Principal Component Analysis (PCA)

- Math
- Intuition
- Code
- Limitations

@end, delhivery case-study (Q&A)

SCALER

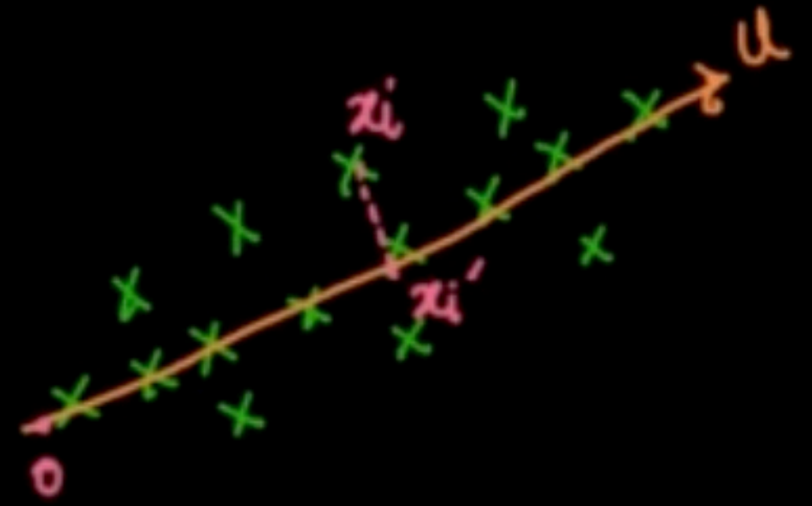
Malhi

$$\max_{u} \quad \frac{1}{n} \sum (u^T x_i^{'2} - u^T \bar{x}_i^{'2})^2$$

} → Let us simplify this

Gradient

$$s.t \quad \|u\| = 1$$



$\bar{x}$ = mean of x_is

mean of x_i's = $u^T \bar{x}$

Detour

Data matrix

features

1 2 3 ... j ... d

$X =$

1
2
3
...
i
...
n

x_i^T

x_{1j}
x_{2j}
x_{ij}
x_{nj}

$\mathbb{R}^d \ni \bar{x}$: $\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ ... $\bar{x}_j$ ... $\bar{x}_d$

