Constrained Optimization → Classification & Regression

&

Dimensionality Reduction → PCA (next class)

class starts @ 9:03PM

t-SNE { ML-2 }

SVD

Deeplearning based {NN}

classification:

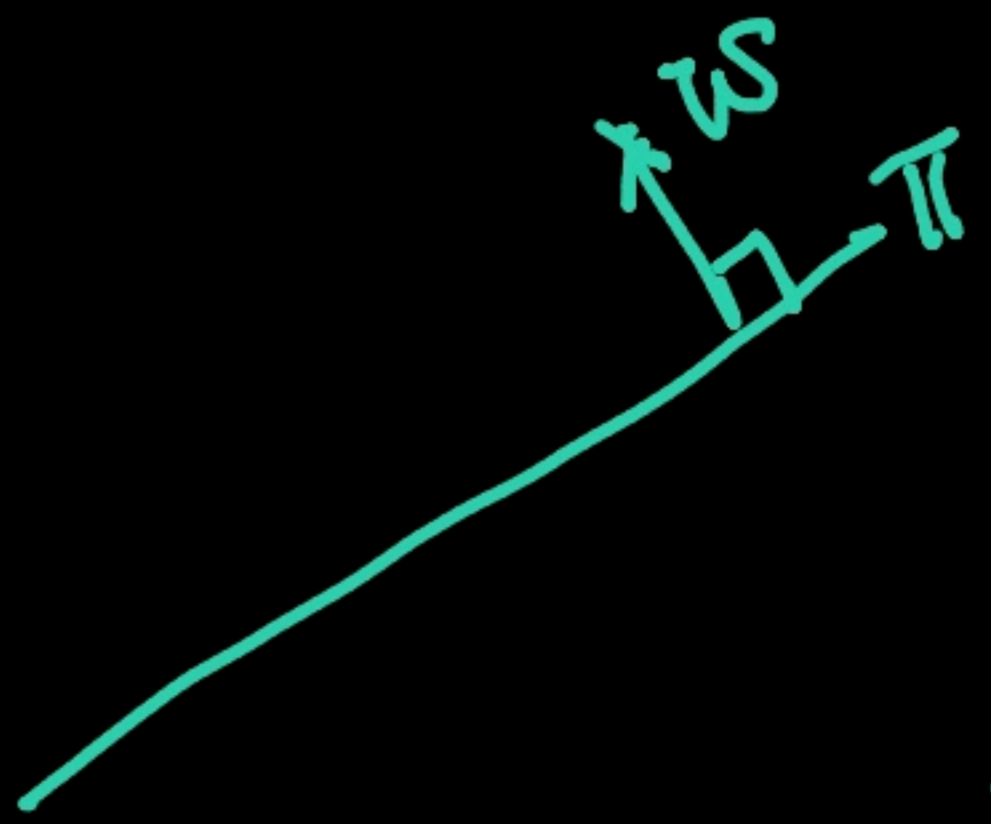
Unconstrained

$$\min_{w,b} \mathcal{L} = -\sum_{i=1}^{n} \frac{y_i(w^T x_i + b)}{\|w\|}$$

hassle to compute gradients

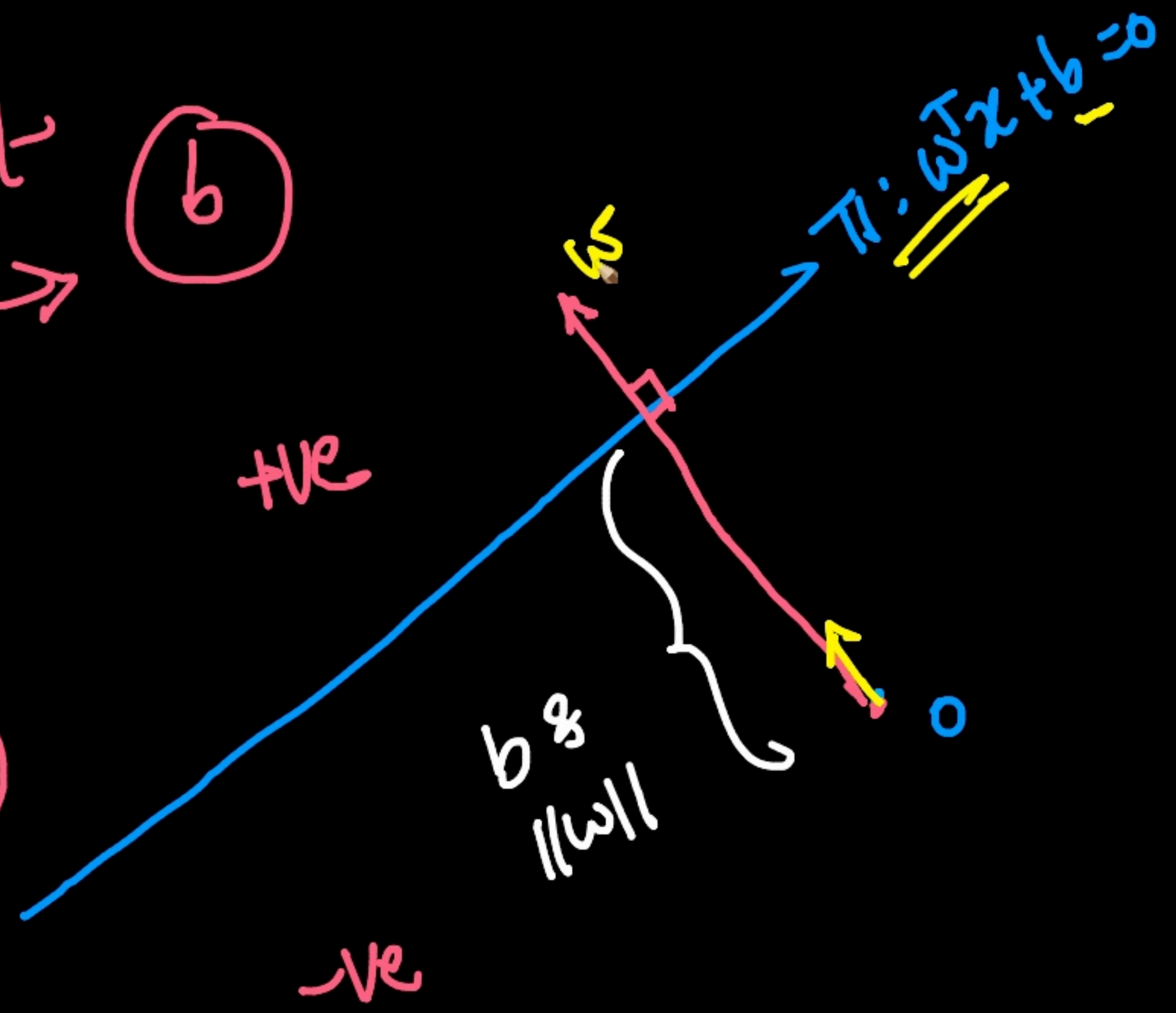
Gradient-descent: $\frac{\partial \mathcal{L}}{\partial w_j}$, $\frac{\partial \mathcal{L}}{\partial b}$

