v}_1, A\vec{v}_2, \dots A\vec{v}_n$ are orthogonal

Proof $A\vec{v}_i \cdot A\vec{v}_j = \vec{v}_i^{\,T} A^T A \vec{v}_j = \vec{v}_i^{\,T} \lambda_j \vec{v}_j$
$= \lambda_j \vec{v}_i^{\,T} v_j$
$= \lambda_j (\vec{v}_i \cdot \vec{v}_j) = 0$

by discussion on orthogonal diagonalization of symmetric matricies.

we then define U in terms of these vectors.

$u_1 = \frac{1}{\sigma_1} A\vec{v}_1$, $u_2 = \frac{1}{\sigma_2} A\vec{v}_2$ ... $u_r = \frac{1}{\sigma_r} A\vec{v}_r$

for each non-zero σ_i

☆ Alarm bells! ☆

if $r=m$? great! $U = \left[\frac{u_1}{\|u_1\|} \; \cdots \; \frac{u_r}{\|u_r\|} \right]$

however $r \leq m$ what do we do if $r < m$?
we complete $u_1, \ldots, u_r$ to a basis of $\mathbb{R}^m$
(more on that friday)