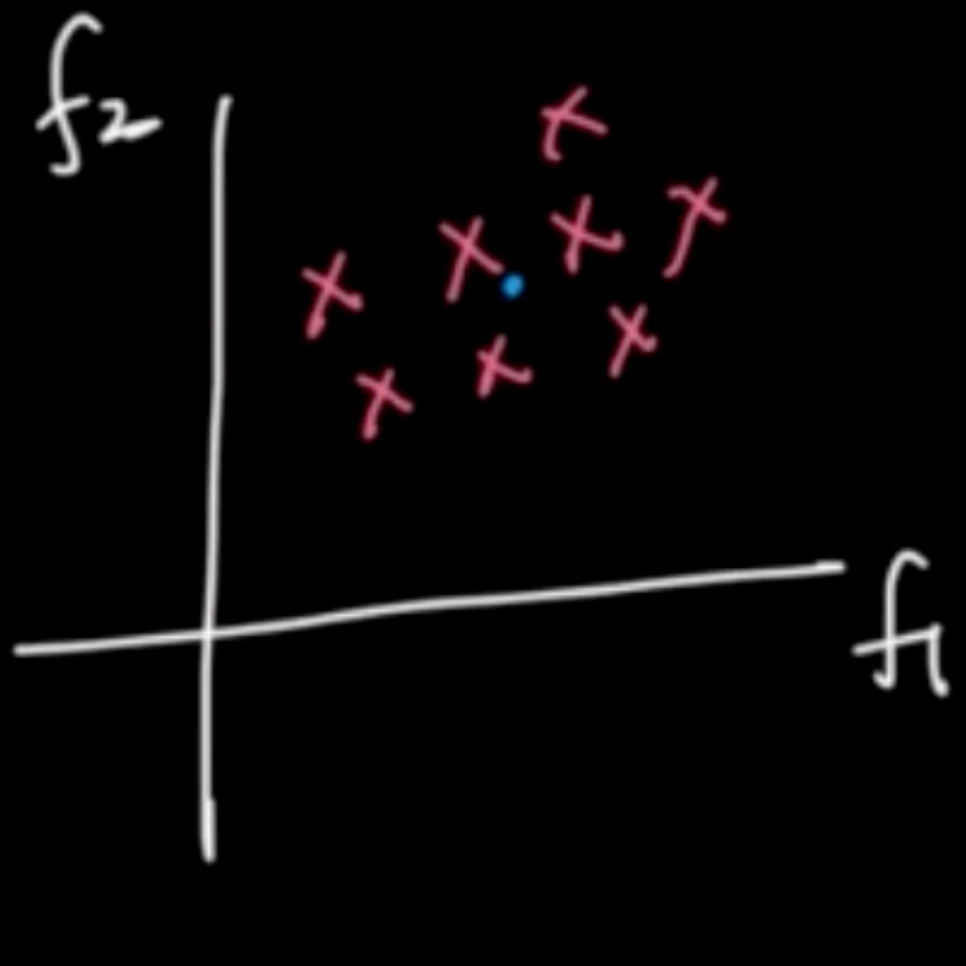
Column Standardization

① Mean Centering

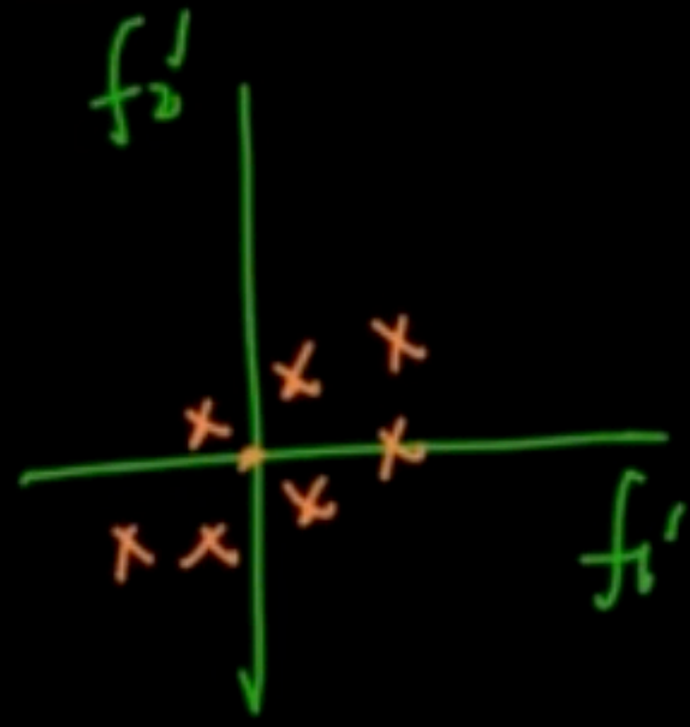
② Variance Scaling

$x_i - \bar{x}$

$\mu_i = \frac{1}{n}$



Mean Centering



e.g: $A = N(\mu, \sigma^2)$

$$\frac{A - \mu}{\sigma} \sim N(0,1)$$

Variance scaling:

$$X = \begin{matrix} & 1 & 2 & \cdots & j & \cdots & d \\ 1 & & & & & & \\ 2 & & & & & & \\ \vdots & & & & & & \\ i & & & & x_{ij} & & \\ \vdots & & & & & & \\ n & & & & & & \end{matrix}$$

x_{1j}, x_{2j}, x_{ij}, x_{nj}

$\sigma_1 \quad \sigma_2 \cdots \quad \sigma_j \quad \cdots \quad \sigma_d$

i^{th} pt ; j^{th} feature

$$x_{ij}^{new} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

Mean-centering & Variance scaling → Standardization

↓

every Column's mean = 0

Column's Var = 1 = std-dev

optimization faster ↑

Simplicity ↑ Math ↓ costly

X^{raw} → pre-processing: Mean centring, Var-scaling → X → optimization problem

$$\max_{u} \frac{1}{n}\sum_{i=1}^{n} \left(u^T x_i - u^T \bar{x} \right)^2$$

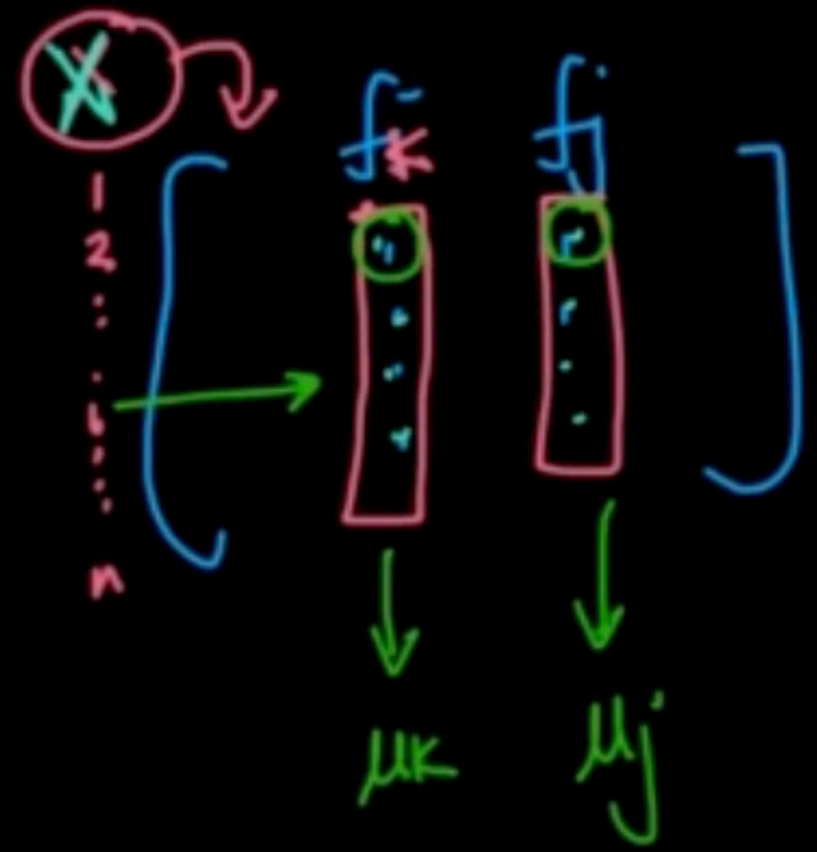
$\bar{x}$ → $[0, 0, 0 \cdots 0]^T$

$$s.t \quad \|u\| = 1$$

let

$$S_{ij} = \text{Covariance}(f_k, f_j)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)$$



$$S_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ik} \cdot x_{ij})$$

(8)

