# Linear Algebra for Computer Science

## Lecture 29

## Low-rank Approximation
## Principal Component Analysis

# Low-rank matrices

K. N. Toosi
University of Technology

$A$ $\qquad \text{rank}(A) = r < \min(m,n)$ $\qquad$ LA29 I

$$A = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} V^T$$

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \in \mathbb{R}^{200 \times n}$$

# Near-low-rank matrices

VIII

$A \in \mathbb{R}^{m \times n}$ is of rank $r < \min(m,n)$

what we have is

$$\tilde{A} = A + N \rightarrow \text{small noise matrix}$$

$$\text{rank}(\tilde{A}) = \min(m,n)$$

$A \in \mathbb{R}^{100 \times 10}$ $\quad$ $\underline{\text{rank}(A) = 4}$

$$\text{rank}(\tilde{A}) = 10$$

$$\tilde{A} = U \begin{bmatrix} 20.1 & & & & & & \\ & 15 & & & & & \\ & & 9 & & & & \\ & & & 4 & & & \\ & & & & 10^{-8} & & \\ & & & & & 10^{-9} & \\ & & & & & & 10^{-9} \end{bmatrix}$$

threshold $\sigma_i < T$

better: $\frac{\sigma_i}{\sigma_1} < T$

# Near-low-rank matrices

# Example

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

$$\operatorname{rank}(A) = 2$$

$$\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}$$

$$\frac{\sigma_3}{\sigma_1}$$

# Matrix Norms



$$\|A-\tilde{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij}-\tilde{a}_{ij})^2}$$

$$\|M\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij}^2} = \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2}$$

$$\|M\|_2 = \max \|Mx\|_2 \qquad \|x\|_2=1$$

$$= \sigma_{max} = \sigma_1$$

# Low-rank Approximation

find $A$ s.t. $A - \tilde{A}$

$$\min_{A} \|A - \tilde{A}\|_F \quad \text{s.t.} \quad \text{rank}(A) \leq r$$

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

low rank approximation (Eckart–Young–Mirsky Lemma / theorem)

$$\tilde{A} = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \end{bmatrix} V^T$$

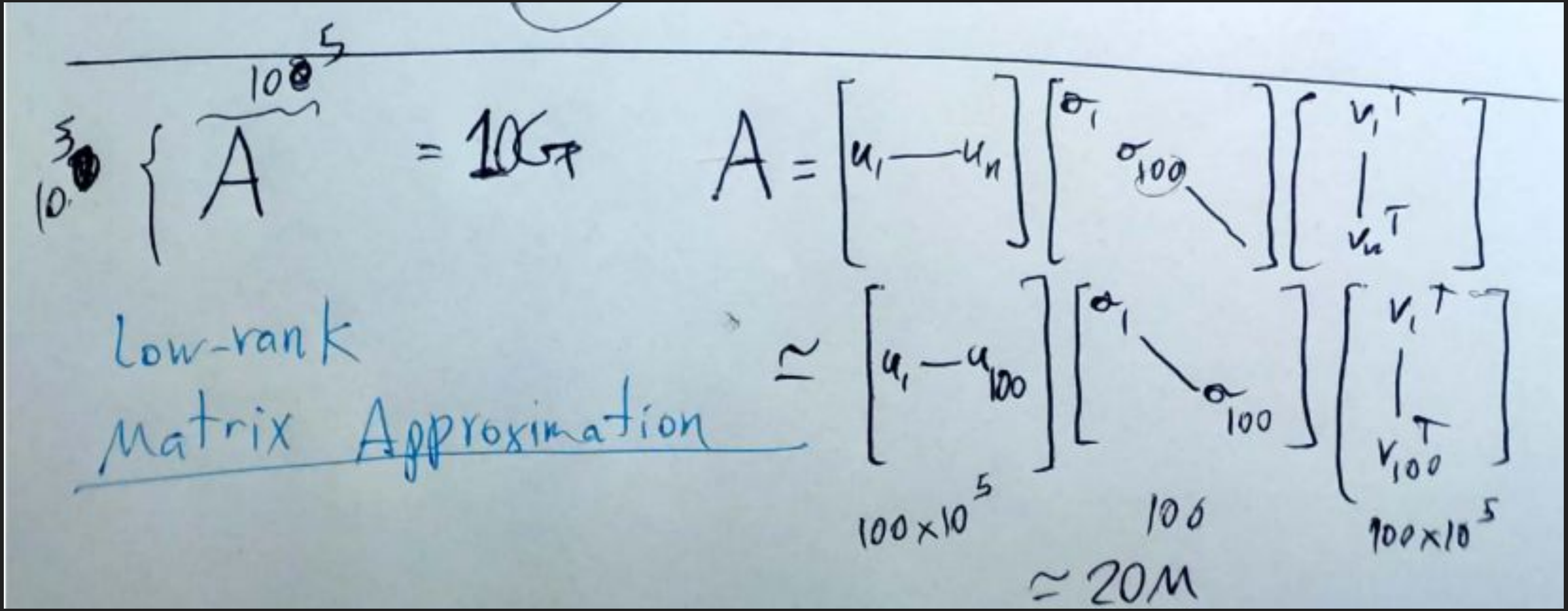
$$A = U \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix} V^T$$

