Principle Component Analysis (PCA)

why? dimensionality reduction

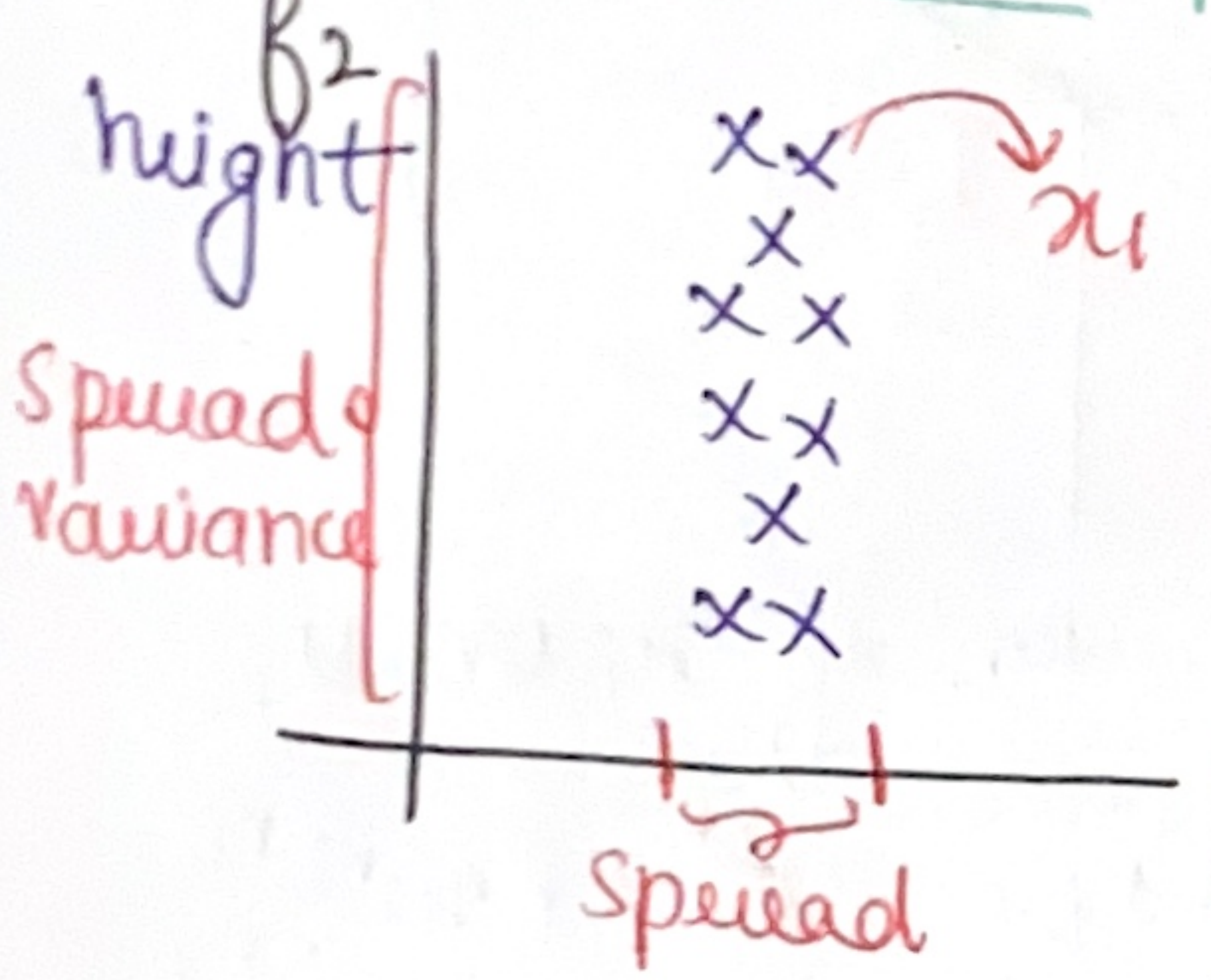
n dim $\rightarrow$ d'-dim

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

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

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

Geometric interpretation of PCA



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

2d $\rightarrow$ 1d

d dim $\rightarrow$ d' dim
$d' < d$

2 dim $\rightarrow$ 1 dim
d $\quad$ d'