$x_1 \quad x_2 \quad x_3 \quad - \quad - \quad - \quad x_{10}$

$10, 20, 30, 40, \quad - - - \quad 62$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$x_i' = x_i - \bar{x}$$

$$\text{mean}(x_i') = 0$$

S_{11}

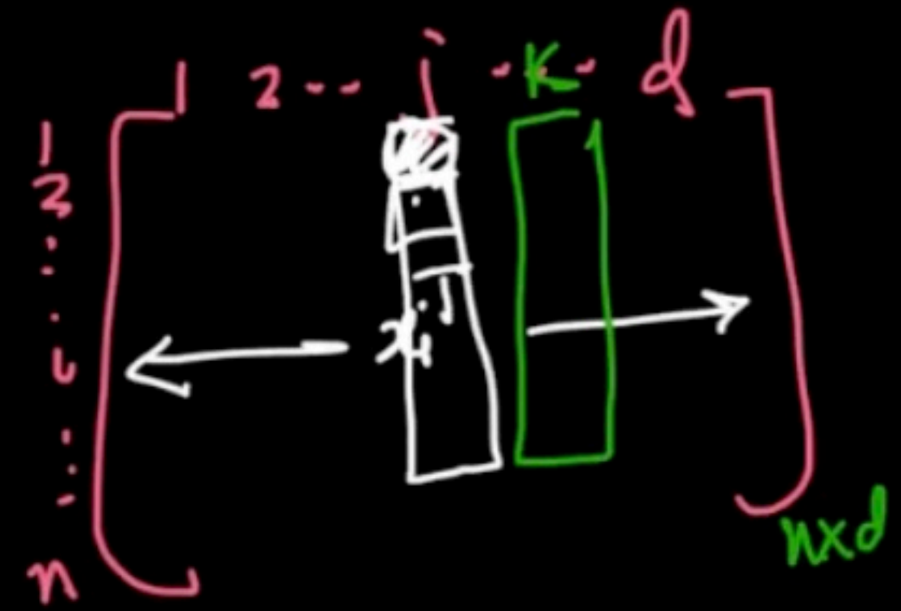
$$S_{kj} = \frac{1}{n} \sum_{i=1}^{n} x_{ik} \cdot x_{ij}$$

kth row jth col

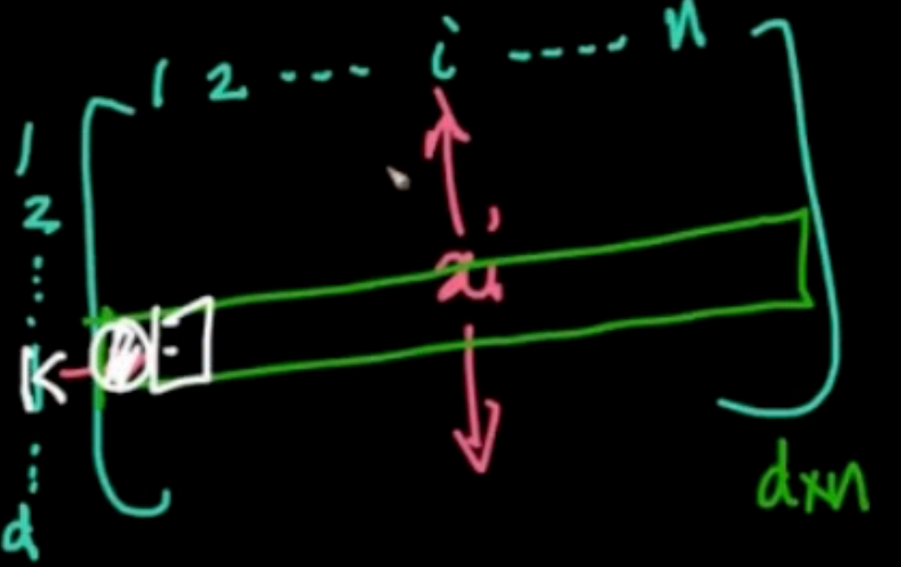
$$S_{dxd} = \frac{1}{n} \cdot X^{T}_{dxn} \cdot X_{nxd}$$

$$S_{kj} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_{ik} \cdot x_{ij}$$