# Low-rank Approximation

$$\tilde{A} = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \sigma_{r+1} & \\ & & & & \sigma \end{bmatrix} V^T$$

$$A = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \emptyset & \\ & & & & \emptyset \end{bmatrix} V^T$$

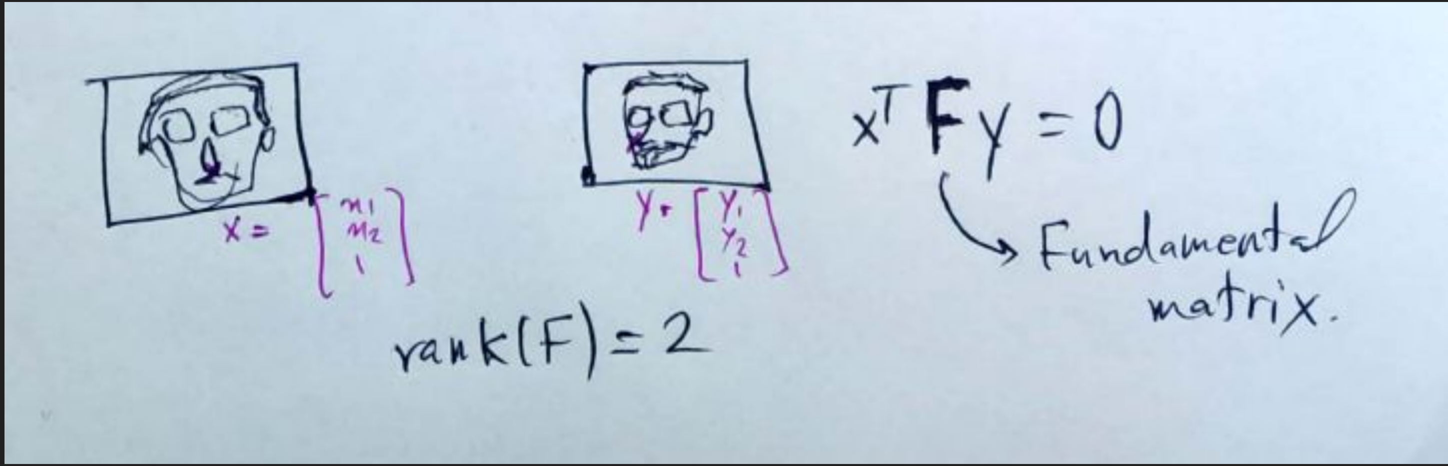
# Application: Reduce memory/computation



# Application: Reduce memory/computation

$$\tilde{A} = \begin{bmatrix} & & & \\ & & & \\ & & & \end{bmatrix} \simeq A' = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \Sigma \end{bmatrix} \begin{bmatrix} & V^T & \end{bmatrix}$$

# Application: Computing the Fundamental Matrix in computer vision



# Eigen-decomposition as an optimization problem

$A$ symmetric

$$\max_{x} \frac{x^TAx}{x^Tx} = \max_{x} \frac{x^TAx}{\|x\|^2}$$

$$x^Tx = 1$$

$$\begin{matrix}\max \\ \min\end{matrix} \; x^TAx \quad \text{subject to} \quad \|x\| = 1$$

$$\vec{v}_1 . \vec{v}_2 \dots \vec{v}_n$$

$$\lambda_1 \geqslant \lambda_2 \dots \geqslant \lambda_n$$

$$\lambda_1 = \lambda_{max} \qquad \lambda_n = \lambda_{min}$$

$$\lambda_{min} \leqslant \frac{x^TAx}{\|x\|^2} \leqslant \lambda_{max}$$

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

$$\lambda_{max} = \max \frac{x^TAx}{x^Tx} \qquad V_{max} = \underset{x}{\operatorname{argmax}} \frac{x^TAx}{x^Tx}$$