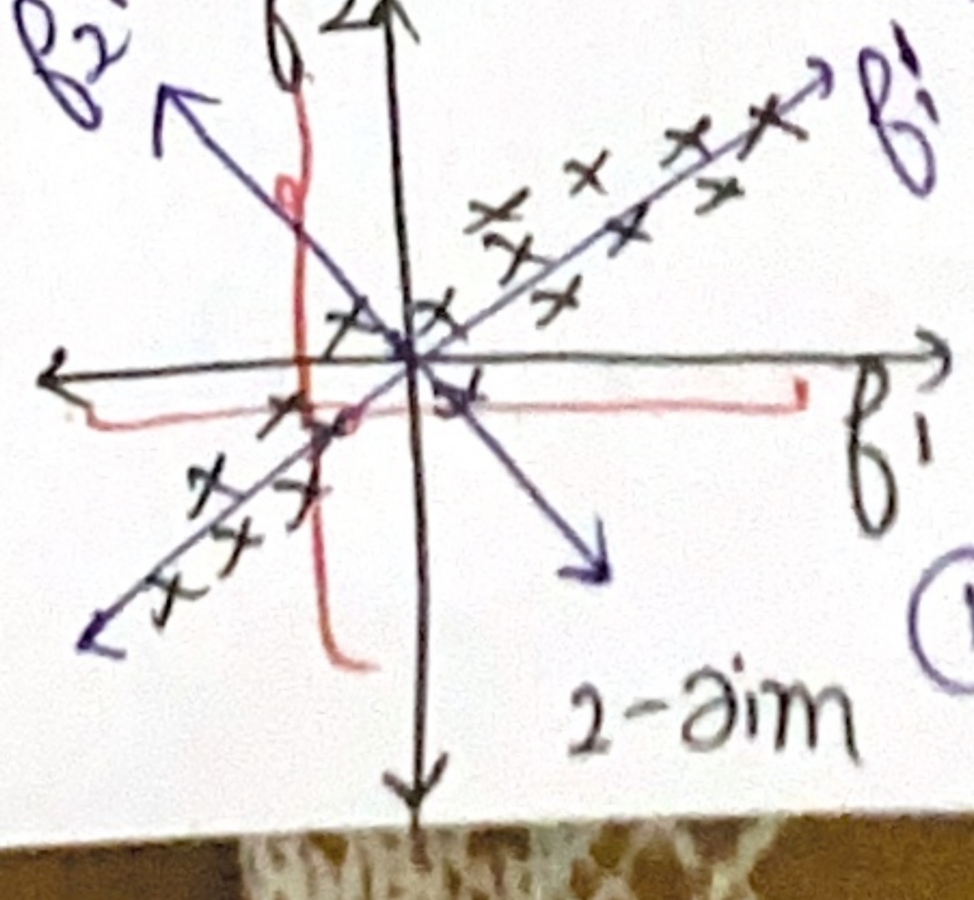
Indian $\leftarrow$ (grey, black)

