Principle Component Analysis (PCA)

Why? dimensionality reduction

n dim $\rightarrow$ d'-dim

$x_i \in \mathbb{R}^d$ $\quad d' < d$

① MNIST $\rightarrow$ 784 dim $\rightarrow$ 2dim (visualize)

② d-dim $\rightarrow$ d' dim $(d' < d)$

Geometric interpretation of PCA

2d → 1d

f_1 = blackness of hair $\in \mathbb{R}$

d dim $\rightarrow$ d' dim

$d' < d$

2dim $\rightarrow$ 1 dim

$d \quad\quad d'$

Indian ← (grey, black)

$$X = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} f_1 \quad f_2 \\ \left[\quad \longrightarrow \quad \right] \end{array} \; ; \quad X' = \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} \left[f_2 \right]$$

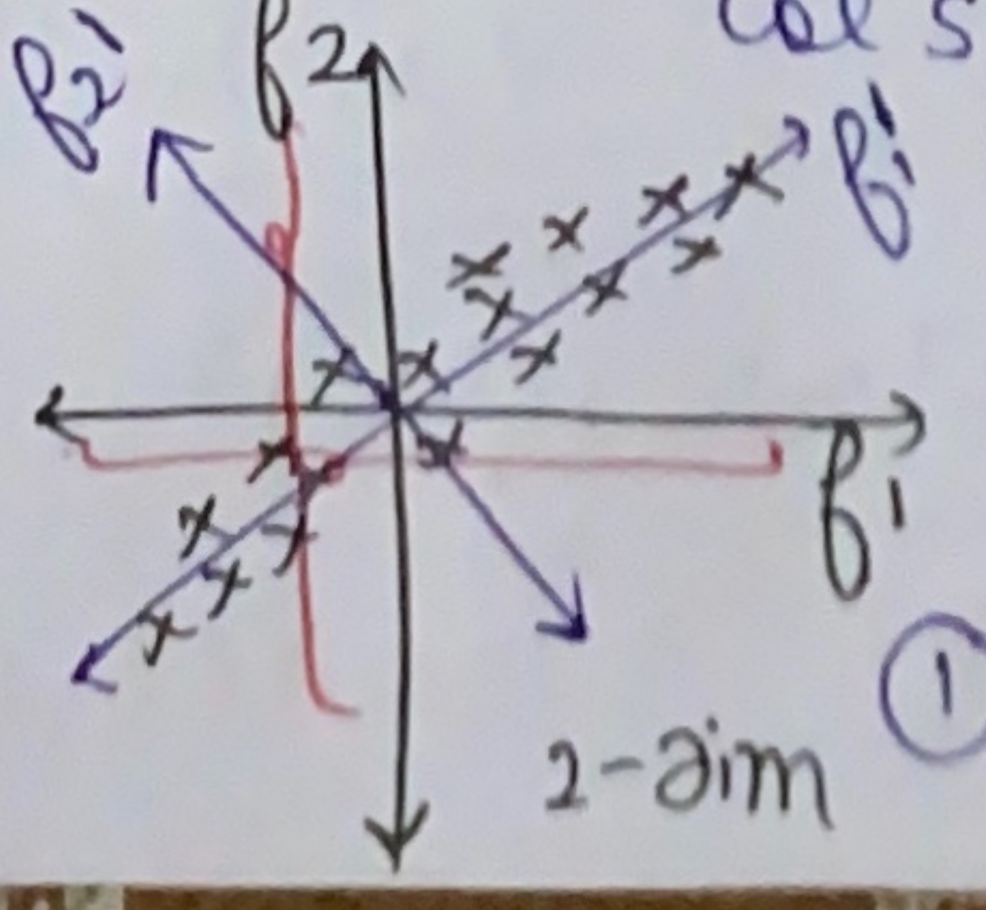
variance = spread = variability

preserving the direction with maximum spread
↳ more information

X = 2-dim dataset

col standarized Mean $\{f_1\}$ = mean $\{f_2\}$ = 0

var $\{f_1\}$ = var $\{f_2\}$ = 1



$\longrightarrow$

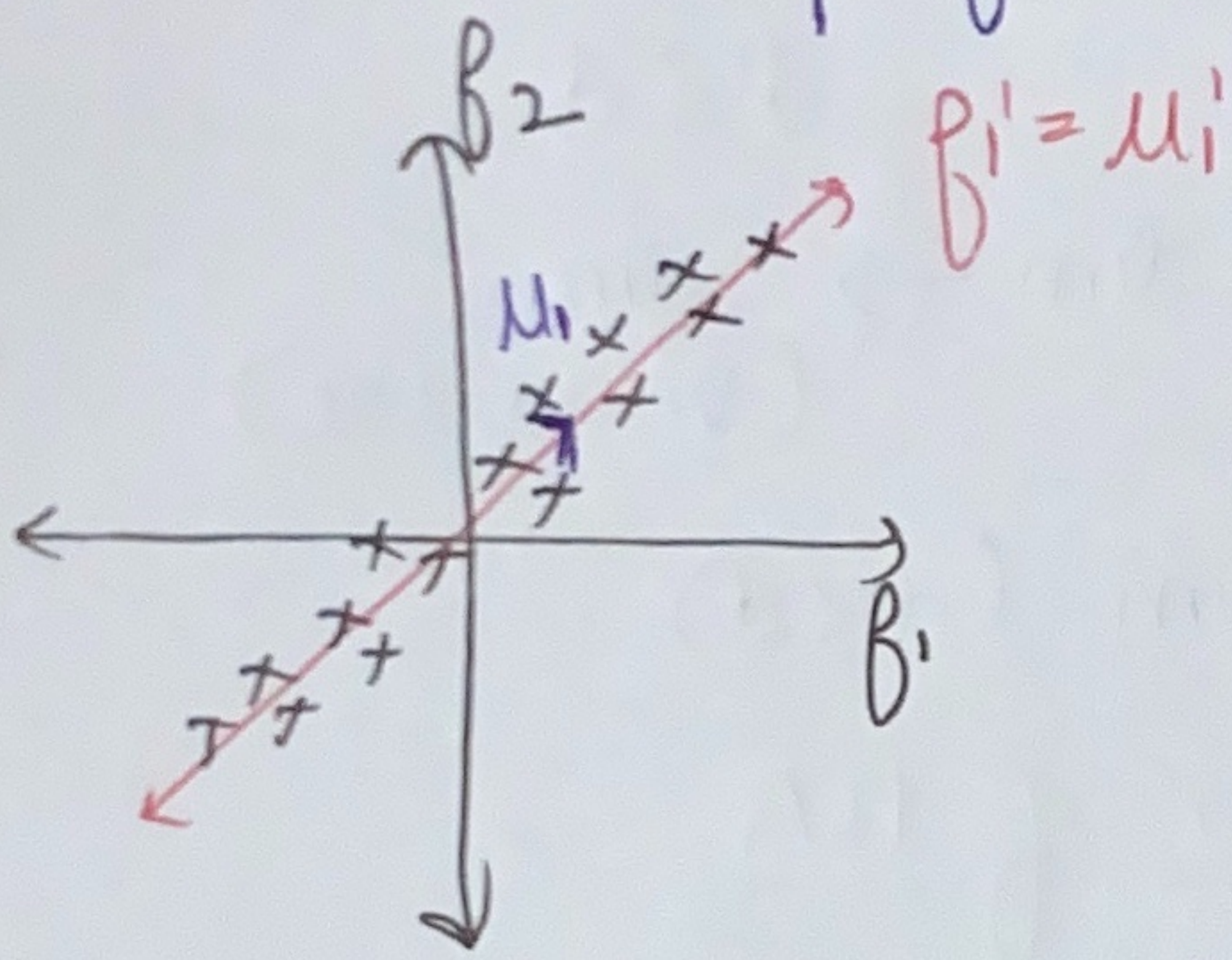
① $f_1' \perp f_2'$ (spread on f_2' << spread on f_1') st f_1' max spread

② drop f_2'