$j = 1 \rightarrow d$



$\bar{w}$/o constraint → b

$\bar{w}$ constraint → b



force
$\|w\| = 1$

$10(w^T x + b) = 0 \times 10$

ML-Classfn

Constrained optimization:

$$\min_{w,b} \mathcal{L}(w,b) = \underbrace{-\sum_{i=1}^{n} y_i(w^T x_i + b)}_{} \rightarrow f(w,b)$$

such that $\|w\| = 1$

Equality constraint

$$g(w,b): \|w\| - 1 = 0$$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:46

Math:

$$\tilde{x}^*, y^* = \underset{x,y}{\operatorname{argMin}} \; f(\underset{\sim}{x}, \underset{\sim}{y}) \;\; \text{s.t.} \;\; g(x,y) = c$$

$\to$ params

$\to$ equality constained optimization

$\hookrightarrow g(x,y) - c = 0$

$\approx$

$$x^*, y^* = \min \left[f(x,y) + \lambda \left(g(x,y) - c \right) \right] \to \mathcal{L}(x,y,\lambda)$$

$\to$ lagrange multiplier

$\lambda \geq 0$

Unconstraint optimzn problem

5 / 26

INTUITION

$$\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda\,(g(x,y) - c)$$

$\lambda \geq 0$

at x^*, y^*, λ^*

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial y} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

$$0 + 1 \cdot g(x,y) - c = 0$$

$$\boxed{g(x,y) = c}$$

NOT-A-PROOF

Linear Regression:

$$\min_{w,b} \sum_{i=1}^{n} \left(y_i - (w^T x_i + b)\right)^2$$

$$\downarrow$$

I donot need to convert this to a constrained setup.

(Q)

$$\min_{w,b} \overbrace{-\sum_{i=1}^{n} y_i(w^T x_i + b)}^{f(w,b)}$$

$f(w,b)$

$d+1$ -params

$$\text{s.t} \quad \underbrace{\|w\| = 1}$$

$$0 \times b + \underbrace{\|w\| - 1}_{g(w,b)} = 0$$

d-params

ML - Module → overfitting & underfitting

→ SVM (inequality constraints)

$$\min_{w,b,\lambda} -\sum_{i=1}^{n} y_i(w^T x_i + b) + \lambda(\|w\| - 1)$$

Dimensionality Reduction

n data points : $x_i \in \mathbb{R}^d$

$$X = \begin{array}{c|ccccc|}
 & f_1 & f_2 & \cdots & f_j \cdots & f_d \\
\hline
1 & & & & & \\
2 & & & & & \\
3 & & & & & \\
\vdots & & & & & \\
i & \longleftarrow & & x_i^T & & \longrightarrow \\
\vdots & & & & & \\
n & & & & & \\
\hline
\end{array}_{\,n \times d}$$

x_i : column-vector

NOT-A-CLASSIFICATION OR REGRESSION

no-class labels (y_i's)

obj: visualize d-dim data of n-points

"100"

→ pair plot : f_1 & f_2 f_2 & f_3

f_1 & f_3 ⋮

f_1 & f_4

⋮

f_1 & f_d

(too many)