$X =$



$X^T =$



Optmzn:

$$\max_{u} \quad \frac{1}{n} \sum_{i=1}^{n} (u^T x_i)^2$$

$$s.t \quad \|u\| = 1$$

$$\|u\| = 1 \rightarrow u^T u = 1$$

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

scalar

$u^T \cdot X = A$

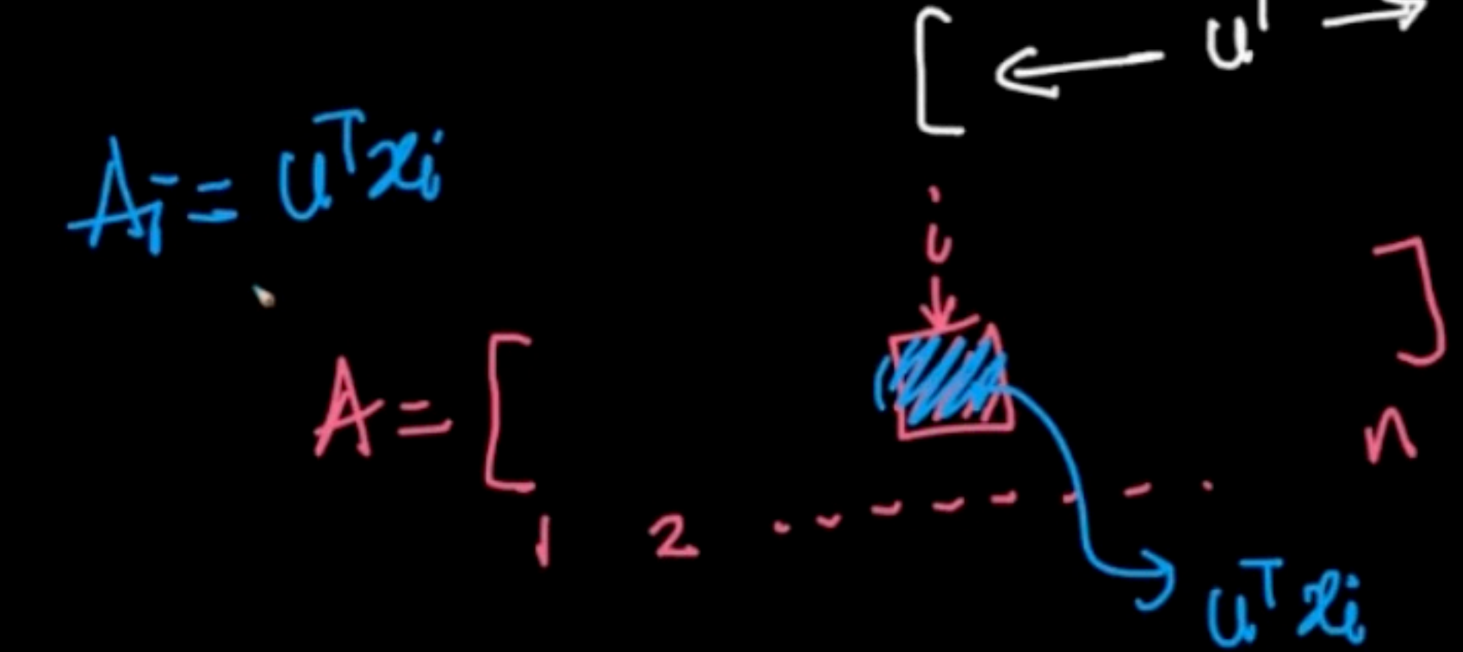
u^T: $1 \times d$, X: $d \times n$, A: $1 \times n$

$u \in \mathbb{R}^d$

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{bmatrix}_{d \times 1}$$

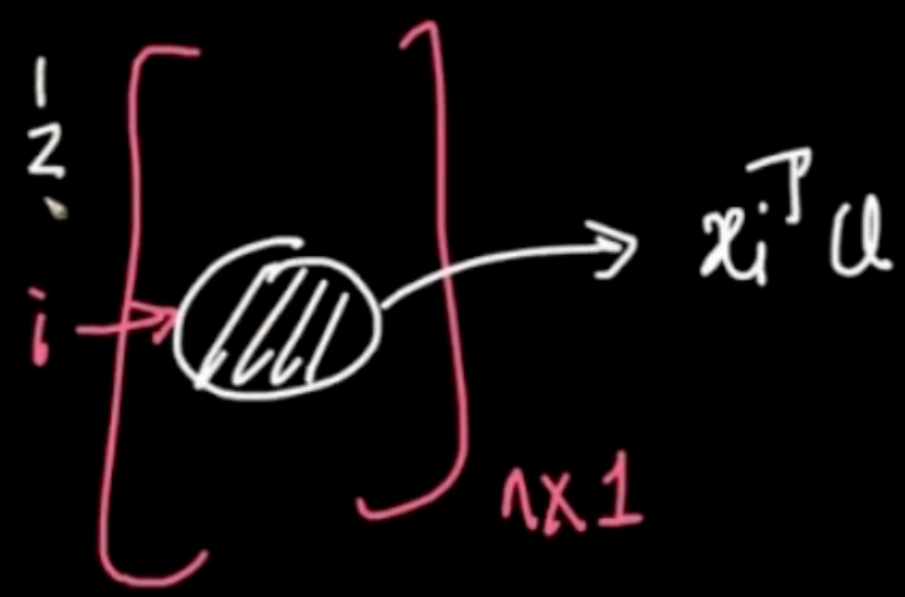
$A_i = u^T x_i$

$A = [\ \leftarrow u^T \rightarrow\ ]$



SCALER

$$B_{n\times 1} = X_{n\times d}\, U_{d\times 1} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ n \end{bmatrix} \begin{bmatrix} \leftarrow & x_i^T & \rightarrow \end{bmatrix} \cdot \begin{bmatrix} \uparrow \\ u \\ \downarrow \end{bmatrix}$$