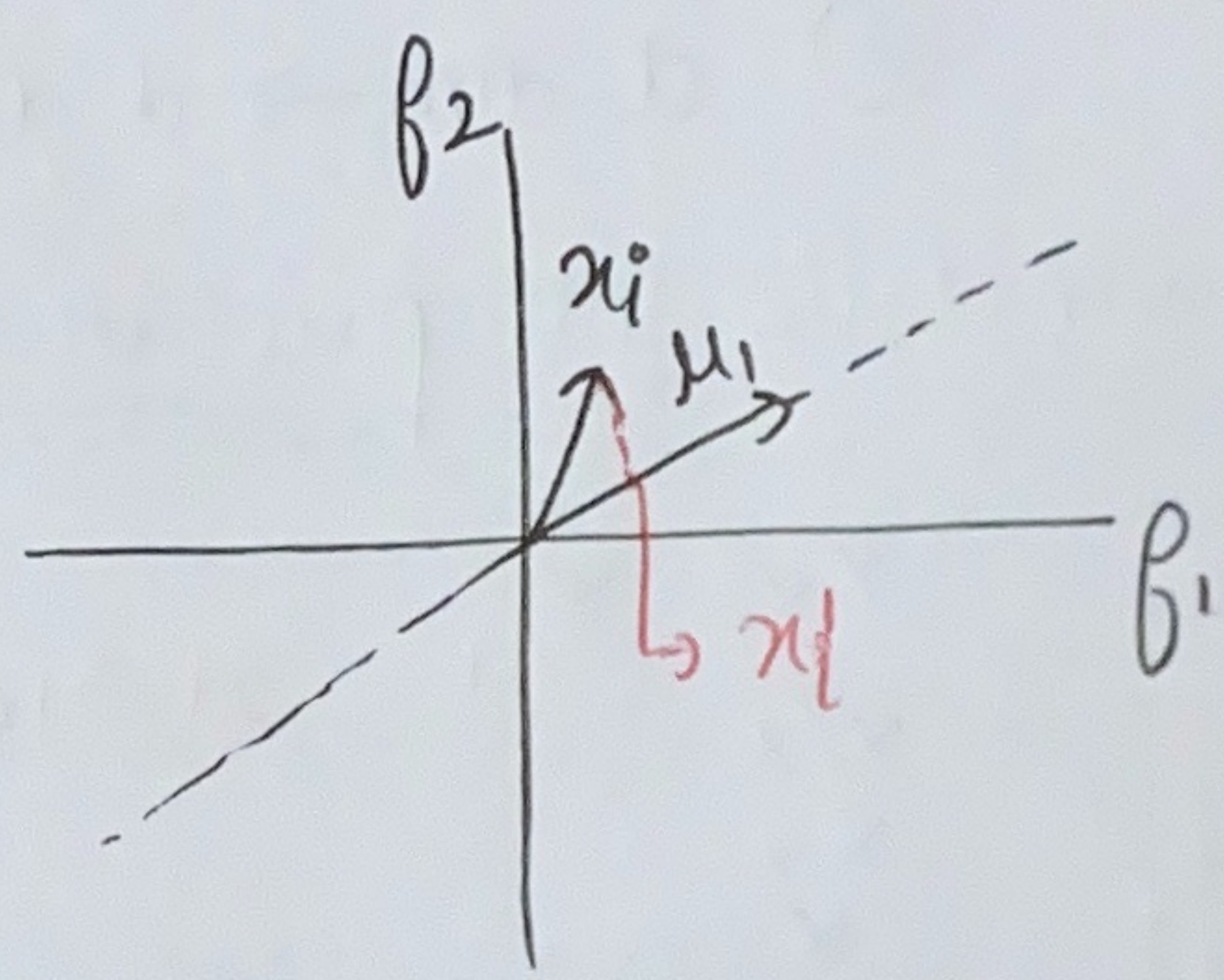
③ project x_is onto f_1' then 2d → 1D

① We want to find a direction f_1' s.t. variance of x_i's projected onto f_1' is maximum

* rotating my axis to find f_1' with max-variance and drop f_2'



u_1 = unit vector : direction
$\|u_1\| = 1$



$$x_i' = \text{proj}_{u_1} x_i = \frac{u_1 \cdot x_i}{\|u_1\|^2} = \boxed{u_1^T \cdot x_i}$$

x_i' = projection of x_i on u_1

$$\mathcal{D} = \{x_i\}_{i=1}^n \to \mathcal{D}' = \{x_i'\}_{i=1}^n$$

$$\boxed{x_i' = u_1^T x_i} \qquad \boxed{\bar{x}_i' = u_1^T \bar{x}}$$

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

(*) find u_1 such that the variance $\underbrace{\{\text{proj}_{u_1} x_i\}_{i=1}^n}$ = max

$$\text{var}\{u_1^T x_i\}_{i=1}^n = \underbrace{\frac{1}{n}\sum_{i=1}^n}_{\text{avg}} (\underbrace{u_1^T x_i}_{x_i'} - \underbrace{u_1^T \bar{x}}_{\text{mean}\{x_i'\}_{i=1}^n})^2$$

