EECS 16B   Module 3, Lecture 3   Apr 8, 2021

- Unsupervised machine learning.
- Principal Component Analysis.

  ↳ Find underlying low dimensional structure.

  e.g. sorting neurons

  movie recommendations.

Useful formula : $\sum_{i=1}^{n} \vec{u}_i \vec{v}_i^T = UV^T$ "Outer product"

Show: $\forall \vec{x} \in \mathbb{R}^n$

$$\sum_{i=1}^{n} \vec{u}_i \vec{v}_i^T \vec{x} = UV\vec{x}$$

Reference: $A = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 8 & 10 \end{bmatrix}_{2\times 4}$

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}}_{\substack{U \\ 2\times 2}} \underbrace{\begin{bmatrix} \sqrt{230} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\substack{\Sigma \\ 2\times 4}} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{46}} & \sqrt{\frac{2}{23}} & 2\sqrt{\frac{2}{23}} & \frac{5}{\sqrt{46}} \\ \frac{-5}{\sqrt{26}} & 0 & 0 & \frac{1}{\sqrt{26}} \\ \frac{-2}{\sqrt{273}} & 0 & \sqrt{\frac{13}{21}} & \frac{-10}{\sqrt{273}} \\ \frac{-1}{\sqrt{483}} & \sqrt{\frac{21}{23}} & \frac{-4}{\sqrt{483}} & \frac{-5}{\sqrt{483}} \end{bmatrix}}_{\substack{V^T \\ 4\times 4}}$$

Neurons:

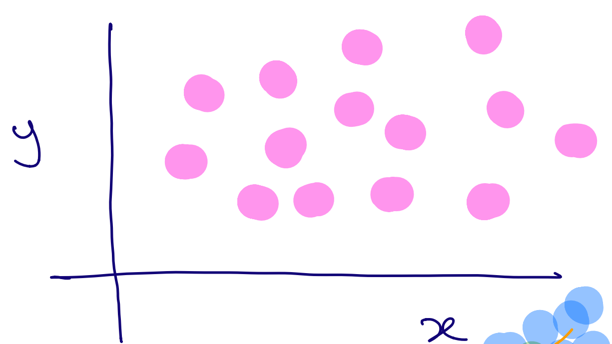
→ features

→ peak ht

peak - trough ht

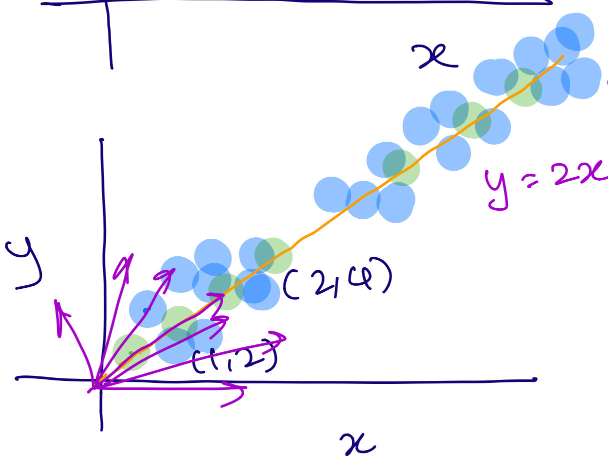
time

"Principal Components" i.e. lower dimensional structure in simple 2D data.



2D

$\mathbb{R}^2$



Underlyin 1D structure.

even ~~too~~ though' the ambient dimension is 2D.

(1, 2)
(2, 4)
(4, 8)
(5, 10)

2D points.

Data in columns.

$$A = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 8 & 10 \end{bmatrix} \longrightarrow \text{span}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Principal Component.

$$A = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \overset{\sigma_1}{\sqrt{230}} & 0 & 0 & 0 \\ 0 & \overset{\sigma_2}{0} & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{46}} & \sqrt{\frac{2}{23}} & 2\sqrt{\frac{2}{23}} & \frac{5}{\sqrt{46}} \\ -\frac{5}{\sqrt{26}} & 0 & 0 & \frac{1}{\sqrt{26}} \end{bmatrix}$$

$$\underset{2\times 2}{U} \qquad \underset{2\times 4}{\Sigma} \qquad \begin{bmatrix} \frac{-2}{\sqrt{273}} & 0 & \sqrt{\frac{13}{21}} & \frac{-10}{\sqrt{273}} \\ \frac{-1}{\sqrt{483}} & \sqrt{\frac{21}{23}} & \frac{-4}{\sqrt{483}} & \frac{-5}{\sqrt{483}} \end{bmatrix}$$