$$(u^T x^T) \cdot (xu)$$

$$u^T x_i = x_i^T u$$

$$Res_{1\times 1} = A_{1\times n} \cdot B_{n\times 1}$$

$$\begin{bmatrix} 1 & 2 & \cdots & i & \cdots & n \\ \cdot & \cdot & \cdot & u^T x_i & \cdot & \cdot \end{bmatrix}$$

$$\begin{pmatrix} \vdots \\ \cdot \\ x_i^T u \\ \cdot \\ \vdots \end{pmatrix} \begin{matrix} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ n \end{matrix}$$

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

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

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

Optmzn:

$$\max_{u} : (u^T x^T)(xu)$$

$$s.t \quad u^T u = 1$$

$$\frac{1}{n} \overbrace{u^T X^T X u}^{\downarrow} = u^T S u$$

opt: (Cov of data)

$$\begin{cases} \max_{u} \; u^T S u \\ s.t \;\; u^T u = 1 \end{cases}$$

X: data matrix

$\downarrow$

S

$$\min \cancel{\max} \; -u^T \downarrow S u + \lambda(u^T u - 1)$$

(1)

$$\cancel{s.t \; u^T u = 1}$$

(2)

$$\min_u \; \underbrace{-u^T S u} + \lambda \, (u^T u - 1) \Rightarrow \mathcal{L}(u, \lambda)$$

$$\frac{\partial \mathcal{L}}{\partial u} = 0 \quad \Rightarrow$$

$$\min_{u} \; -u^T \underset{\downarrow}{S} u + \lambda(u^T u - 1)$$

(1)

~~s.t $u^T u = 1$~~

(2)

$$\min_{u} \; \underbrace{-u^T S u + \lambda\,(u^T u - 1)}_{\mathcal{L}(u,\lambda)}$$

Gradient descent

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial u} = 0 \end{cases} \Rightarrow -2(u^T S)^T + 2\lambda u = 0$$

