Lecture 4

Spectral Theorem

$A \in S^n \qquad A = U \Lambda U^T$

U : orthonormal (unitary)

Λ : diagonal

$$A = \Big[\ \underbrace{u_1 \quad u_2 \cdots u_r}_{\text{range}} \ \overbrace{u_{n-1} \quad u_n}^{\text{null}}\ \Big] \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ & & & & 0 \end{bmatrix}$$

Variational Characteristics of eigenvalue symmetric matri

$$R = \frac{\vec{x}^T A \vec{x}}{\underbrace{\vec{x}^T \vec{x}}_{\text{scalar - Raleigh coeff}}}, \quad A \in S$$

↳ quadratic form

$$\lambda_{min}(A) \le \frac{\vec{x}^T A \vec{x}}{\|\vec{x}\|_2^2} \le \lambda_{max}(A)$$

$$\lambda_{max}(A) = \max_{\|\vec{x}\|_2 = 1} \vec{x}^T A \vec{x}$$

↳ use spectral thm to write eqn for $\lambda_{min}(A)$

$A = U^T \Lambda U$

$\lambda_{max}(A) = \qquad \vec{x}^T A \vec{x}$

$$\vec{x}^T A \vec{x} = \vec{x}^T U^T \Lambda U \vec{x}$$

unit vector

orthonormal matrix $\Rightarrow$ preserves norm

$U\vec{x} = \vec{y}$

$\vec{y}$ is a unit vector

$$= \vec{y}^T \Lambda \vec{y}$$

dot prod

$$= \underbrace{\sum_{i=1}^{n} \lambda_i y_i^2}_{\text{want to bound this quantity}}$$

$\forall i,\ \lambda_i \leq \lambda_{max}$

$\|\vec{x}\|_2^2 = 1$

$\Rightarrow \|\vec{y}\|_2^2 = 1$

$$= \sum_{i=1}^{n} \lambda_i y_i^2 \leq \lambda_{max} \sum_{i=1}^{n} y_i^2 = \lambda_{max} \|\vec{y}\|_2^2$$

$$= \lambda_{max} \|\vec{x}\|_2^2$$

Q/ What $\vec{x}$ achieves the max?

↳ eigenvector that corresponds to λ_{max}

$\Rightarrow$ choose this eigenvector

PCA

- have p-dim data
  - ↳ want q-dim structure $\quad q < p$

$\vec{x}_1, \ldots, \vec{x}_n \leftarrow p$-dim $(\vec{x}_i \in \mathbb{R}^p)$

$$X = \begin{bmatrix} — & \vec{x}_1^T & — \\ — & \vec{x}_2^T & — \\ & \vdots & \\ — & \vec{x}_n^T & — \end{bmatrix}$$

$n \times p$

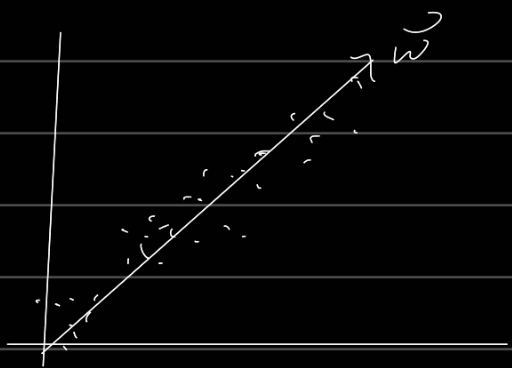
"Data matrix"

$$X^T = \begin{bmatrix} | & & | \\ \vec{x}_1 & \cdots & \vec{x}_n \\ | & & | \end{bmatrix}$$

$$C = \frac{1}{n} X X^T = \frac{1}{n} \underbrace{\begin{bmatrix} \|\vec{x}_1\|^2 & \langle \vec{x}_1, \vec{x}_2 \rangle & \cdots \\ \vdots & \|\vec{x}_2\|^2 & \\ \vdots & & \ddots \\ & & & \|\vec{x}_n\|^2 \end{bmatrix}}_{\text{Symmetric}}$$

$$D = \frac{1}{n} X^T X \leftarrow \text{also symmetric}$$

Want to find : $\vec{w}$



given $\vec{x}_1, \dots, \vec{x}_n$ want one-dim $\vec{w}$, $\|\vec{w}\| = 1$
s.t. projected vectors are as close as possible to o.g. vectors $\rightarrow \min_{\vec{w}} \frac{1}{n} \sum e_i^2$

proj: $\langle \vec{x}_i, \vec{w} \rangle \vec{w}$ $\rightarrow$ want to minimize error of projection