$$\text{scalar} = \underset{(1\times n)}{(u_1)^T} \underset{(n\times 1)}{x_i}$$

X: colm standarized

$\bar{x} = [0, 0 - - - 0]$

$$\boxed{var\{x_i'\}_{i=1}^{n} = \frac{1}{n}\sum_{i=1}^{n}(u_1^T x_i)^2}$$

$$\boxed{\max_{u_1} \frac{1}{n}\sum_{i=1}^{n}(u_1^T x_i)^2}$$ data matrix

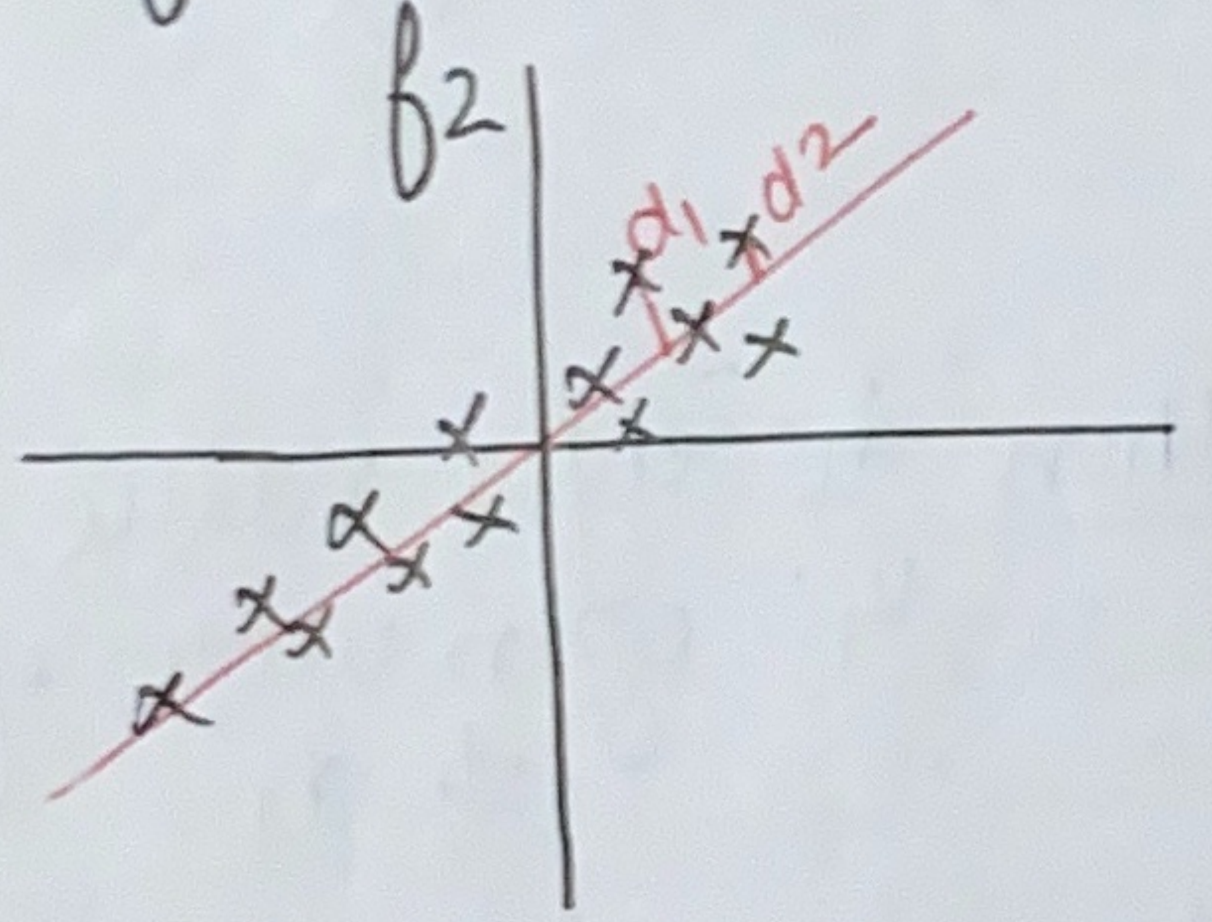
s.t $u_1^T u = 1 = ||u||^2$

↓

u_1 is unit vector

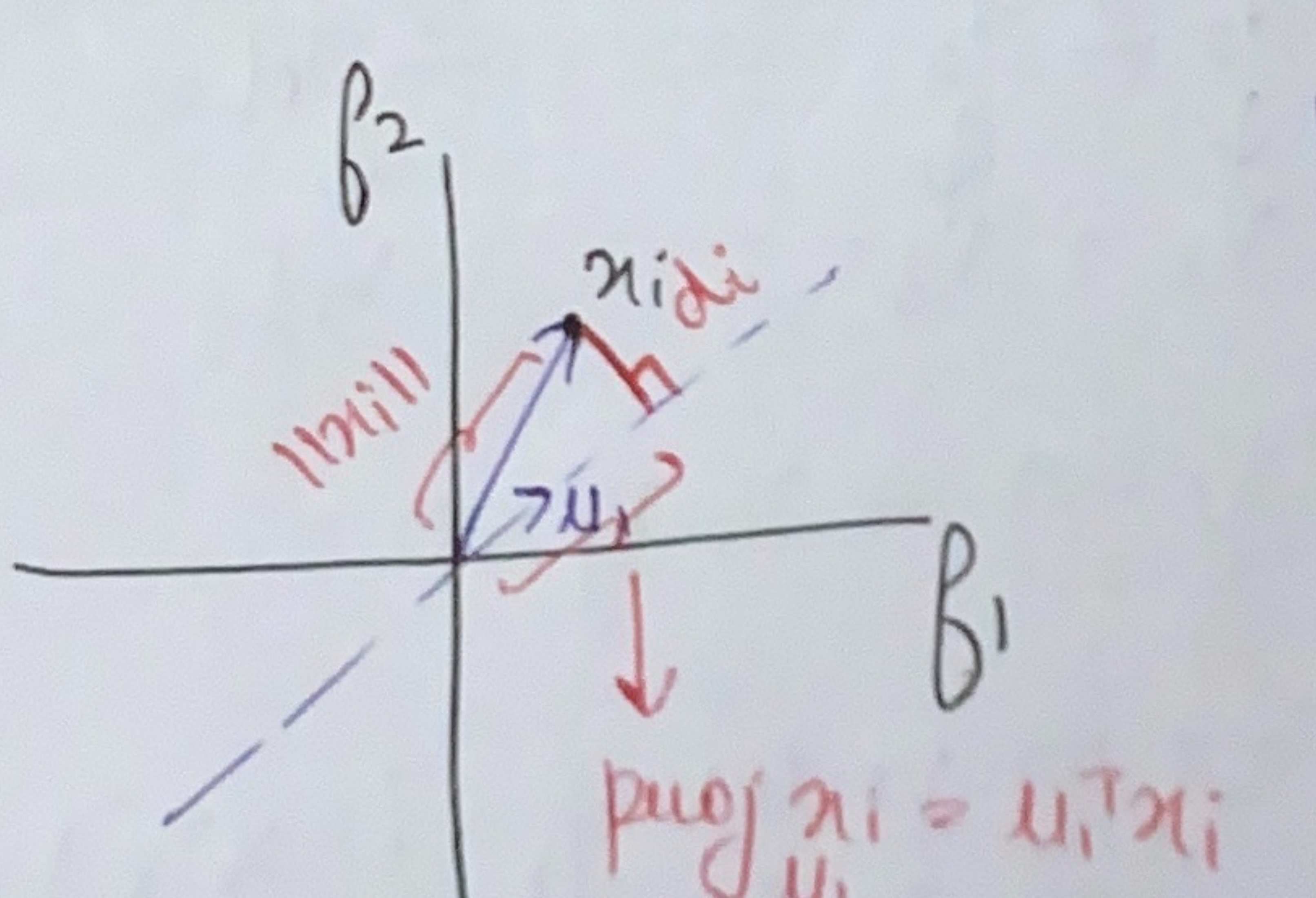
Alternative formulation of PCA: dist. minimization

→ find u_1 which maximizes projected variance



$x_i \to d_i$: dist. from x_i to u_1

$$\boxed{\min_{u_1} \sum_{i=1}^{2} d_i^2}$$



u_1: unit vector

$u_1^T u_1 = 1 = ||u_1||^2$

$d_i^2 = ||x_i||^2 - (u_1^T x_i)^2$

$d_i^2 = x_i^T x_i - (u_1^T x_i)^2$

dist min PCA

$$\boxed{\min_{u_1} \sum_{i=1}^{n} \underbrace{\left[(x_i^T x_i) - (u_1^T x_i)^2\right]}_{d_i^2}}$$ s.t $u^T u = 1$

Solution to our optimization problems: λ_1, v_1

$$X = \begin{array}{c|c} & 1\ 2\ 3 \ldots d \\ \begin{matrix}1\\2\\3\\|\\|\\n\end{matrix} & \begin{bmatrix} \quad & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array}$$

$X \rightarrow$ col std

covariance matrix X $S = \; = X^T X$

S: $d \times d$ sq. symm matrix; X^T: $d \times n$; X: $n \times d$

eigen values $(\lambda_1, \lambda_2 ---- \lambda_d)$

eigen vector $(v_1, v_2 --- v_d)$

$S_{d \times d} =$

max eigen value: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 ---- \geq \lambda_d$

(eigenvalues) of $S = \lambda_1, \lambda_2 --- \lambda_d$

(eigen vector) of $S = v_1, v_2 ---- v_d$

def: If $\lambda_1 v_1 = S_{d \times d} v_1^{d \times 1}$

λ_1: scalar; v_1: $d \times 1$ vector

then λ_1: eigen value of S

v_1: eigen vector of S corr to λ_1

$S_{d \times d}$

$$\begin{matrix} \lambda_1 \geq & \lambda_2 \geq & \lambda_3 & --- & \geq \lambda_d \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ v_1, & v_2, & v_3 & ---- & v_d \end{matrix}$$

$\boxed{v_i \perp v_j}$: $v_i^T v_j = 0 = v_i \cdot v_j = 0$

$u_1 = v_1$ = eigen vector of $S (= X^T X)$ corr to largest eigen value $(= \lambda_1)$