DIM RED { d-dim → 2 or 3 dim

↓

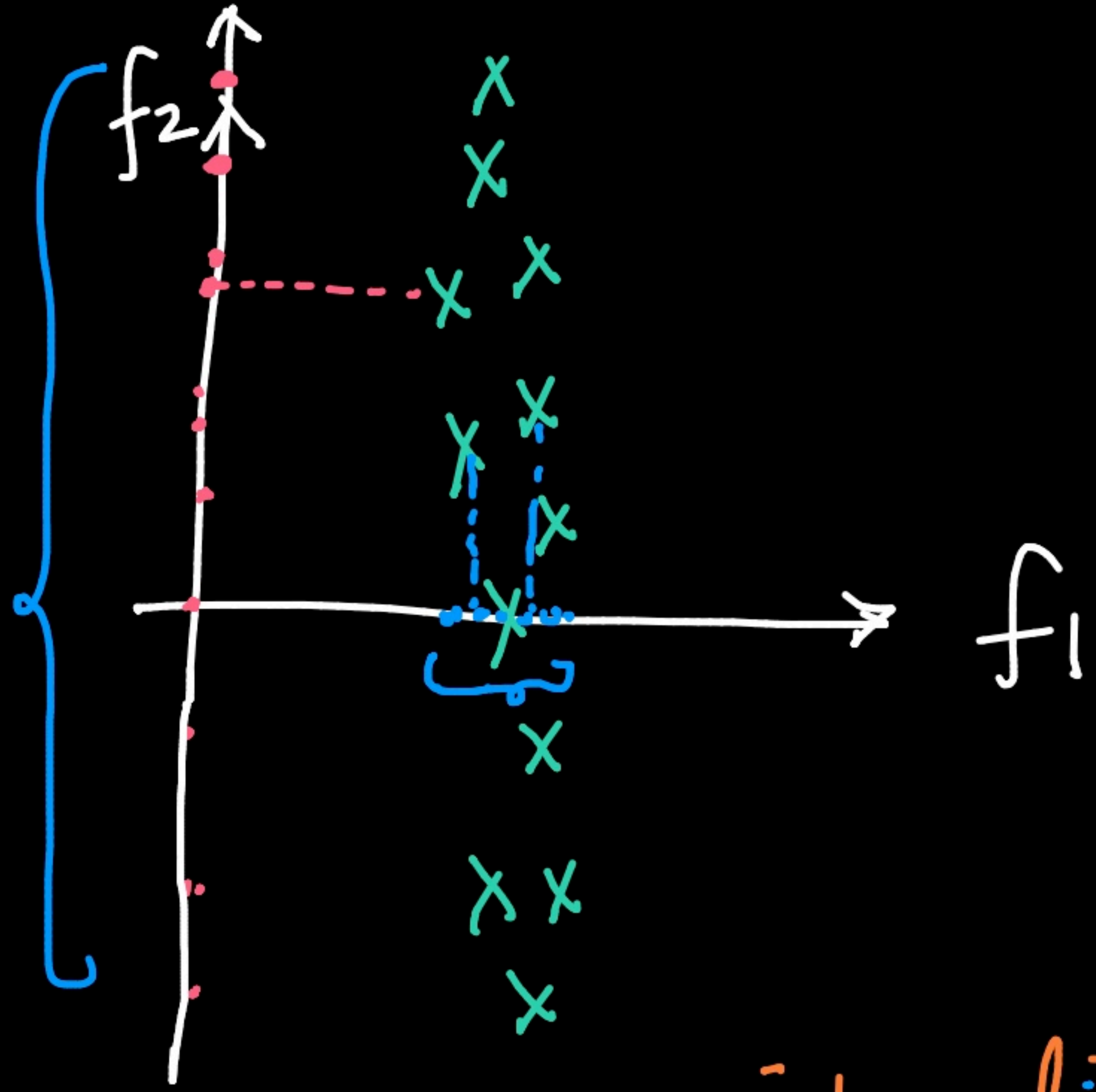
visualize (scatterplot)