Error: $\| \vec{x}_i - \langle \vec{w}, \vec{x}_i \rangle \vec{w} \|^2 = e_i^2$

Avg error: $\frac{1}{n} \sum_{i=1}^{n} e_i^2$

$$\| \vec{x}_i - \langle \vec{w}, \vec{x}_i \rangle \vec{w} \|^2 = (\vec{x}_i - \langle \vec{w}, \vec{x}_i \rangle \vec{w})^T (\vec{x}_i - \langle \vec{w}, \vec{x}_i \rangle \vec{w})$$

$$= \|\vec{x}_i\|_2^2 + \langle \vec{w}, \vec{x}_i \rangle^2 \cancel{\|\vec{w}\|_2^2}^{1} - 2 \langle \vec{w}, \vec{x}_i \rangle^2$$

$$= \|\vec{x}_i\|_2^2 - \langle \vec{w}, \vec{x}_i \rangle$$

Avg error: $\frac{1}{n} \sum_{i=1}^{n} \|\vec{x}_i\|_2^2 - \frac{1}{n} \sum_{i=1}^{n} \langle \vec{w}, \vec{x}_i \rangle^2$

minimize over $\vec{w}$
so the 1st term is neglible

$$\min_{\vec{w}} \frac{1}{n} \sum_{i=1}^{n} \|\vec{x}_i\|_2^2 - \frac{1}{n} \sum_{i=1}^{n} \langle \vec{w}, \vec{x}_i \rangle^2$$

$$= \min_{\vec{w}} -\frac{1}{n} \sum_{i=1}^{n} \langle \vec{w}, \vec{x}_i \rangle^2$$

$$= \max_{\vec{w}} \frac{1}{n} \sum_{i=1}^{n} \langle \vec{w}, \vec{x}_i \rangle^2$$

$$= \max_{\vec{w}} \frac{1}{n} \| X \cdot \vec{w} \|^2 \qquad X = \begin{bmatrix} - & \vec{x}_1^T & - \\ & \vdots & \\ - & \vec{x}_n^T & - \end{bmatrix}$$

$$= \max_{\vec{w}} \frac{1}{n} (X \vec{w})^T (X \vec{w})$$

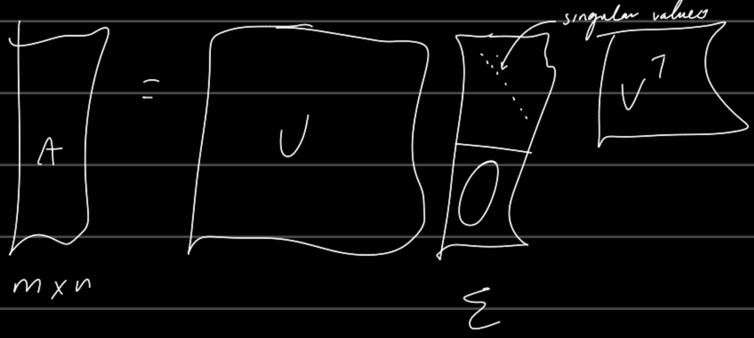
$$= \max_{\vec{\omega}} \vec{\omega}^T \underbrace{X^T X}_{} \vec{\omega}$$

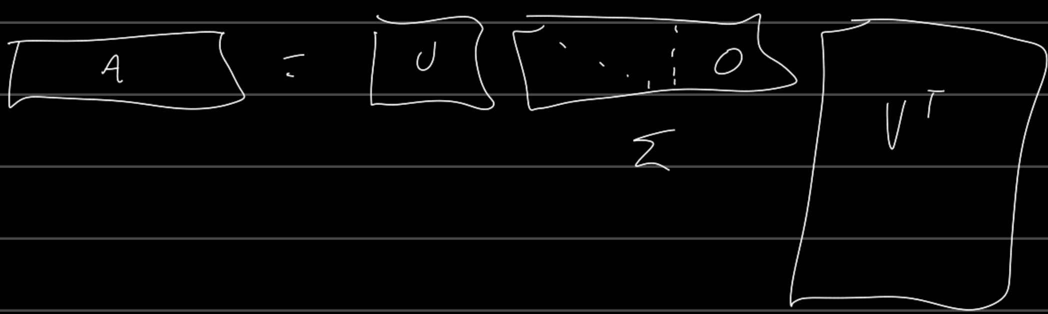
1st PC is eigenvector corr. to λ_{max} of $X^T X$

SVD

SVD of a matrix $A \in \mathbb{R}^{m \times n}$ is given by

$$\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times n}{\Sigma}\,\underset{n\times n}{V^T} \quad \leftarrow \text{full SVD}$$





→ can be written in dyadic form:

$$A = \underbrace{\sigma_1 \vec{u}_1 \vec{v}_1^T}_{\text{dyad}} + \sigma_2 \vec{u}_r \vec{v}_2^T + \dots + \sigma_r \vec{u}_r \vec{v}_r^T$$

$\text{Rank}(A) = r$

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

Consider $A^T A$ (symmetric)

$$\hookrightarrow \underset{\vec{v}_1}{\lambda_1} \geq \underset{\vec{v}_2}{\lambda_2} \cdots \geq \underset{\vec{v}_r}{\lambda_r} > 0$$

$$\lambda_{r+1} \cdots \lambda_n = 0$$

$$A^T A \vec{v}_i = \lambda_i \vec{v}_i$$

$$V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$$

Define : $\sigma_i = \sqrt{\lambda_i}$, $A\vec{v}_i = \sigma_i \vec{u}_i$

Turns out : $\vec{u}_i, \vec{v}_j$ orthonormal