$$\downarrow$$

$$S^T u$$

$$\frac{d\left(-\overbrace{u^T s u}^{A}\right)}{du} = -s^T u - su$$

$$= -2su$$

$$\frac{d\,Au}{du} = A^T$$

$$\frac{d(u^T A)}{du} = A$$

$$-2Su + 2\lambda u = 0$$

$$\Rightarrow \lambda u = Su$$

{ u's are eig-vec of S
λ's are eig-val of S

① Data matrix

pre-processing: Mean centering & var-scaling

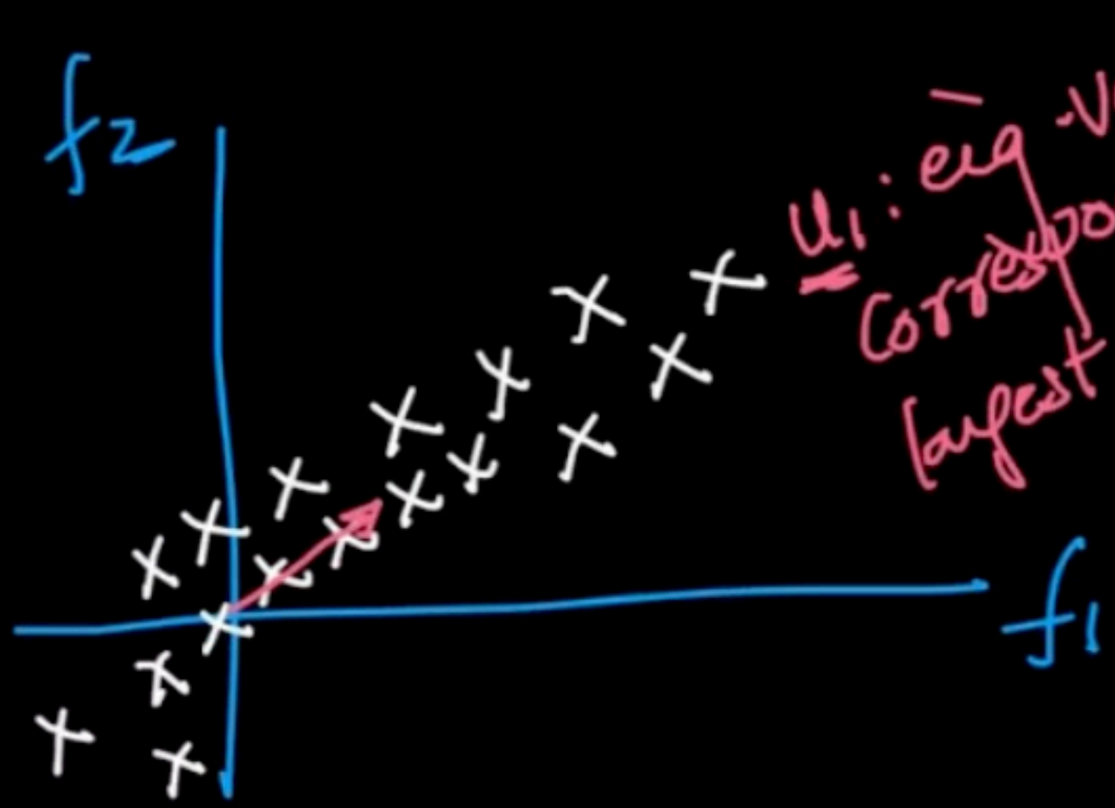
$X_{n \times d}$

② Compute $S_{d \times d}$ = Cov. matrix of X

③ eig-vec & eig val of S

$u_1, u_2, u_3 \dots u_d$

$\lambda_1, \lambda_2, \lambda_2 \dots \lambda_d$



$X_{n \times d}$

↓