INTUITION

$2D \rightarrow 1D$

ignore f_1

& retain f_2 ✓



information

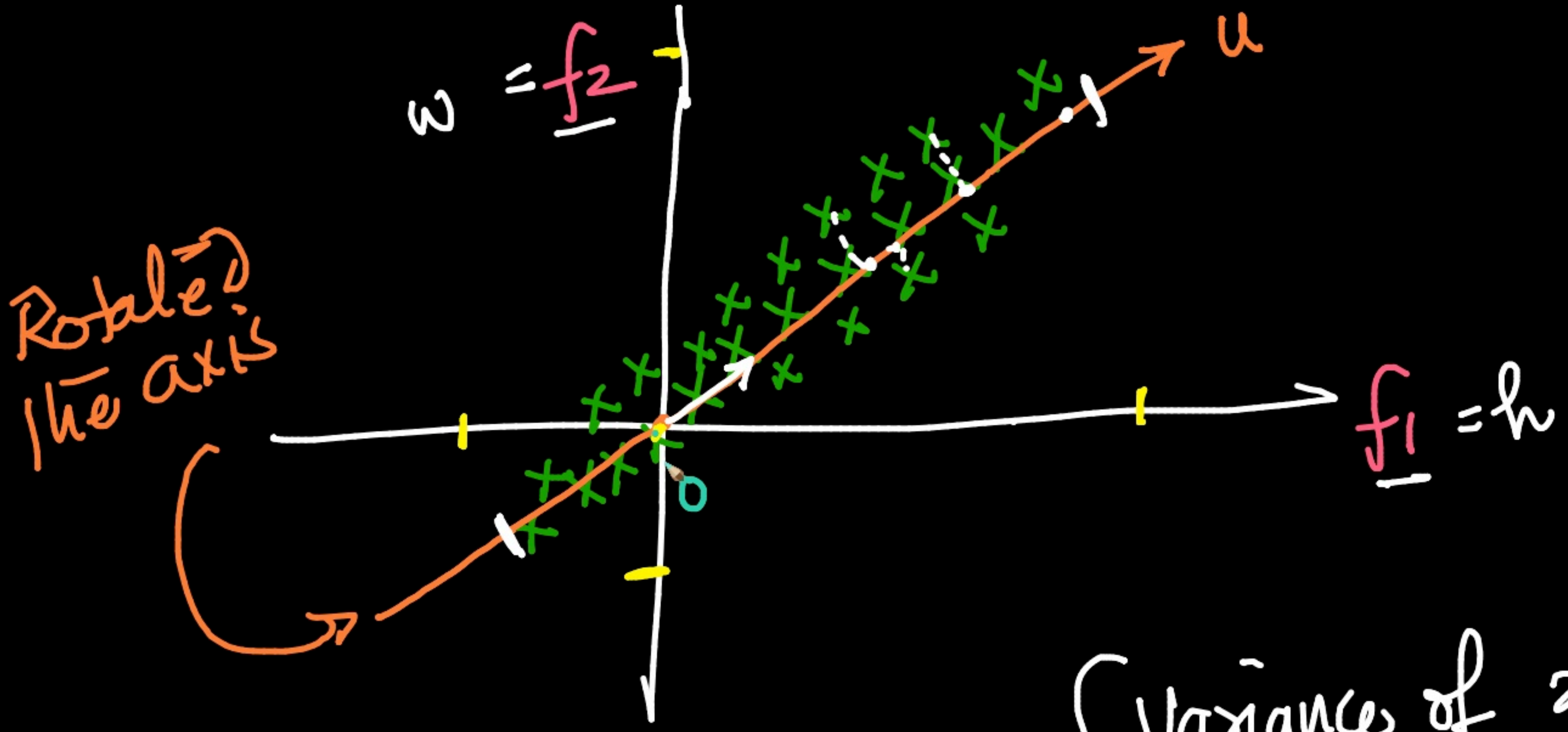
↓

spread / variance

DIM Red

find a vector (u) s.t the variance of projected (on u) points is maximal

$$2D \xrightarrow[\text{RED}]{\text{DIM}} 1D$$



$$u : \alpha f_1 + \beta f_2$$

Variance of x_i projected onto u is maximal

DIM-Red

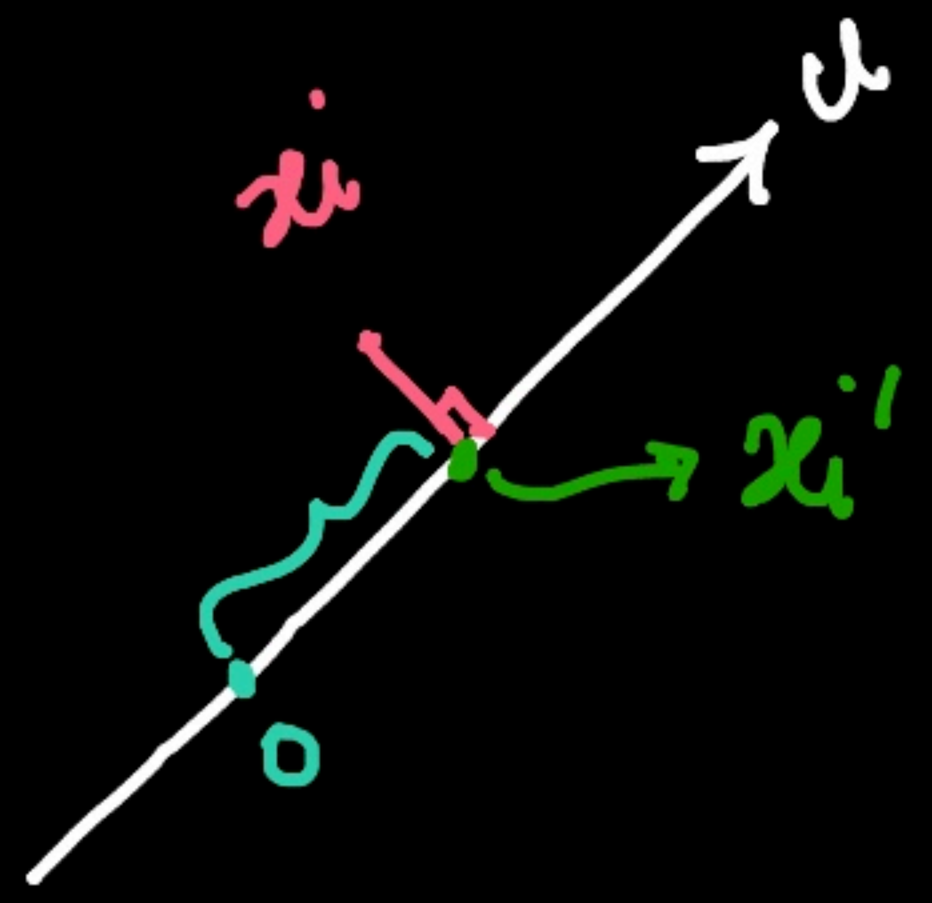
$\|u\| = 1$

Find $\vec{u}$ s.t when x_i's are projected onto u, the variance is maximal

x_i' = projection of x_i onto u

✓ scalar or ~~vector~~

$$x_i' = \frac{x_i \cdot u}{\cancel{\|u\|}} \quad \text{s.t } \|u\| = 1$$



"how far away is the projection from origin in the direction of u"

$$d \xrightarrow[\text{Red}]{\text{DIM}} 1$$

$$d \rightarrow 2$$

$$\max_{u} \ \text{Variance}\{x_i's\}$$

$$\text{s.t.} \quad \|u\| = 1$$

$$x_i' = u^T x_i$$

$$\text{Variance}\{x_i'\}_{i=1}^{n}$$

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

mean of x_i's

$$\frac{1}{n}\sum_{i=1}^{n}\left(u^T x_i - \bar{x}'\right)^2$$

$\bar{x}'$ = avg/mean of all x_i'

$$\frac{1}{n}\sum_{i=1}^{n} x_i' = \frac{1}{n}\sum_{i=1}^{n} u^T x_i$$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:49