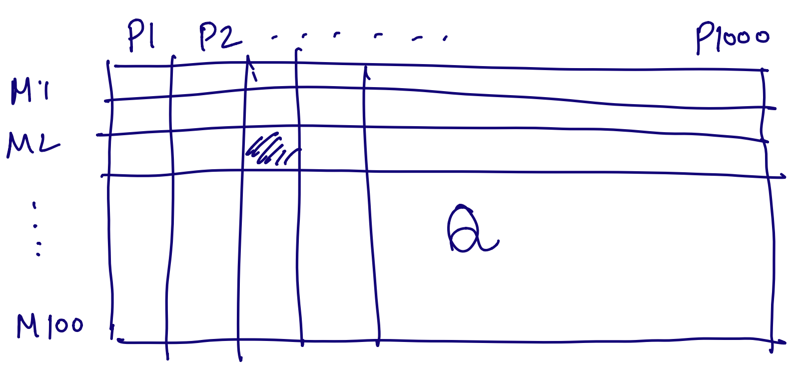
$$= \underbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}}_{U_r} \begin{bmatrix} \sqrt{230} \end{bmatrix} \underbrace{\begin{bmatrix} \quad \end{bmatrix}}_{V_r^T}$$

U_r: 2×1 matrix

U_r is a basis for the columnspace of A.

$\begin{bmatrix} x \\ y \end{bmatrix} \in \text{Col}(A)$ if $y = 2x$.

<u>Generalizing:</u> <u>Movie recommendation problem.</u>



Q :

Q :

q_{ij} : rating of person j for movie i

Goal: Understand different types of movies
↳ recomendations for people.

16A style:
- "Make a model"
- "Learn the model"
- "Use model to make predictions"

Say every movie is represented by 4 numbers.

Score for m_i

$$\begin{bmatrix} a_i & b_i & c_i & d_i \end{bmatrix}$$

"action" "bechdel test" "comedy" "drama"

Person j: sensitivity's to components.

Person j : $[\, s_{aj} \quad s_{bj} \quad s_{cj} \quad s_{dj} \,]$

$$\boxed{q_{ij} = s_{aj} \cdot a_i + s_{bj} b_i + s_{cj} \cdot c_i + s_{dj} \cdot d_i}$$

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{100} \end{bmatrix}$$ action scores of all movies

$\in \mathbb{R}^{100}$

$\vec{b}$, $\vec{c}$, $\vec{d}$ -

$$\vec{S_a} = \begin{bmatrix} Sa_1 \\ Sa_2 \\ \vdots \\ Sa_{1000} \end{bmatrix} \quad \text{"sensivity"} \quad \in \mathbb{R}^{1000}$$

$\vec{S_b}, \vec{S_c}, \vec{S_d}, \ldots$

Consider:

$$\vec{a} \cdot \vec{S_a}^T \qquad (100 \times 1)\,(1 \times 1000)$$

"outer product"

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{100} \end{bmatrix} \begin{bmatrix} Sa_1 & Sa_2 & \cdots & Sa_{1000} \end{bmatrix}$$

$$= \begin{bmatrix} a_1 Sa_1 & a_1 Sa_2 & \cdots & a_1 Sa_{1000} \\ a_2 Sa_1 & & & \vdots \\ \vdots & & & \\ a_{100} Sa_1 & \cdots & & a_{100} Sa_{1000} \end{bmatrix}$$

$$Q = \vec{a} \cdot \vec{S_a}^T + \vec{b} \cdot \vec{S_b}^T + \vec{c} \cdot \vec{S_c}^T + d \cdot \vec{S_d}^T$$

almost diagonal

$$Q = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i \vec{u_i} \vec{v_i}^T$$

r: rank of the matrix Q.

The $\overset{k}{2}$ principal components of matrix Q.

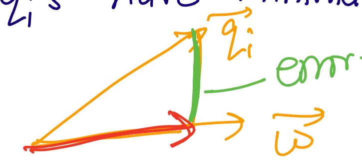
- along the columns : $\vec{u}_1, \vec{u}_2 \cdots, \vec{u}_k$
- along the rows : $\vec{v}_1, \vec{v}_2, \ldots \vec{v}_k$

$$Q = \begin{bmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & & & | \end{bmatrix}$$

"Data" is organized by columns.

Goal: Is to find the first principal component, i.e. that direction $\vec{w}$ such that it captures most of the data.

$\rightarrow$ We want a $\vec{w}$ such that, when I project $\vec{q}_i$ onto $\vec{w}$, then the q_i's have minimum error.



Consider : $\|\vec{w}\|^2 = 1$

Projection of $\vec{q}_i$ onto $\vec{w}$: $\langle \vec{q}_i, \vec{w} \rangle \vec{w}$

(because $\|\vec{w}\|^2 = 1$)

Squared-norm of Error: $\left\|\left(\vec{q_i} - \langle \vec{q_i}, \vec{\omega} \rangle \vec{\omega}\right)\right\|^2$

Summing error over all points:

$$\boxed{\underset{\vec{\omega}}{\text{minimize}} \sum_{i=1}^{n} \underbrace{\left\| \vec{q_i} - \langle \vec{q_i}, \vec{\omega} \rangle \vec{\omega} \right\|^2}_{\text{simplify.}}}$$

find $\vec{\omega}$ that achives the min.