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

variance = spread = variability

perserving the direction with maximum spread
↳ more information

X = 2-dim dataset



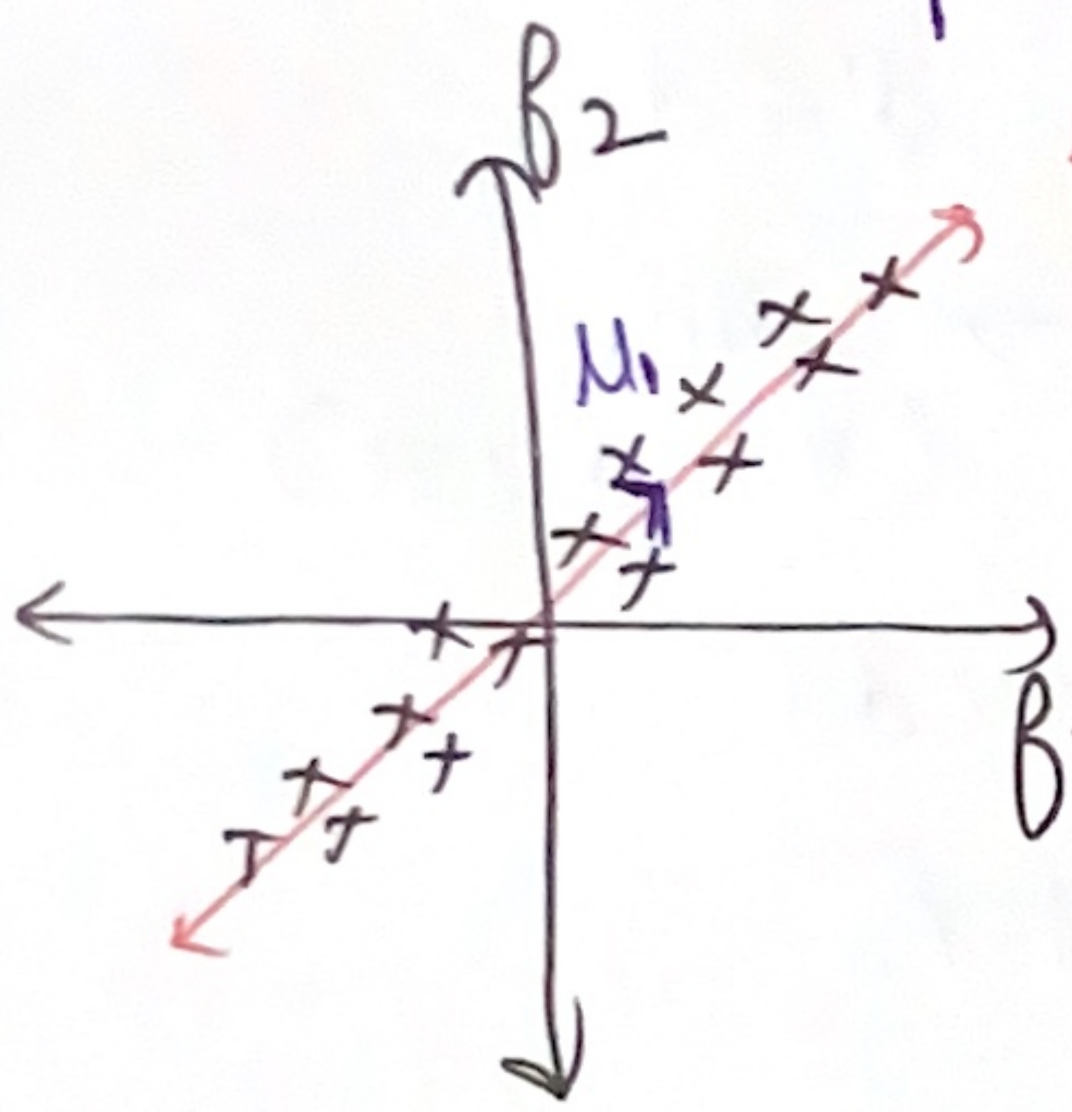
Col standardized Mean$\{f_1\}$ = mean$\{f_2\}$ = 0
Var$\{f_1\}$ = Var$\{f_2\}$ = 1