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_j \\ \vdots \\ u_d \end{bmatrix}$$

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

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

Gradient Descent ✓

DETOUR

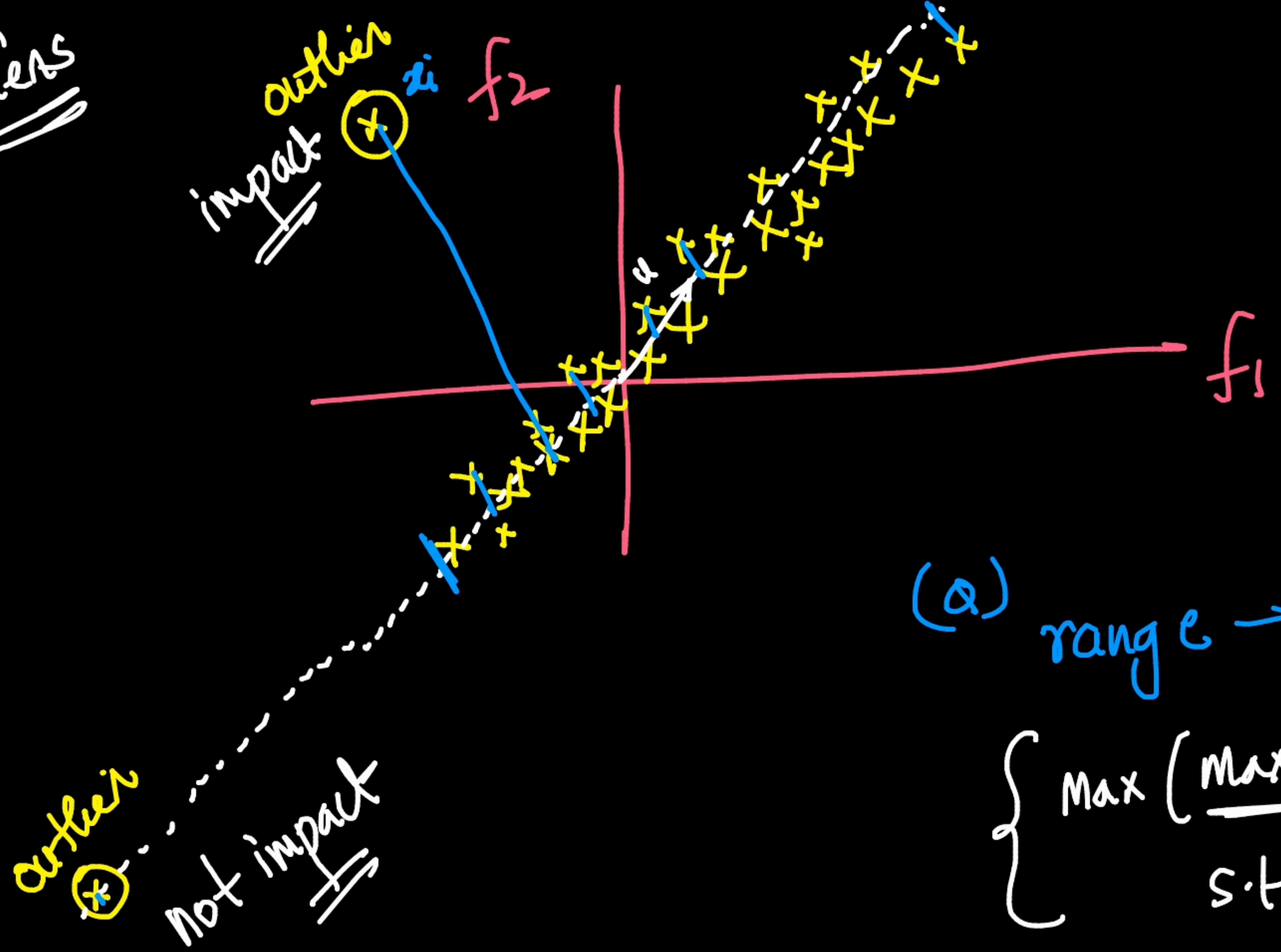
without using GD (closed-form) → (Matrix Theory) → eigen val & eigen vec

params → d-params + 1 (λ)

18 / 26

Outliers

PCA



(Q) range → max − min

$$\begin{cases} \max\left(\max x_i' - \min x_i'\right) \\ \text{s.t.} - \end{cases}$$