"arg min"

Simplifying: $\left\| \vec{q_i} - \langle \vec{q_i}, \vec{\omega} \rangle , \vec{\omega} \right\|^2$

$$= \left(\vec{q_i} - \langle \vec{q_i}, \vec{\omega} \rangle \cdot \vec{\omega}\right)^T \left(\vec{q_i} - \langle \vec{q_i}, \vec{\omega} \rangle . \vec{\omega}\right)$$

$$= \|\vec{q_i}\|^2 + \langle \vec{q_i}, \vec{\omega} \rangle^2 \underbrace{\|\vec{\omega}\|^2}_{1}$$

$$- 2 \cdot \vec{q_i}^T \langle \vec{q_i}, \vec{\omega} \rangle \vec{\omega}.$$

$$= \|\vec{q_i}\|^2 + \langle \vec{q_i}, \vec{\omega} \rangle^2 - 2 \langle \vec{q_i}, \vec{\omega} \rangle^2$$

$$= \boxed{\|\vec{q_i}\|^2 - \langle \vec{q_i}, \vec{\omega} \rangle^2}$$

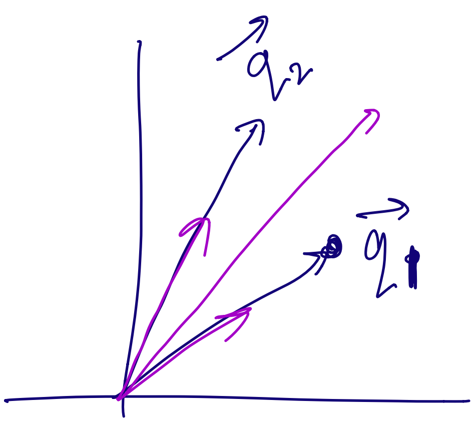
Rewrite as:

$$\text{Minimize}_{\vec{w}} \; \sum_{i=1}^{n} - \langle \vec{q_i}, \vec{w} \rangle^2$$

OR

$$\boxed{\text{Maximize}_{\vec{w}} \; \sum_{i=1}^{n} \langle \vec{q_i}, \vec{w} \rangle^2}$$

Choose a $\vec{w}$ such that $\sum_{i=1}^{n} \langle \vec{q_i}, \vec{w} \rangle^2$ is as large as possible.



$$\max_{\vec{w}} \; \sum_{i=1}^{n} \langle \vec{q_i}, \vec{w} \rangle^2$$

$$= \max_{\vec{w}} \left[\sum_{i=1}^{n} \langle \vec{w}, \vec{q_i} \rangle \langle \vec{q_i}, \vec{w} \rangle \right]$$

$$= \max_{\vec{w}} \left[\sum_{i=1}^{n} \vec{w}^T \vec{q_i} \, \vec{q_i}^T \vec{w} \right]$$

do not depend on i

$$= \max_{\vec{w}} \left[\vec{w}^T \underbrace{\left(\sum_{i=1}^{n} \vec{q_i} \, \vec{q_i}^T \right)}_{\text{outer product!!}} \vec{w} \right]$$

$$\sum_{i=1}^{n} 2 i = 2 \left(\sum_{i=1}^{n} i \right)$$

$$= \max_{\vec{w}} \vec{w}^T Q \cdot Q^T \vec{w}$$

$$= \max_{\vec{w}} \vec{w}^T U \Sigma V^T (U \Sigma V^T)^T \vec{w}$$

Formula:

$$\sum \vec{u_i} \vec{v_i}^T = U V^T$$

$$= \max_{\vec{w}} \vec{w}^T U \Sigma V^T V \Sigma^T U^T \vec{w}$$

$$= \max_{\vec{w}} \vec{w}^T U \Sigma \Sigma^T U^T \vec{w}$$

$$= \max_{\vec{w}} \ \vec{w}^T U \cdot \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{100}^2 \end{bmatrix} \underbrace{U^T \vec{w}}_{\tilde{\vec{w}}}$$

Define : $U^T \vec{w} = \tilde{\vec{w}}$

$\Rightarrow \vec{w}^T U = \tilde{\vec{w}}^T$

$$= \max_{\tilde{\vec{w}}} \ \tilde{\vec{w}}^T \begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sigma_{100}^2 \end{bmatrix} \tilde{\vec{w}}$$

unit

$$\tilde{\vec{w}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{e}_1$$

"Eckhart-Yong" Thm.

$= \cancel{\max} \ \sigma_1^2$ if $\tilde{\vec{w}} = \vec{e}_1$

$\tilde{\vec{w}} = \vec{e}_1$

Then $\vec{w} = U \cdot \tilde{\vec{w}} = U \cdot \vec{e}_1 = \vec{u}_1$

$$\Sigma = {\scriptstyle 100}\underbrace{\begin{bmatrix} \ddots & 0 \end{bmatrix}}_{1000}$$

$$\Sigma^T = \overset{100}{\begin{bmatrix} \Sigma \\ 0 \end{bmatrix}} \Big\} {\scriptstyle 1000}$$

$$\Sigma \cdot \Sigma^T = 100 \times 100$$

diagonal matrix.

$$\begin{bmatrix} \Sigma & 0 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix}$$

$$= \Sigma_1^2 \quad 100 \times 100$$