$u_1 \rightarrow$ max variance direction

$X = \begin{bmatrix} & \\ & \end{bmatrix}$

Steps

① Colmn std of X is done

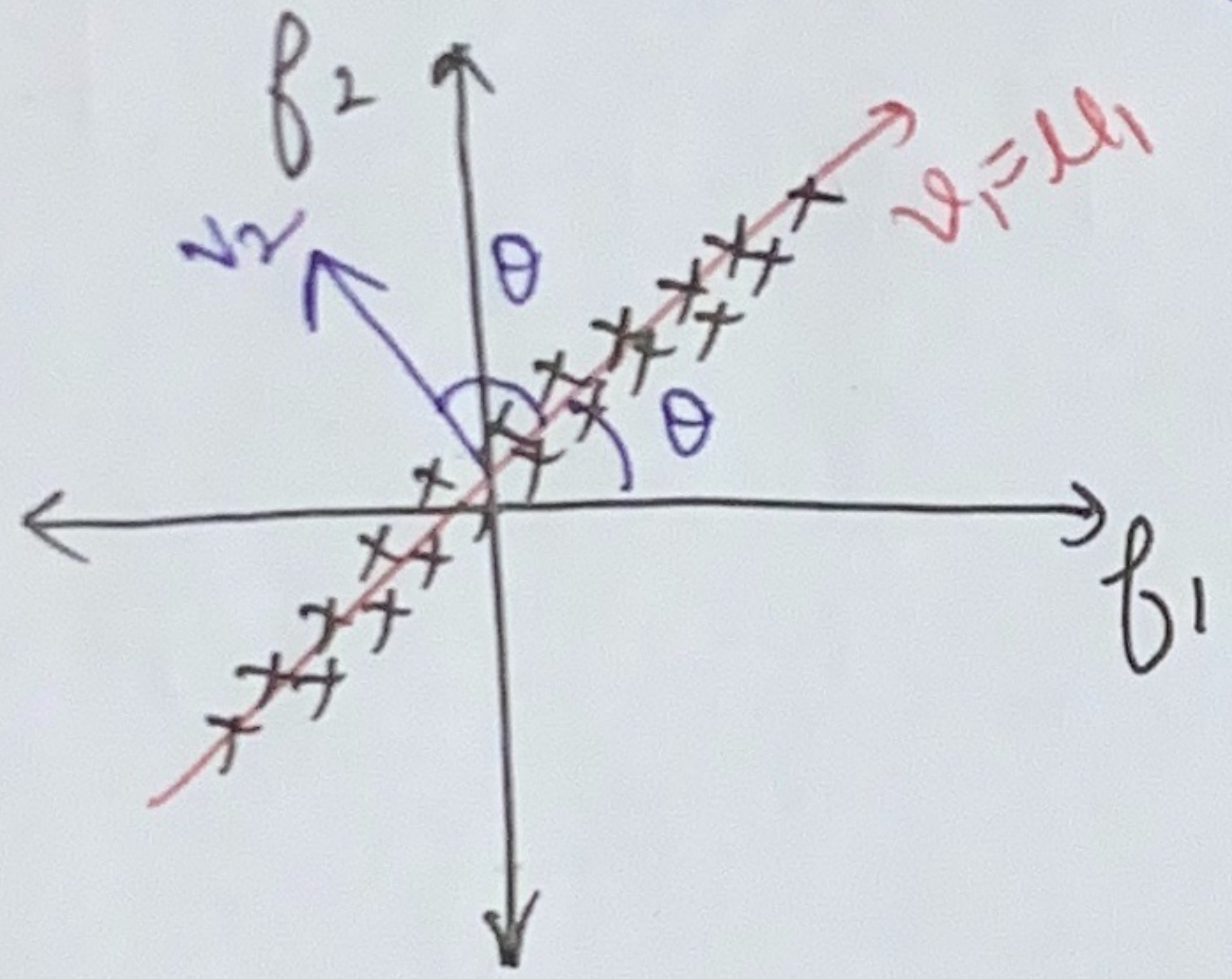
② $S = X^T X$

③ eigen values & vectors of S

$\lambda_1 \geq \lambda_2 --- \lambda_d$

$v_1, v_2 ---- v_d$

④ $u_1 = v_1$



2-dim

$d=2$

$\lambda_1 \geq \lambda_2$

$v_1 \perp v_2$

$x_i \in \mathbb{R}^{10}$ $d=10$

$\lambda_1 \geq \lambda_2 \geq \lambda_3 --- \geq \lambda_{10}$

$v_1, v_2, v_3 ---- v_{10}$ ← least variance in v_{10}

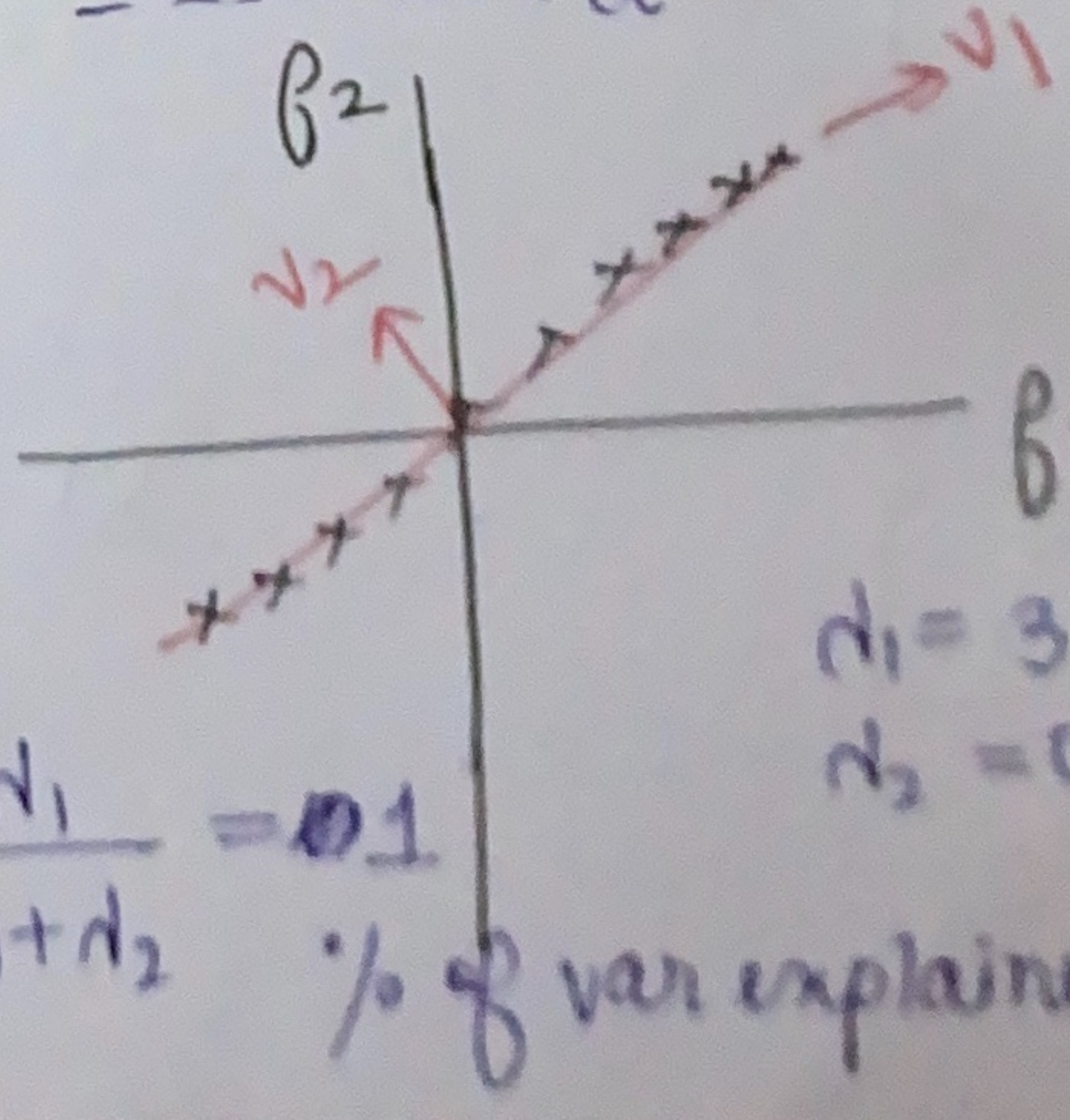
v_1: direction with max var

v_2: 2nd direc with most variance

v_3: 3rd maximal direction

$\lambda_1 \geq \lambda_2 \geq \lambda_3 ---- \geq \lambda_d$