$S_{d \times d}$

↓

$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$

$u_1, u_2 \cdots u_d$

largest

Viz

100-dim → 2-dim

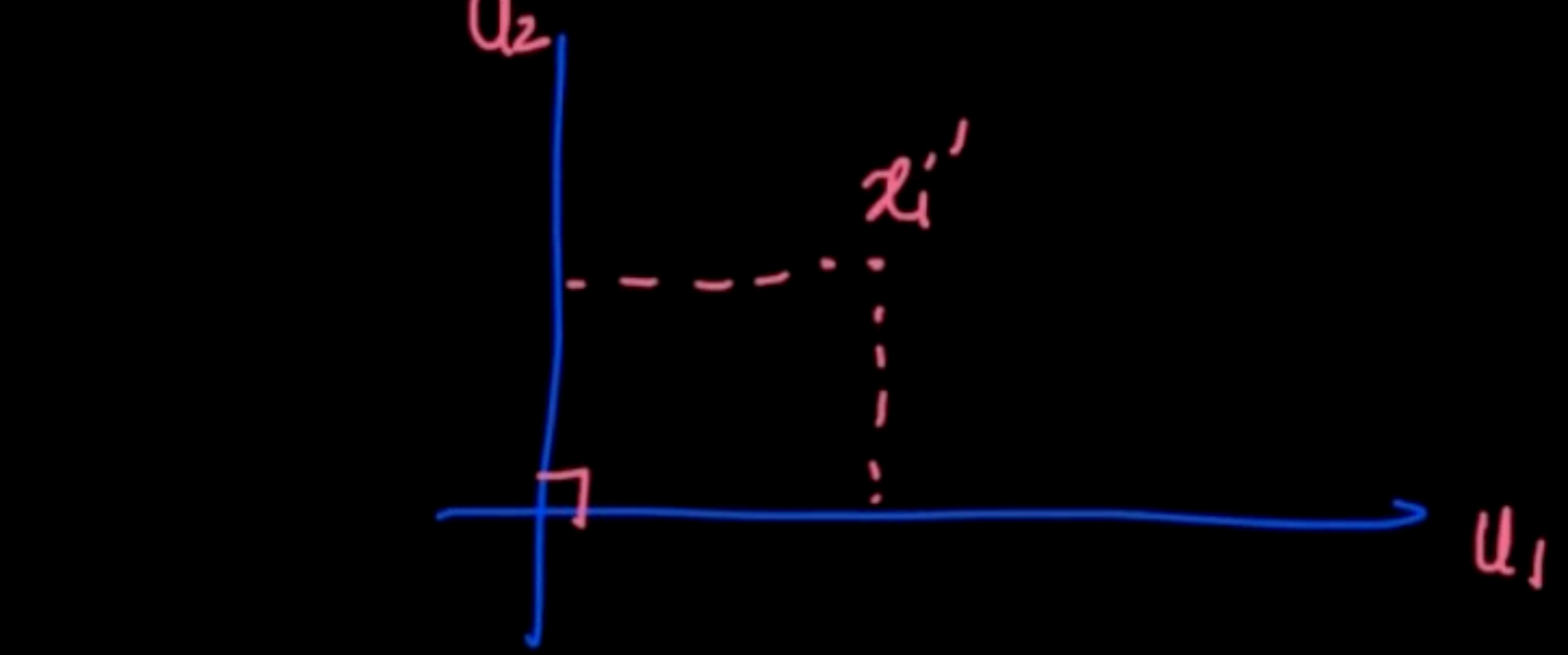
X

↳ S → $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{100}$

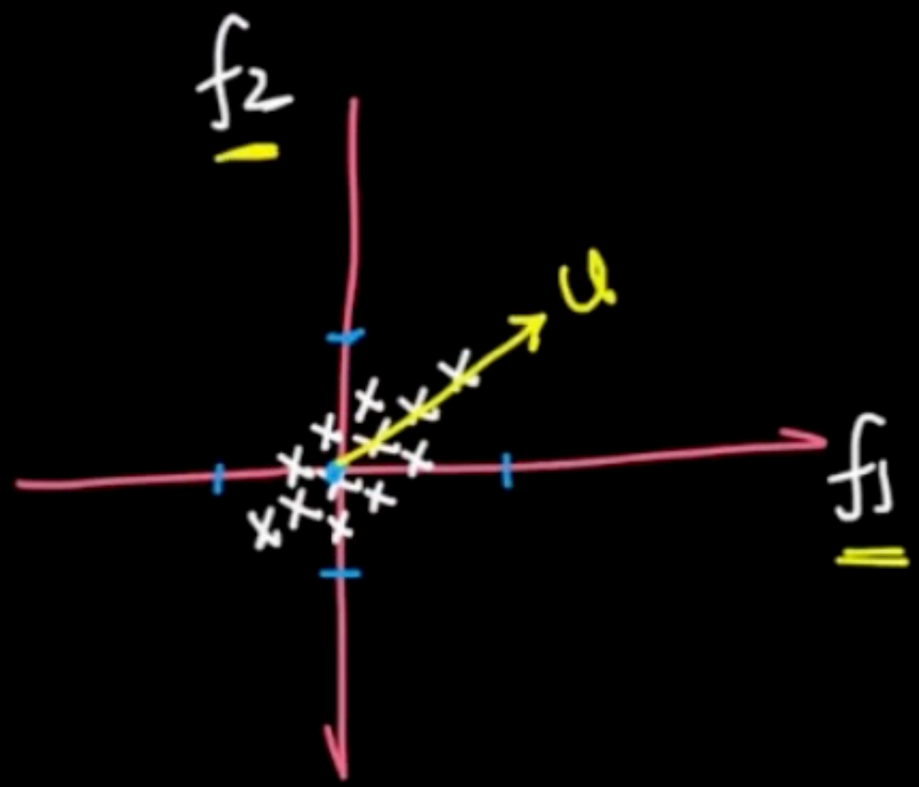
u_1 u_2 ... u_{100}

$$x_i \cdot u_1 = x_i'_1$$

$$x_i \cdot u_2 = x_i'_2$$



(Q)



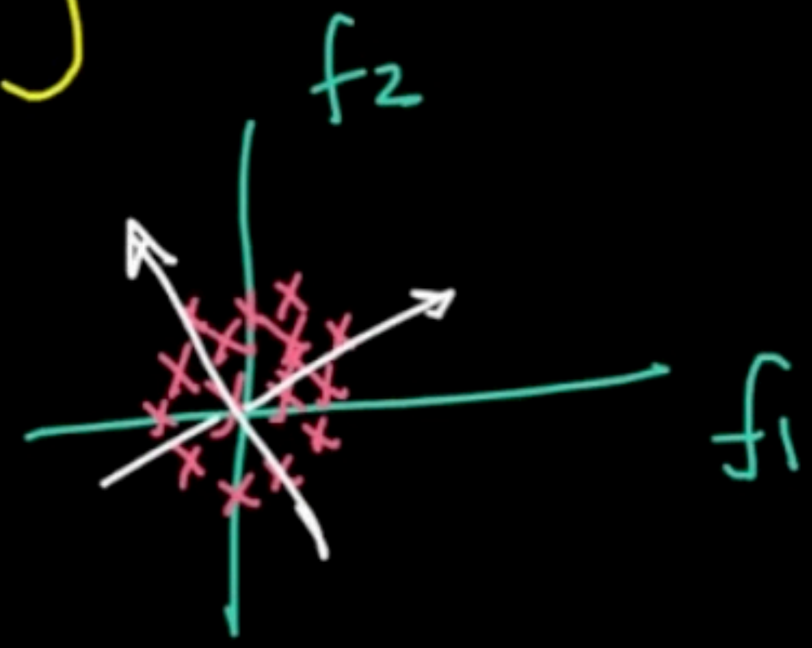
✓ mean centered

✓ var scaling

(Q)

Identity matrix

$$S = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \\ & & & 1 \\ & & & & 1 \end{bmatrix}$$