PCA

$\Rightarrow$

$\downarrow$

Variance-max

PCA

(Dim-Red)

find u s.t the variance of x_i's projected onto u is maximal

$\downarrow$

$opt_{var\text{-}max}$

$\cong$

find u s.t the sum of dist of each x_i to u is minimal

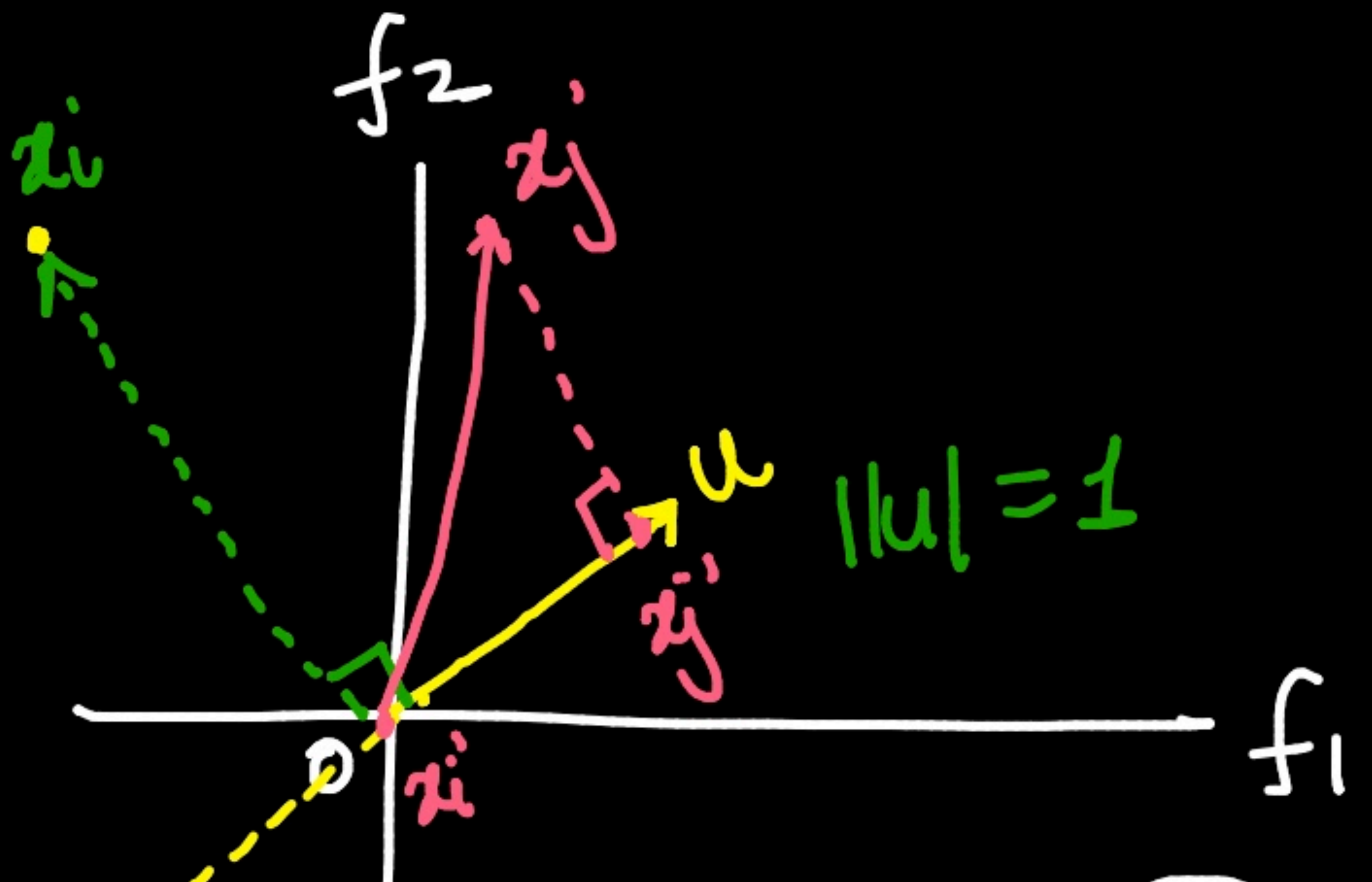
||

dist-min

$\cong$

opt_{dist}

(8)