$\sum_{i=1}^{d} \lambda_i =$

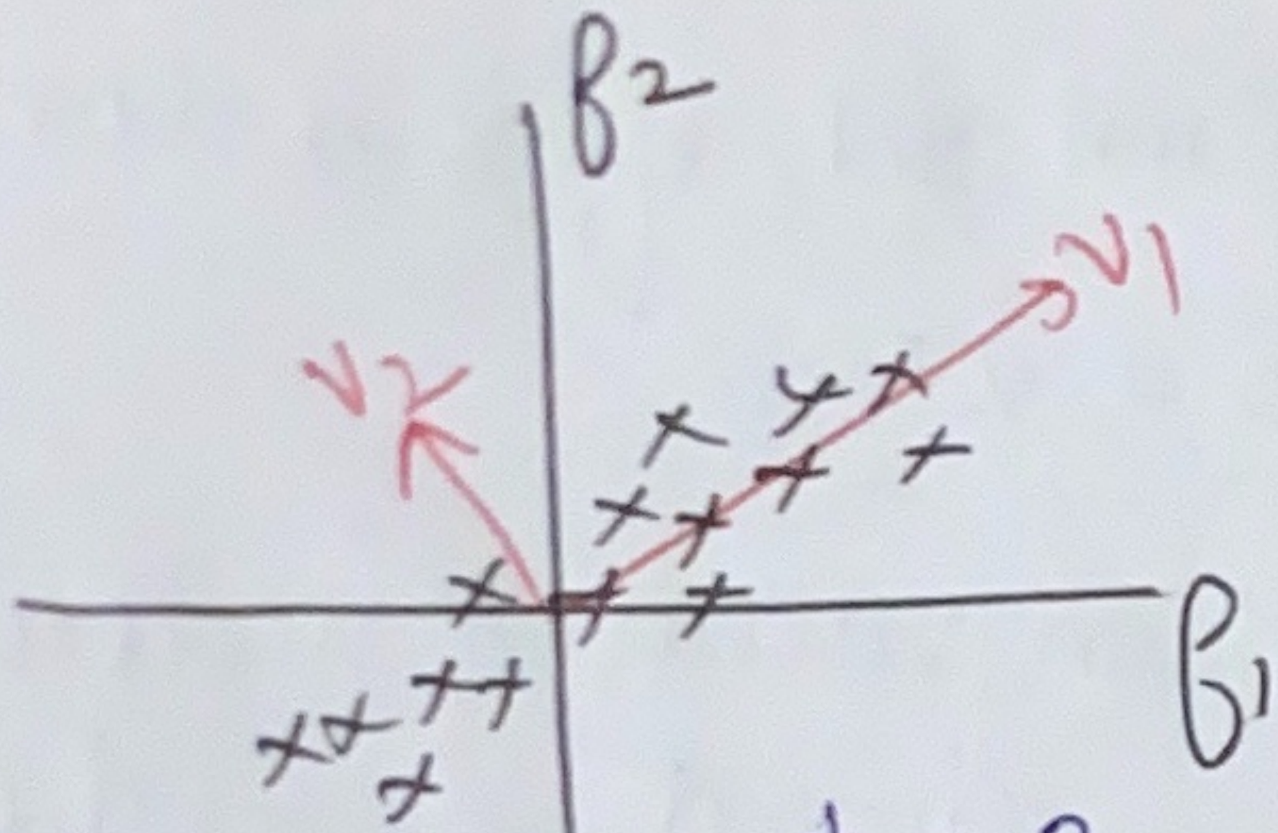


$\lambda_1 = 3$

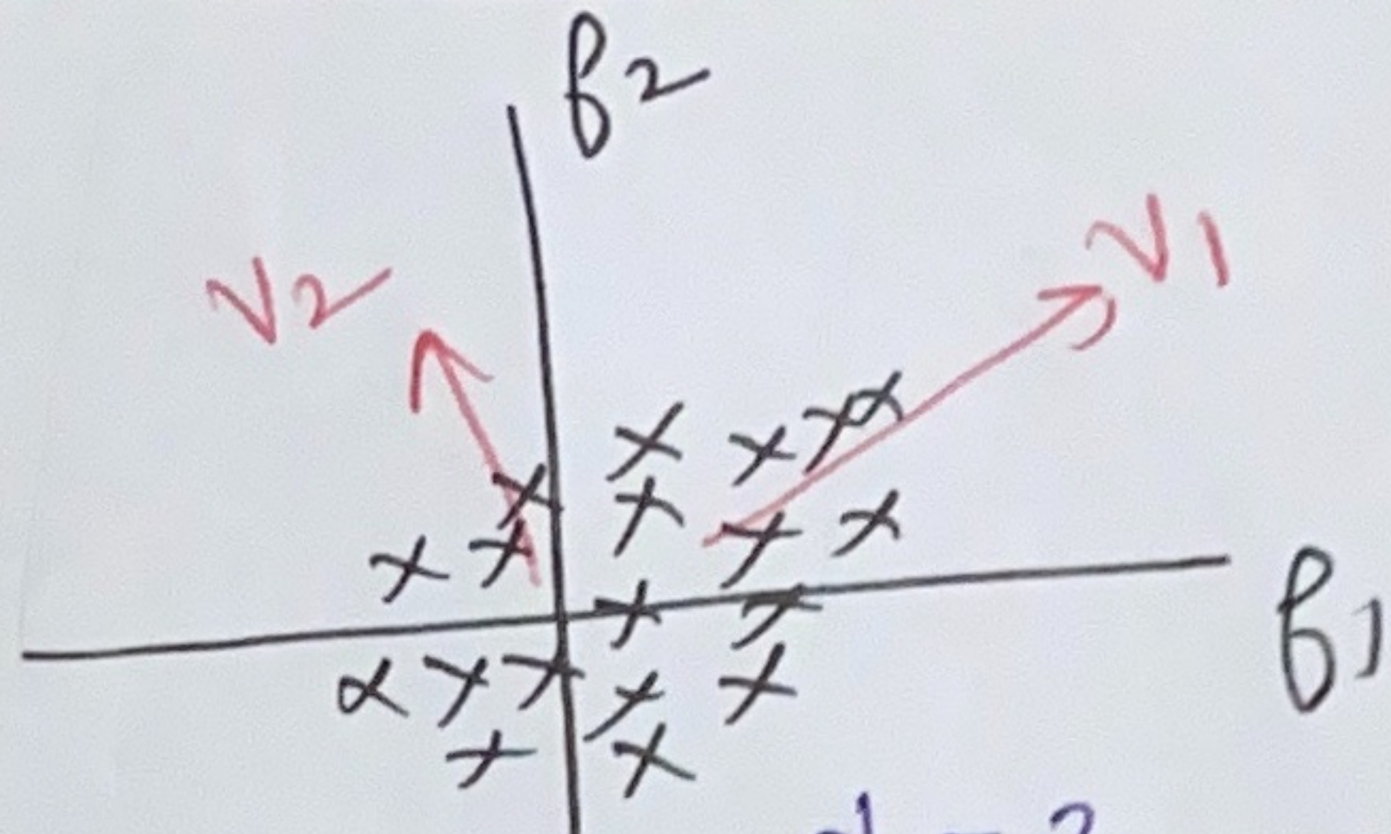
$\lambda_2 = 0$

$\frac{\lambda_1}{\lambda_1 + \lambda_2} = 1$

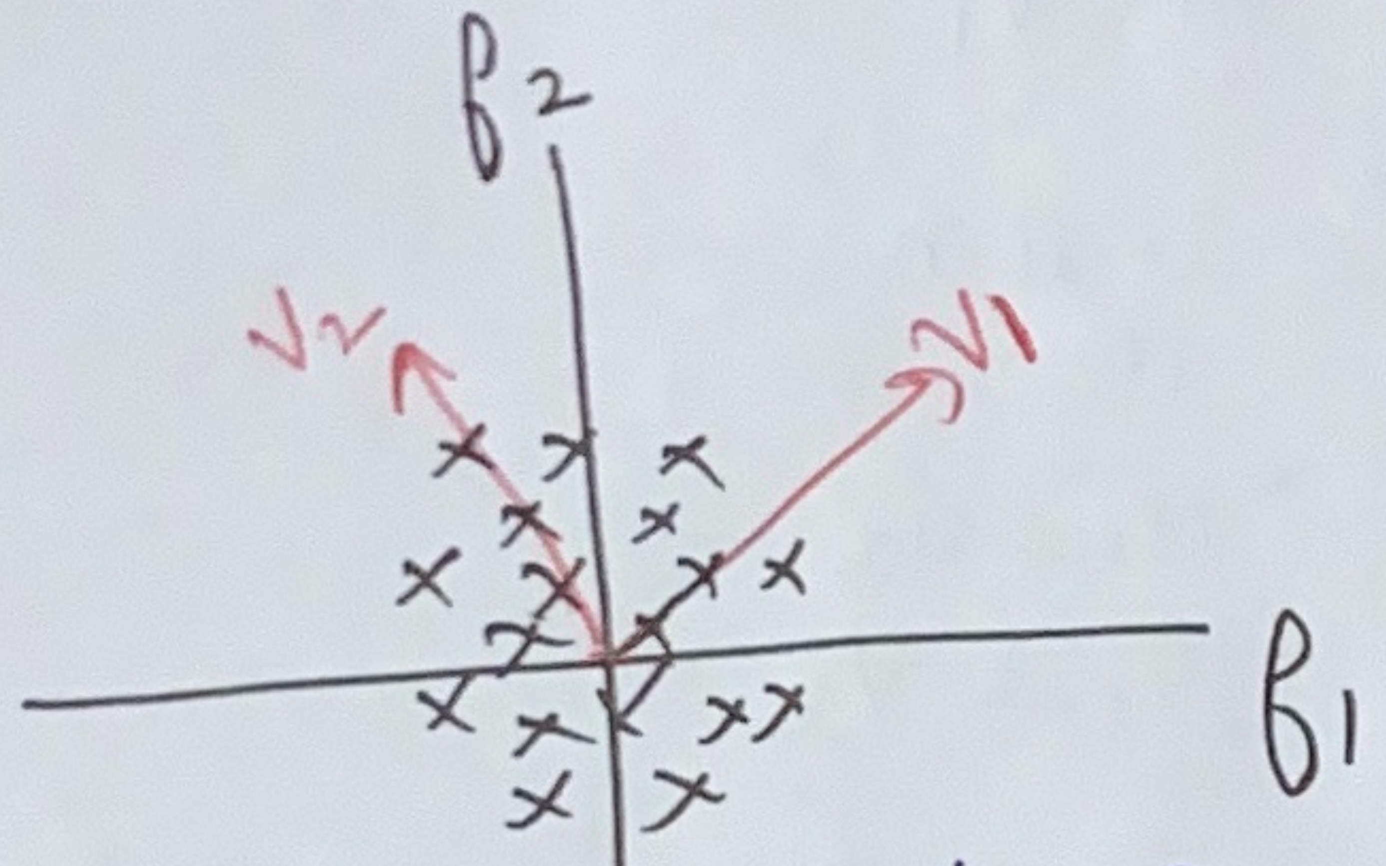
% of var explained



$$\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{3}{4} \qquad \lambda_1 = 3,\ \lambda_2 = 1$$

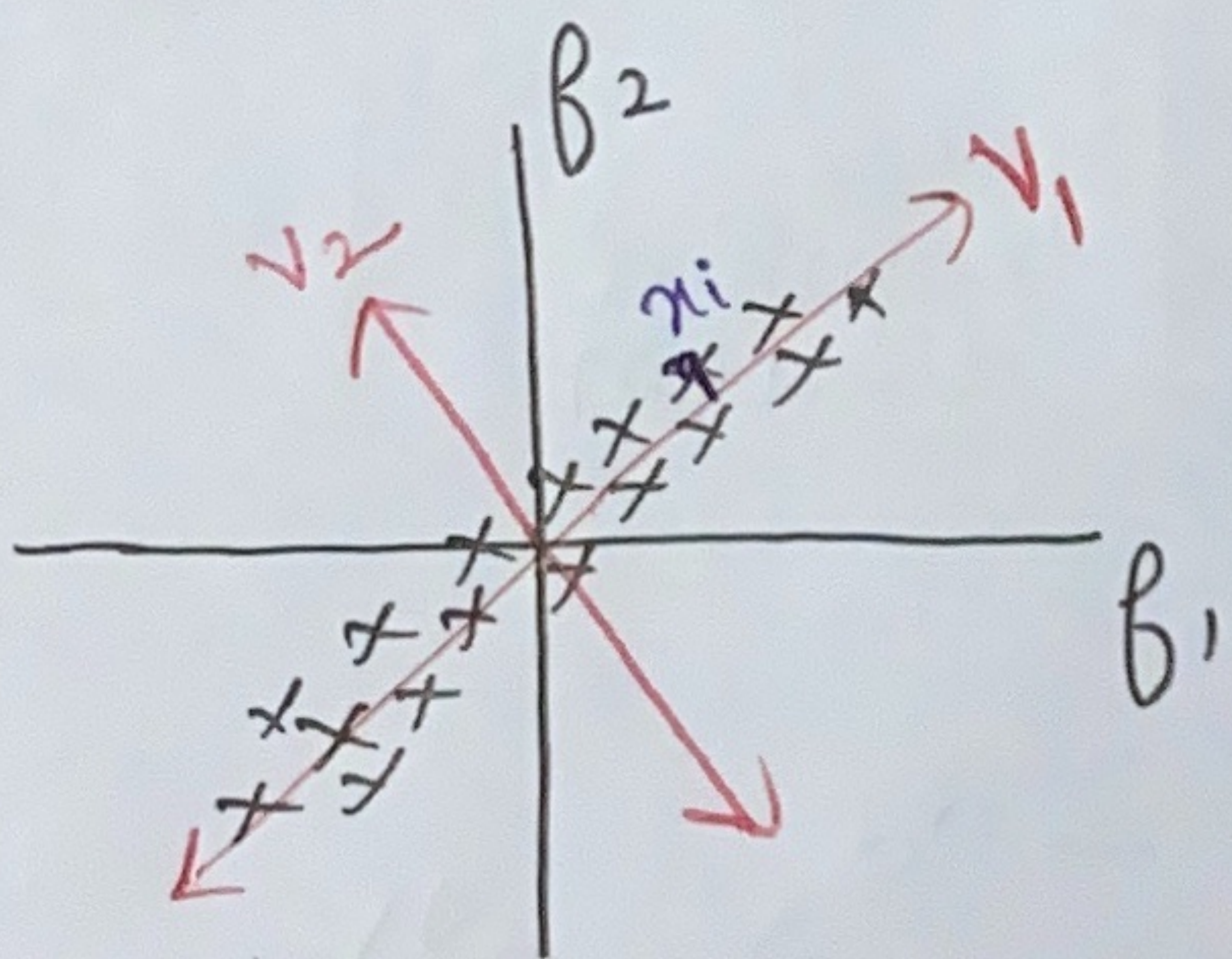


$$\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{3}{5} \qquad \lambda_1 = 3,\ \lambda_2 = 2$$



$$\frac{\lambda_1}{\lambda_1+\lambda_3} = \frac{3}{6} = \frac{1}{2} \qquad \lambda_1 = 3,\ \lambda_2 = 3$$

PCA for dim-reduction



$$X = \begin{array}{c|cc} & f_1 & f_2 \\ \hline 1 & & \\ 2 & & \\ \vdots & \leftarrow x_i^T \rightarrow & \\ n & & \end{array} \quad \text{2-D}$$

max Varia Method (PCA)

$$S = X^T X$$

1-D

$$X' = \begin{array}{c|c} & v_1 \\ \hline 1 & \\ 2 & \\ 3 & \\ \vdots & x_i' \\ n & \end{array}$$

$$\boxed{x_i' = x_i^T v_1}$$

$$X = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} f_1 \; f_2 \; f_3 \; \cdots \; f_{10} \\ \left[\begin{array}{c} \\ \longleftarrow x_i^T \longrightarrow \\ \\ \end{array} \right]_{n \times 10} \end{array} \xrightarrow[\text{(PCA)}]{\text{dim red}} X' = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} \begin{array}{c} v_1 \; v_2 \\ \left[\begin{array}{c} \\ \longleftarrow x_i' \longrightarrow \\ \\ \end{array} \right]_{n \times 2} \end{array}$$