$$\lambda_{min} = \min \frac{x^TAx}{x^Tx}$$

K. N. Toosi
University of Technology

# Eigen-decomposition as an optimization problem

$A$ symmetric $\longrightarrow$ ~~eigen~~ $n$ real eigenvector

Eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n$

Corresponding Eigenvector $\vec{v}_1 \quad \vec{v}_2 \quad \cdots \quad v_{n-1} \quad v_n$

$$x_{max} = v_1$$

$$x_{min} = v_n$$

$\|x\| = 1$

$$\vec{v}_1 = \underset{x}{\arg\max}\; x^T A x \quad \text{subject to } x^T x = 1$$

$$\vec{v}_2 = \underset{x}{\arg\max}\; x^T A x \quad \text{s.t. } x^T x = 1,\; x \perp v_1$$

$$\vec{v}_3 = \underset{x}{\arg\max}\; x^T A x \quad \text{s.t. } x^T x = 1,\; x \perp v_1,\; x \perp v_2$$

$$A = V \Lambda V^T = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \cdots + \lambda_n v_n v_n^T$$

$$Ax = \lambda_1 v_1 v_1^T x + \lambda_2 v_2 v_2^T x + \cdots + \lambda_n v_n v_n^T x$$

# SVD as an optimiztion problem

K. N. Toosi
University of Technology

$$\max_{x} \|Ax\| \quad \text{subject to} \quad \|x\|=1$$

$$\frac{v_1^T A v_1}{v_1^T v_1} = \frac{v_1^T(\lambda_1 v_1)}{1}$$

$$\lambda_1 v_1^T v_1 = \lambda_1$$

$$\max_{x} \frac{\|Ax\|}{\|x\|} = ?$$

$$\max \frac{\|Ax\|^2}{\|x\|^2} = \frac{(Ax)^T(Ax)}{x^Tx} = \frac{x^T(A^TA)x}{x^Tx}$$

$$\max_{x} \frac{\|Ax\|^2}{\|x\|^2} = \lambda_{max}(A^TA) = \sigma_{max}^2 = \sigma_1^2$$

$$\max_{x} \frac{\|Ax\|}{\|x\|} = \sigma_1 = \sigma_{max}$$

$$\operatorname*{argmax}_{x} = \frac{\|Ax\|}{\|x\|} = \vec{v}_1 \qquad A = U\Sigma \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

# SVD as an optimiztion problem

LA29 (IV)

$\underbrace{y = Ax}_{} \in \mathbb{R}^m$, $A$: $m \times n$, $x \in \mathbb{R}^n$

$$x_{max} = \underset{x}{\operatorname{argmax}} \|Ax\|_2 \quad \text{subject to } \|x\| = 1$$

$$= \underset{x}{\operatorname{argmax}} \|Ax\|_2^2 \quad \text{subject to } x^T x = 1$$

$$= \operatorname{argmax} (Ax)^T (Ax) \quad \text{"} \quad \text{"} \quad \text{"}$$

$$= \operatorname{argmax} x^T \underbrace{A^T A}_{} x \quad \text{s.t. } x^T x = 1$$

$x_{max}$ = the eigenvector of $A^T A$ corresponding to the largest eigenvalue.

$A^T A$ symmetric & Positive Semidefinite

$\Longrightarrow x_{max} = v_1$ = the right singular vector of $A$ correspond to largest singular value.

# SVD as an optimiztion problem

K. N. Toosi
University of Technology

$$\min_{x} \|Ax\| \quad \text{s. t.} \quad \|x\|=1 \quad = \sigma_{min}$$

$$\operatorname*{argmin}_{x} \|Ax\| \quad \text{s. t.} \quad \|x\|=1 \qquad V_{min}$$

$$A = U \begin{bmatrix} \sigma_{max} & & \\ & \sigma_2 & \\ & & \ddots & \\ & & & \sigma_{min} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \\ \vdots \\ V_{min}^T \end{bmatrix}$$

# SVD as an optimiztion problem

K. N. Toosi
University of Technology

$A \in \mathbb{R}^{m\times n}$ $\qquad x_{min} = v_n$

$x_{max}$

$$A = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{\min(m,n)} \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$x_{min}$

$$x = v_1 \implies \|Ax\| = \|\sigma_1 u_1\| = \sigma_1$$

# Remember: Solving Homogeneous Equations

K. N. Toosi
University of Technology

IX

$$\begin{cases} 2x + 3y + z = 0 \\ -4x + y - z = 0 \\ x - y = 0 \end{cases}$$

homogeneous equations

if $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is a solution, so is $\begin{bmatrix} \alpha x \\ \alpha y \\ \alpha z \end{bmatrix}$ for all $\alpha \in \mathbb{R}$

$$\begin{bmatrix} 2 & 3 & 1 \\ -4 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$A\vec{x} = \vec{0}$$

find $N(A)$ $\vec{x} \in N(A)$

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \equiv \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