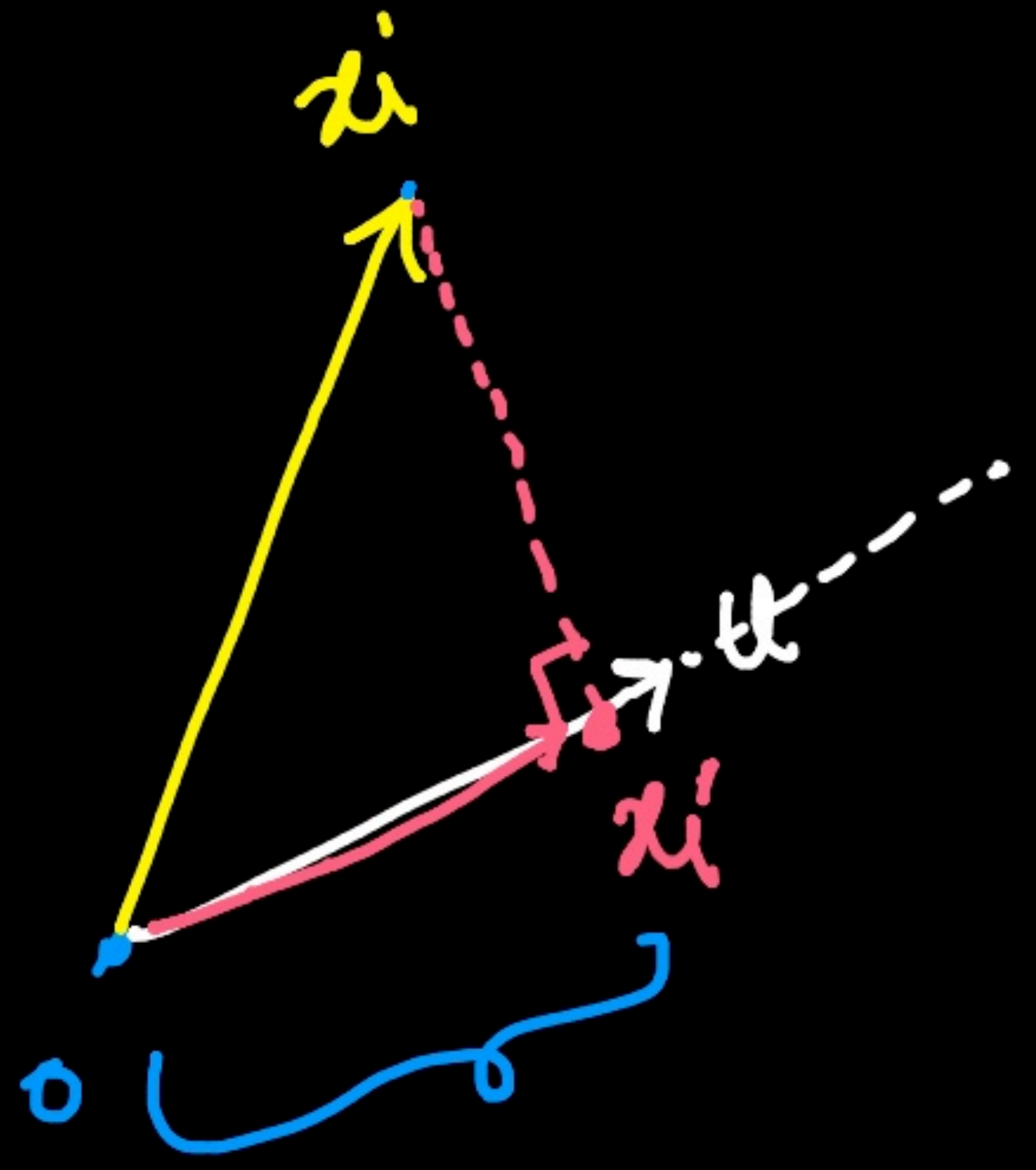


$\|u\| = 1$

$x_i' = u^T x_i = 0$

- Geometrically (2D)
- Algebrically (LA & Alg),
- optimization

Q



$$\text{proj}_u \, x_i = x_i'$$

$$= \frac{u \cdot x_i}{\|u\|} \hat{u}$$

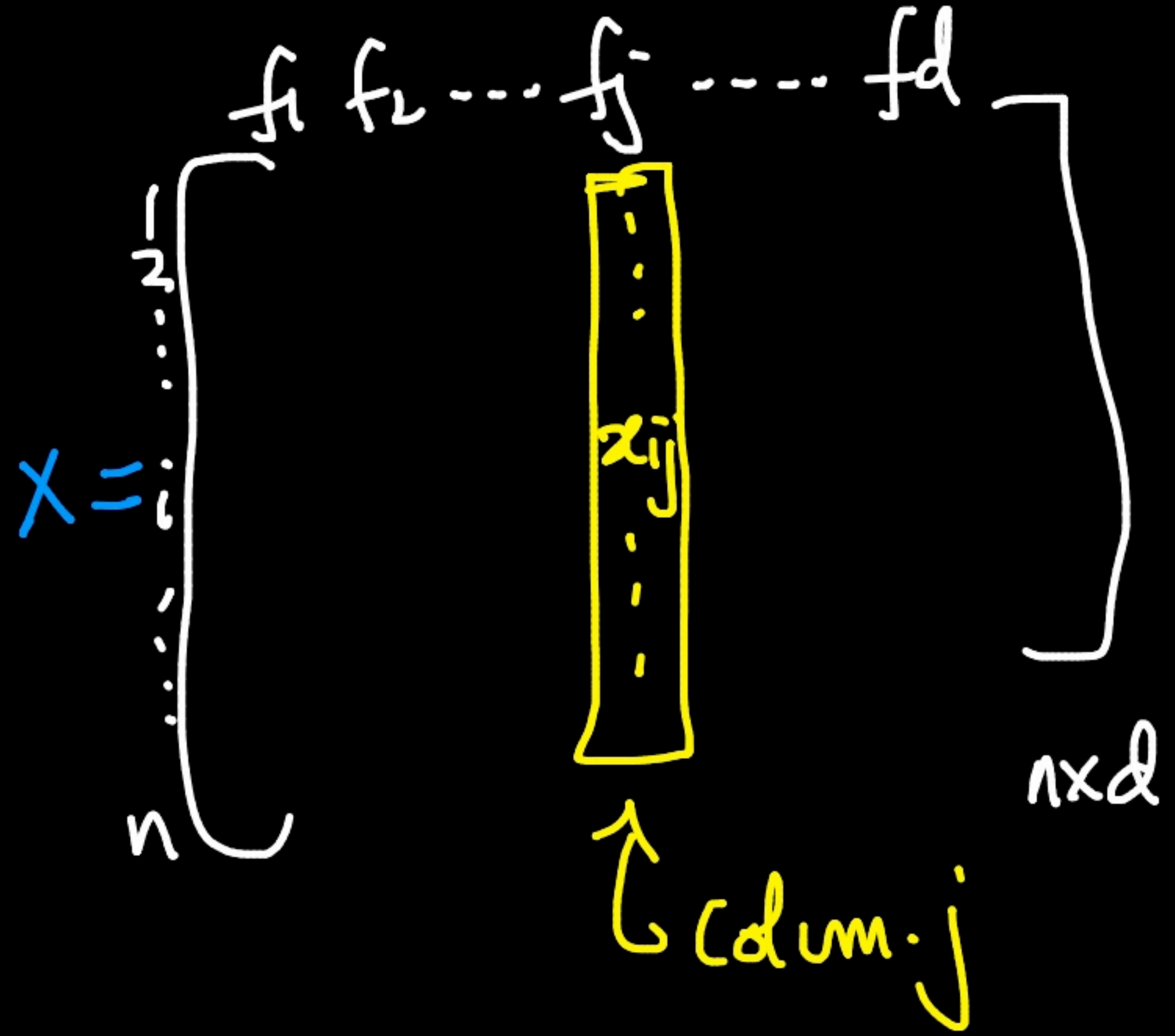
DETOUR

PCA - optimization problem $\xrightarrow{\text{Math}}$ Matrix optimzn problem

✓ Column standardized data

↳ pre-processing for any optimization (MUST)

[Mean-centering
& Variance-scaling]



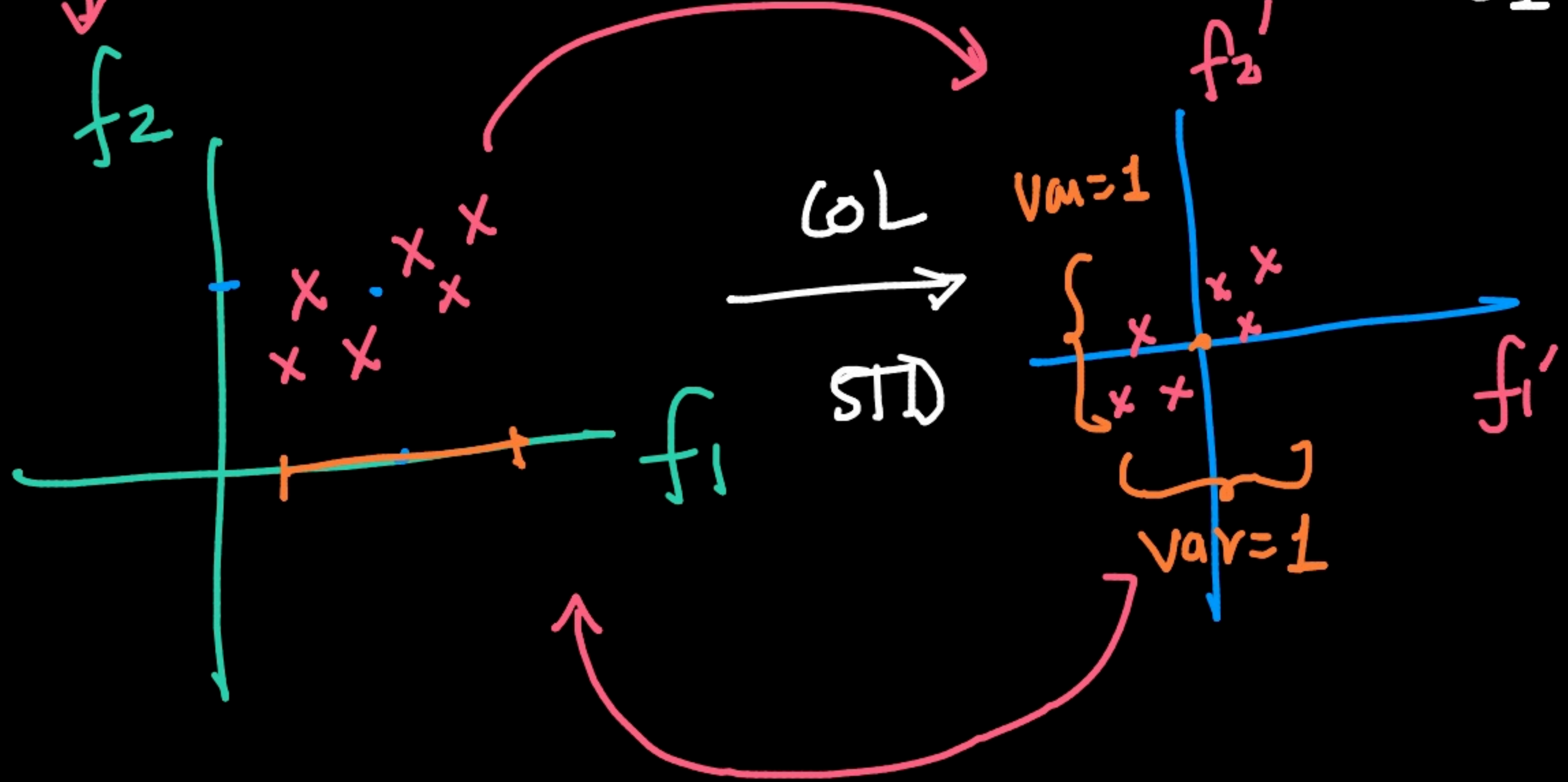
INTUITION / GEOM

$X \xrightarrow{\text{col-std}} X'$

given

X' → mean of each feature = 0 (Moving)

X' → variance of each feature = 1



next-class

Mean-Centering:

Given: $x_1, x_2 \ldots x_i, \ldots x_n$

$\forall i = 1 \rightarrow n$

① mean-vec

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\begin{bmatrix} \text{mean-}f_1 \\ \text{mean-}f_2 \\ \vdots \\ \text{mean-}f_d \end{bmatrix}$$

② $x_i' = x_i - \bar{x}$

mean $x_i' = \vec{0}$

= TO-BE-CONTINUED =