$S = X^T X =$

$\text{eigen}(S) = d_1 \geq d_2 \geq d_3 \cdots \geq d_{10}$

$$\begin{array}{ccccc} \downarrow & \downarrow & \downarrow & \cdots & \downarrow \\ v_1 & v_2 & v_3 & & v_{10} \end{array}$$

$x_i' = [x_i^T v_1, x_i^T v_2]$

v_1, v_2

$$X = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} f_1, f_2 \; \cdots \; f_{100} \\ \left[\begin{array}{c} \\ \\ \\ \end{array} \right]_{n \times 100} \end{array} \xrightarrow{\text{PCA}} X' = \begin{array}{c} v_1 \; v_2 \; \cdots \; v_{50} \\ \left[\begin{array}{c} \\ \\ \\ \end{array} \right]_{n \times 50} \end{array}$$

$\underset{100 \times 100}{S} = X^T X \qquad d_1 \geq d_2 \geq d_3 \cdots \geq d_{100}$

$$\begin{array}{ccccc} \downarrow & & & & \\ v_1 & v_2 & v_3 & \cdots & v_{100} \end{array}$$

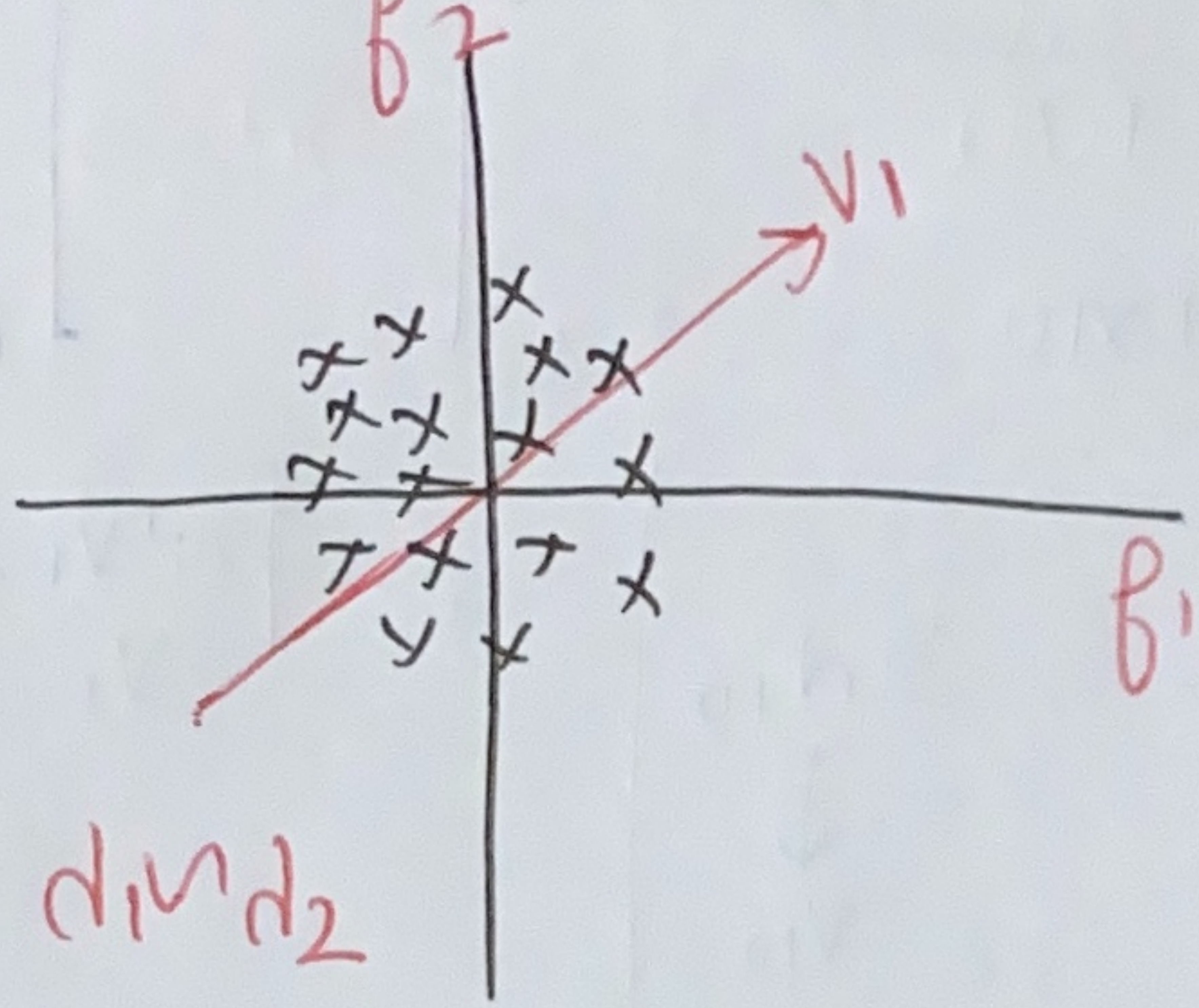
* $x_i \in \mathbb{R}^{100}$; $x_i \in \mathbb{R}^{d'}$ $[d' < 100]$

preserve 99% of variance

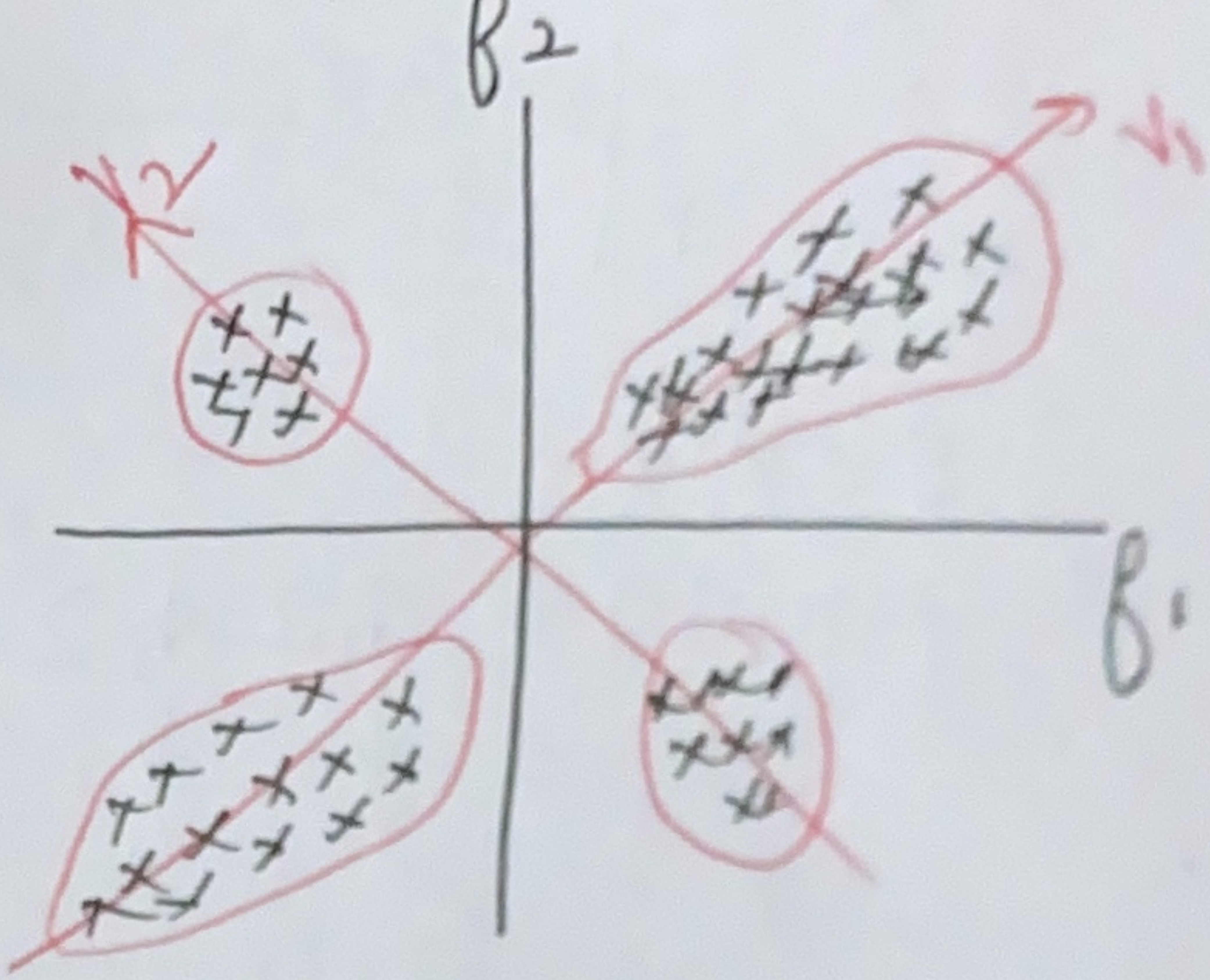
let $\dfrac{d_1 + d_2 \cdots d_{51}}{\sum_{i=1}^{100} d_i} = 0.99$