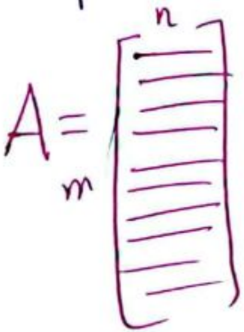
homography transform

Example!

# Application: Noisy Homogeneous Equations

K. N. Toosi
University of Technology

In practice $A$ has full rank, $N(A) = \{0\}$




$m \gg n$ in most situations.

$r = \text{rank}(A) = n$

$Ax = 0 \Rightarrow$ no solution!

find $x_* = \underset{x}{\text{argmin}} \|Ax\|$ subject to $\|x\| = 1$

$$= \text{argmin} \frac{\|Ax\|}{\|x\|} = \text{argmin} \frac{\|Ax\|^2}{\|x\|^2} = \frac{x^T A^T A x}{x^T x}$$

$$A = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$x_* = v_n$

singular vector corr. to smallest $\sigma_i$

# Zero centering data

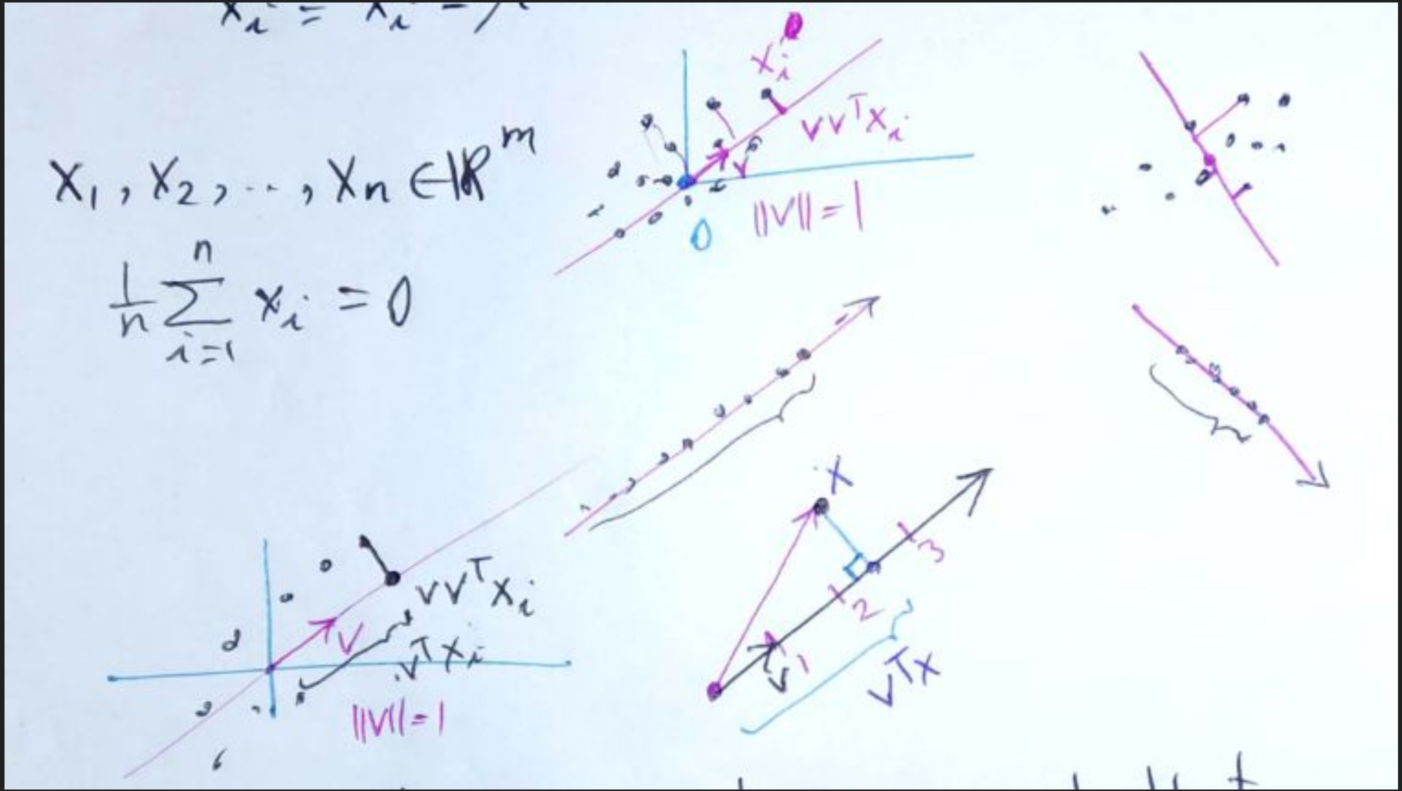
$$x_1', x_2', \dots, x_n' \in \mathbb{R}^m$$