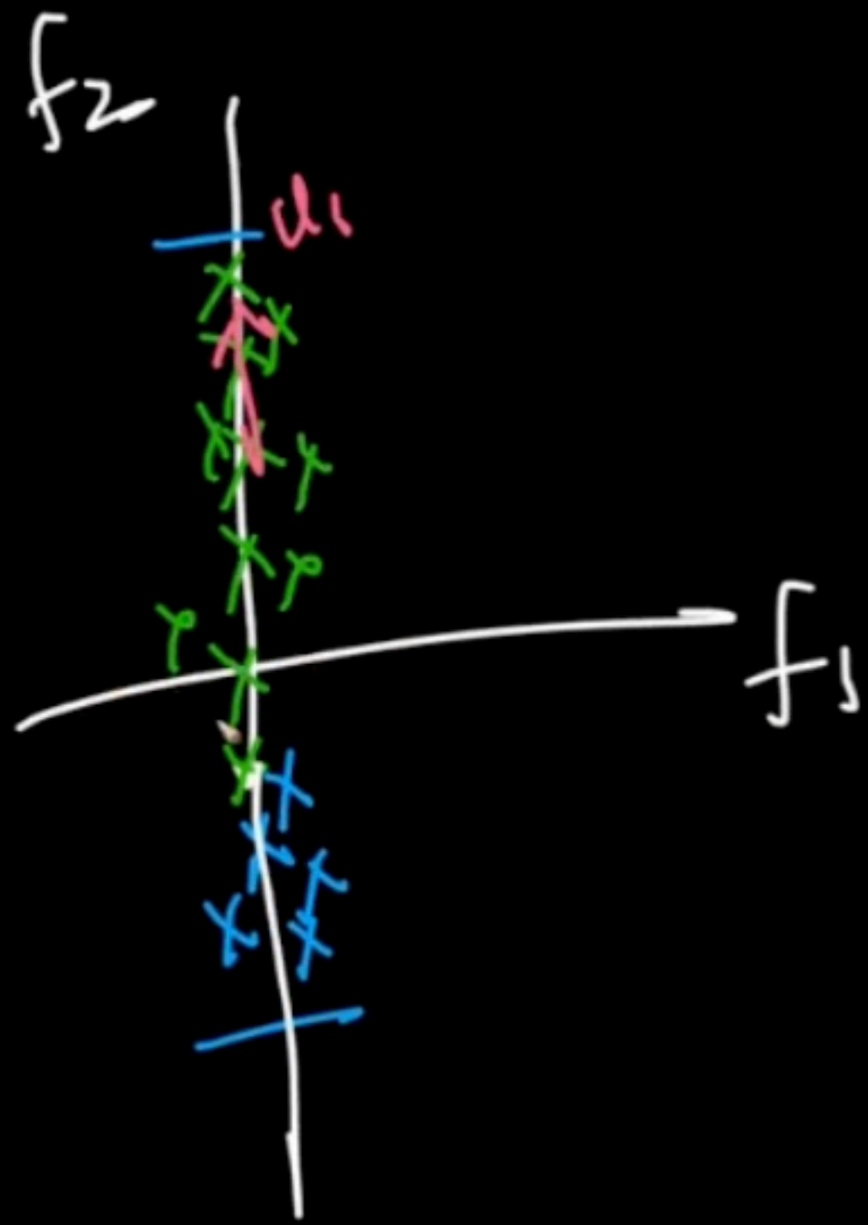
u_1 u_2 ---

↺ principal components

SCALER

Chrome File Edit View History Bookmarks Profiles Tab Window Help

(Q)



30 / 31

PCA → dim-reduction model

PCA → optmzn

↳ Malᵢₙ optmzn → finding λ's & u's

(Q)

100 -d ⟶ Classification or regression

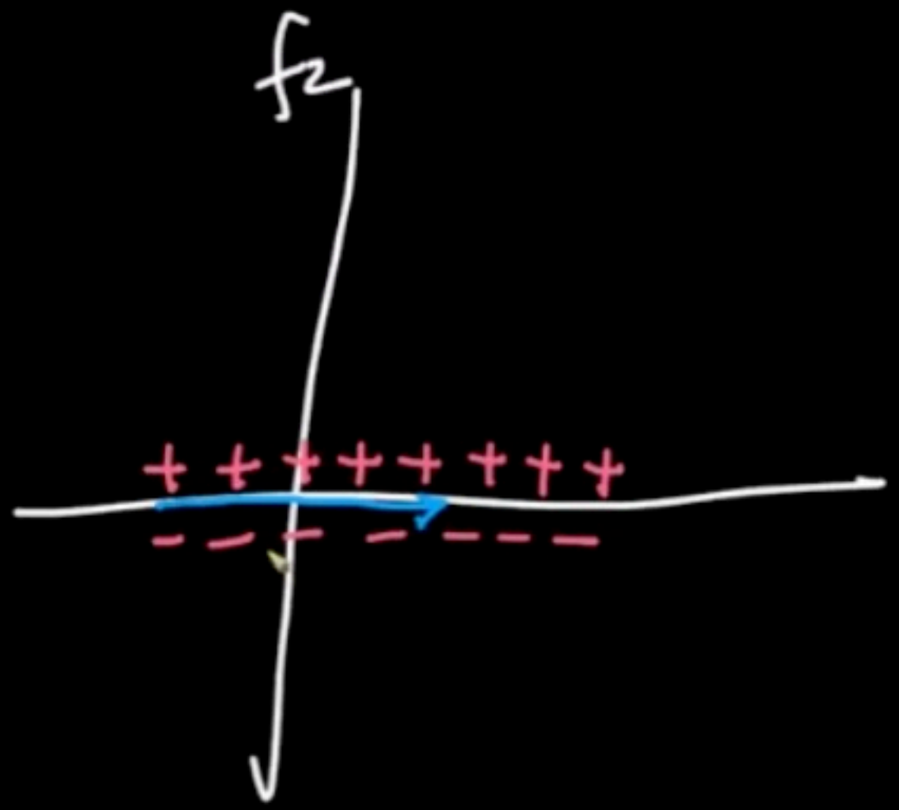
- Linear Reg or Logistic Reg or KNN

5-dim ⟶ ML (your can) (why ?)

$(u_1, u_2 \dots u_5)$

↓

loss of information



2D

PCA

doesnot use Class labels

Classfn?

(Q) Feature Imp

100 -d ⟶ Classification or regression

- Linear Reg or Logistic Reg or KNN

100 -d → 5-dim → ML

(you can) (why ?)

(u1, u2 .. u5)

↓

loss of information