$\rightarrow$ ② drop f_2'
③ project x_is onto f_1' then 2d $\rightarrow$ 1d

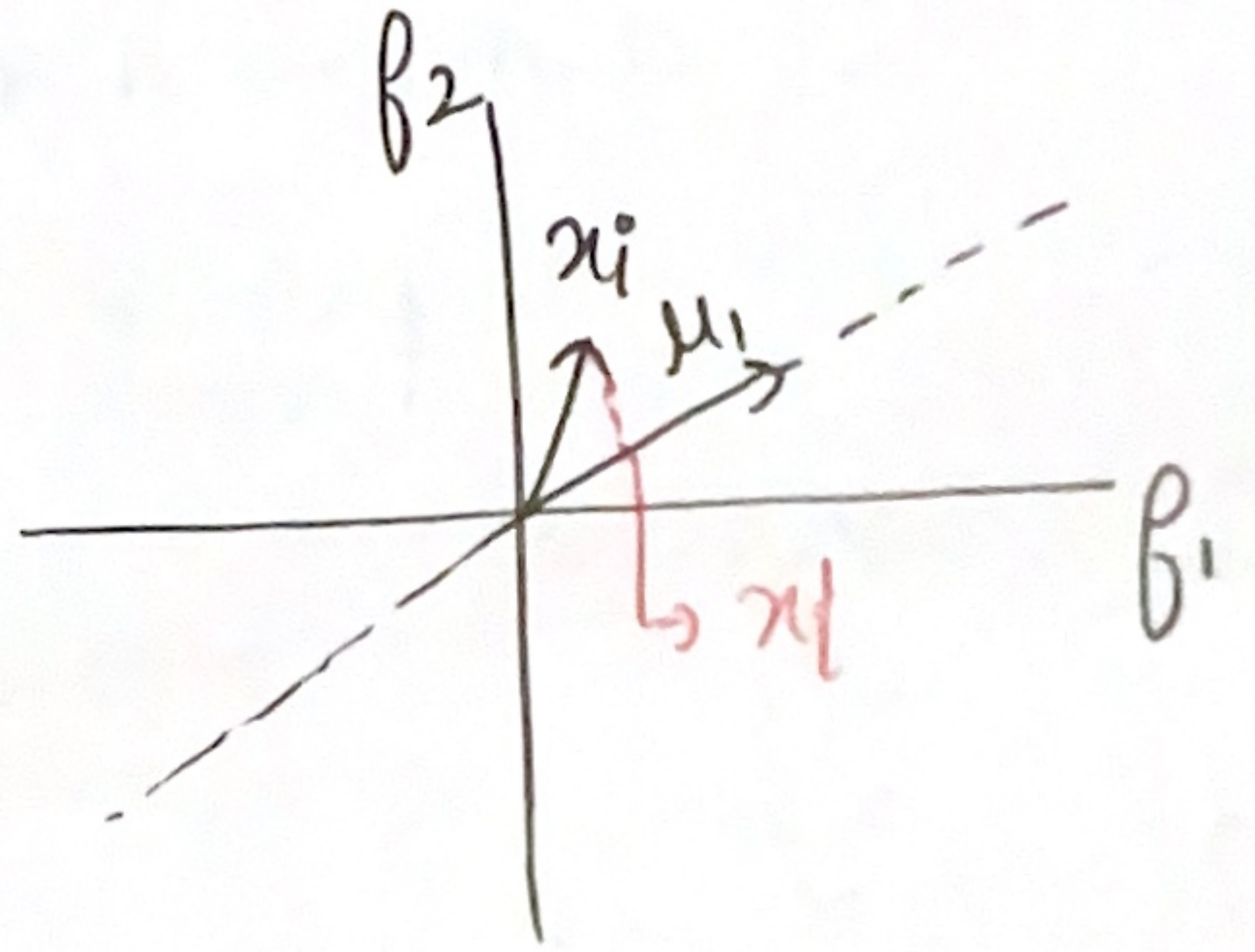
① $f_1' \perp f_2'$ (spread on f_2' << spread on f_1')
st f_1' max spread

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

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

$\boxed{x_i' = u_1^T x_i}$

$$\underbrace{\boxed{\bar{x}_i' = u_1^T \underbrace{\bar{x}}_{\text{mean}\{x_i\}_{i=1}^n}}}_{\text{mean}\{x_i'\}_{i=1}^n}$$