$$X' = [x_1', x_2' \dots x_n']$$

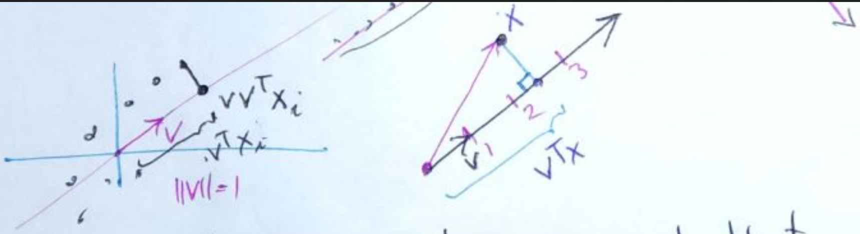

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i'$$

$$x_i = x_i' - \mu$$

# Projecting points on a line



# Find direction maximizing variance




find the direction v such that the variance of the projected points is maximized

$$v = \arg\max_{v} \sum_{i=1}^{n} (v^T x_i)^2 \quad \text{s.t. } \|v\|=1$$

$$\sum_{i=1}^{n} (v^T x_i)^2 = \sum_{i=1}^{n} (v^T x_i)(x_i^T v)$$

$$= \sum_{i=1}^{n} v^T (x_i x_i^T) v$$

$$= v^T \Big( \sum_{i=1}^{n} x_i x_i^T \Big) v$$

covariance matrix

# Find direction maximizing variance

K. N. Toosi
University of Technology

LA29 (VI)


$$\mathrm{Var}(x_1 \;\text{—}\; x_n)$$

$$\mathrm{var} = v^T \Big( \underbrace{\sum \underset{n\times 1}{x_i} \underset{1\times n}{x_i^T}}_{n \times n} \Big) v$$

→ covaricon

$$v^T \sum_{i=1}^{n} (x_i' - \mu)(x_i' - \mu)^T v$$

$$X = [x_1, x_2, \cdots x_n]$$

$$C = X X^T = [x_1 \;\text{—}\; x_n] \begin{bmatrix} x_1^T \\ | \\ x_n^T \end{bmatrix} = \sum_{i=1}^{n} x_i \cdot x_i^T$$

covaricnce matrix

# Find direction maximizing variance

$X = [x_1, x_2, \cdots x_n]$

$$C = XX^T = [x_1 — x_n] \begin{bmatrix} x_1^T \\ | \\ x_n^T \end{bmatrix} = \sum_{i=1}^{n} x_i \cdot x_i^T$$

covariance matrix

$$v_{max} = \arg\max_{v} \; v^T (XX^T) v \quad \text{s.t.} \quad \|v\| = 1$$

$v_{max}$ = the eigenvector of $XX^T$ ~~covar~~ corresponding to the largest eigenvalue

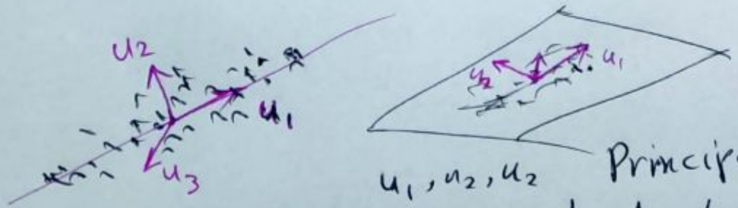
= the first ~~left~~ righ singular vector of $X^T$

= the first left singular vector of $X = u_1$

# Principal Component Analysis (PCA)

max

$v_{max}$ = the eigenvector of $XX^T$ ~~corres~~
corresponding to the largest eigenvalue
= the first ~~lef~~ righ singular vector
of $X^T$
= the first left singular vector
of $X$ $= u_1$




$u_1, u_2, u_2$ Principal Components

Principal Component Analysis (PCA)