$d' = 51$

Limitations of PCA

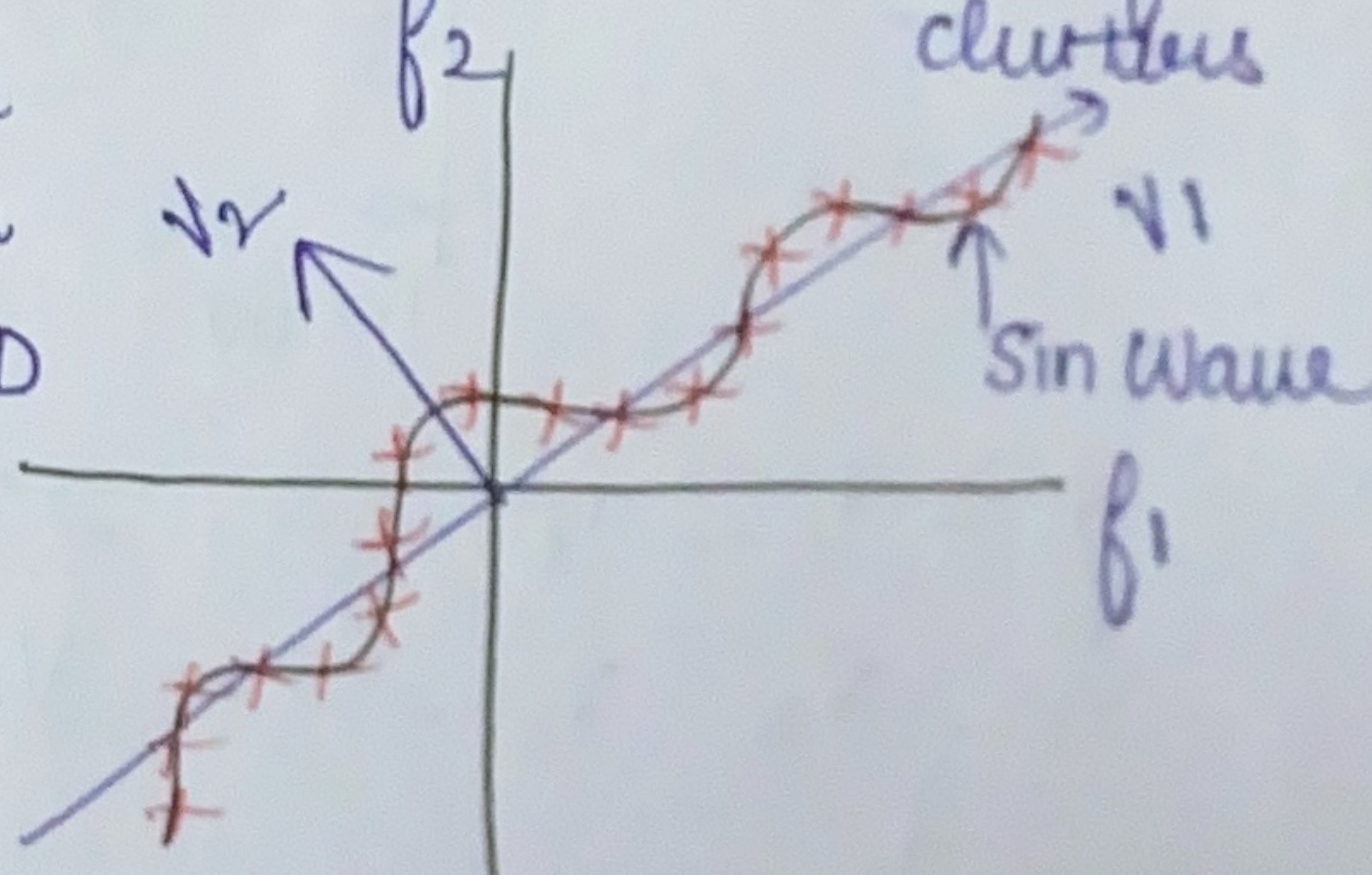


* (info lost is very high)



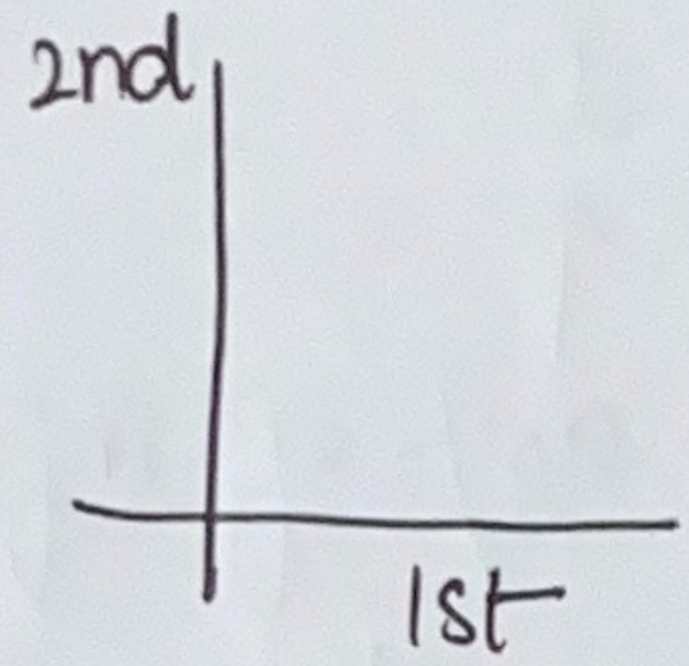
* unable to distinguish data from different clusters

* will never see sine wave when converted to 1D



PCA for dimensionality reduction (not viz)

PCA: $784 \rightarrow 2$



$\rightarrow$ Viz

$784 \xrightarrow{PCA} 10 \rightarrow$ ML-models $[d' \leq d]$

* $784 \rightarrow 200$

how

$$X_{15000 \times 784} \, v_{784 \times 200} = X'_{15K \times 200}$$

$$\uparrow$$

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & 200 \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}_{784 \times 200}$$

$S = X^T X$

$\hookrightarrow d_i \quad i: 1 \rightarrow 784$

$v_i \quad i: 1 \rightarrow 784$

$v_i \in \mathbb{R}^{784}$

$784 \rightarrow 10$

? $\left.\begin{array}{l} 20 \\ 50 \\ 100 \\ 200 \end{array}\right\}$

PCA: max-variance of proj. points

$784 \rightarrow 10 \rightarrow$ original variance explained? $(784 \rightarrow 10)$

PCA:

$C = X^T X$

$\downarrow$

d_i, v_i

$d_1 \leq d_2 \cdots d_{784}$

$784 \rightarrow 10$ dim

Variance explained in 10 dim

$$\boxed{\% \text{ of variance explained} = \frac{d_1 + d_2 \cdots d_{10}}{\sum_{i=1}^{784} d_i}}$$

$$784 \xrightarrow{PCA} d'$$

$\uparrow$

90% of info variance

$$\left\{ \frac{d_1 + d_2 \text{ - - - - - } d_{d'}}{\sum_{i=1}^{d} d_i} = 0.9 \right\}$$