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

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

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

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

data matrix

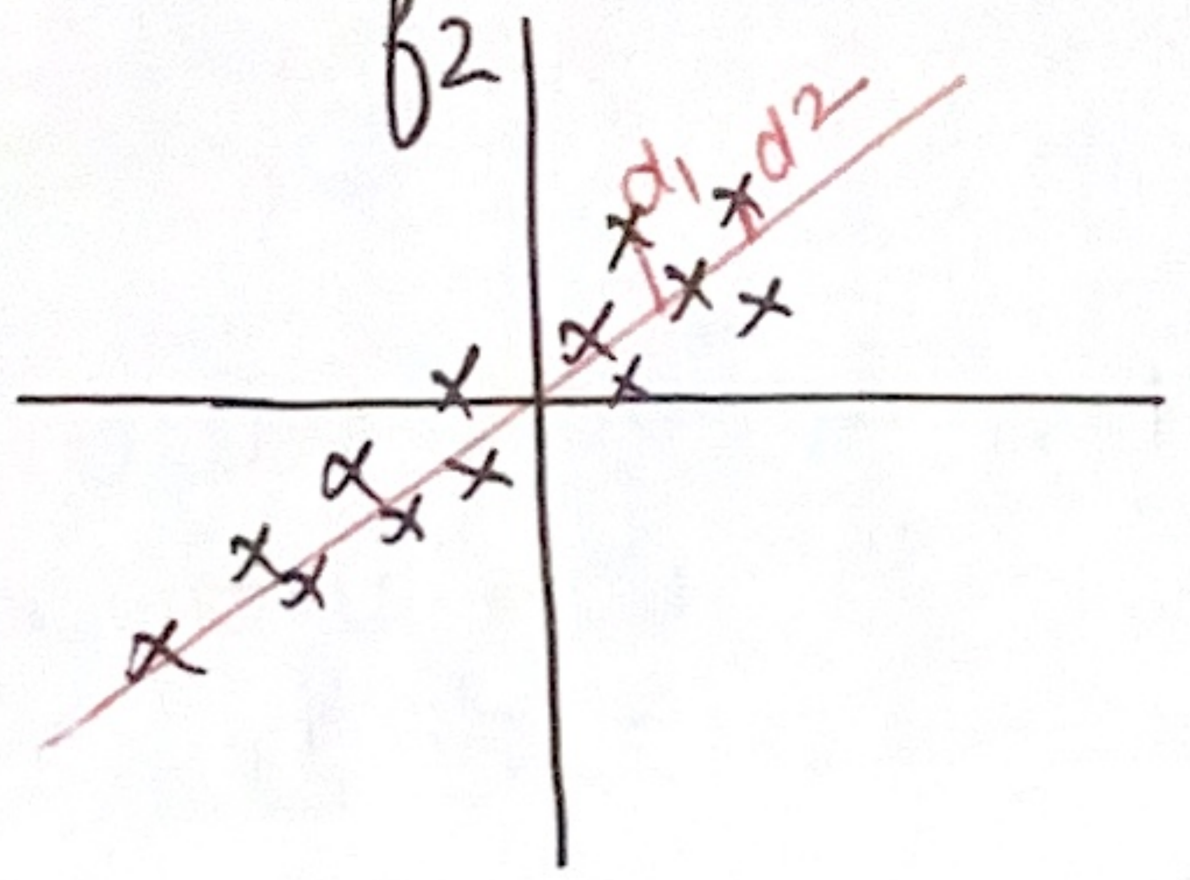
s.t $u_1^T u = 1 = \|u\|^2$

↓

u_1 is unit vector

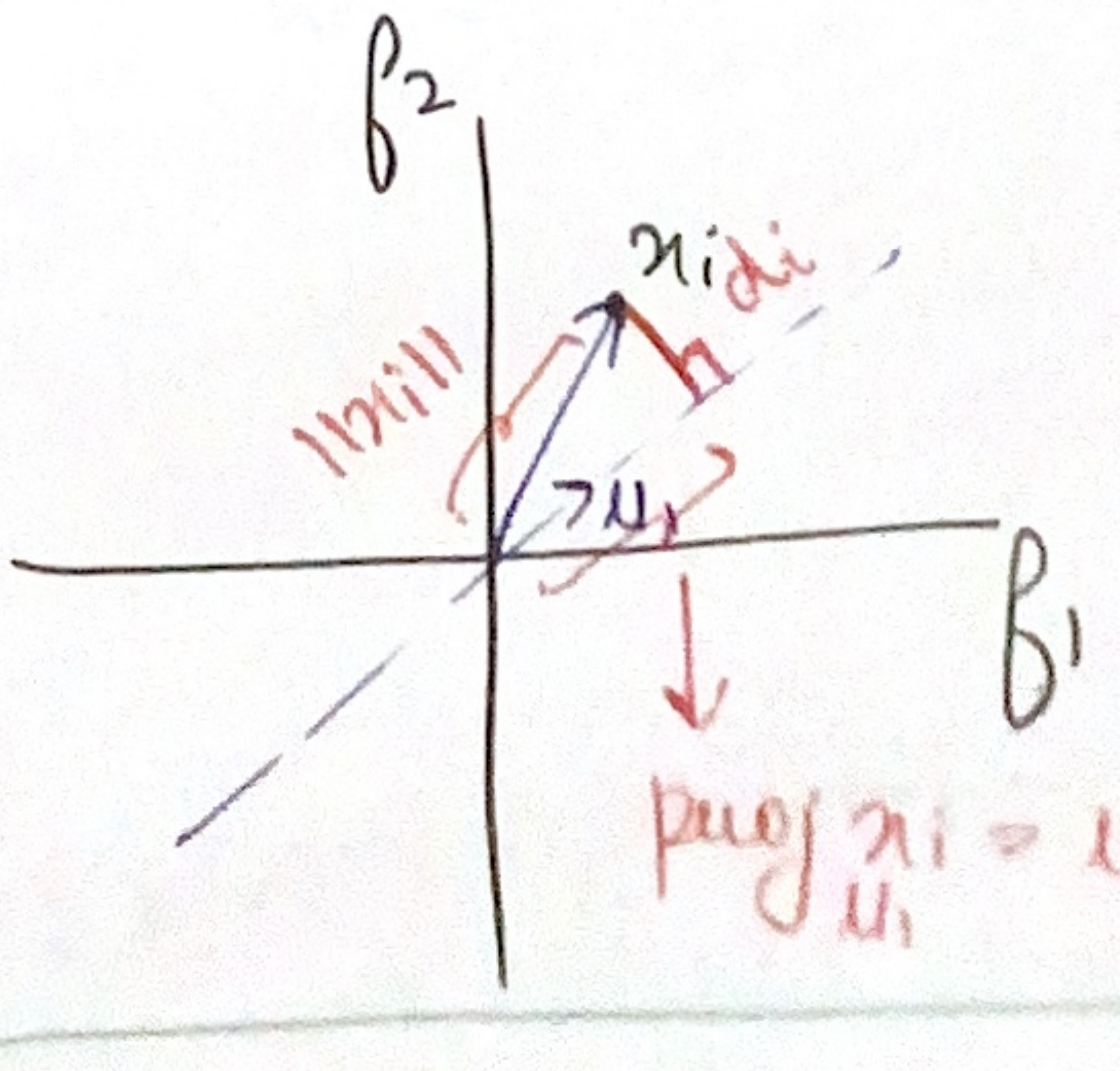
Alternative formulation of PCA: dist. minimization

→ find u_1 which maximizes projected variance



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

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



$\text{proj}_{u_1} x_i = u_1^T x_i$

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

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

s.t $u^T u = 1$

Solution to our optimization problems: λ_1, V_1

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

col std

covariance matrix X

$S = \; = X^T X$

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

eigen values $(\lambda_1, \lambda_2 \text{----} \lambda_d)$

eigen vector $(V_1, V_2 \text{---} V_d)$

max eigen value

$S_{d \times d} = \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \ldots \geq \lambda_d$

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

(eigen vector) of $S = V_1, V_2 \text{----} V_d$

def: $\boxed{\text{If } \lambda_1 V_1 = S V_1 \text{ then } \lambda_1: \text{eigen value of } S,\ V_1: \text{eigen vector of } S \text{ corr to } \lambda_1}$

λ_1: scalar; V_1: $d \times 1$ vector; S: $d \times d$; SV_1: $d \times 1$

$S_{d \times d} \quad \begin{array}{ccccc} \lambda_1 \geq & \lambda_2 \geq & \lambda_3 & \text{---} & \geq \lambda_d \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ V_1, & V_2, & V_3 & \text{------} & V_d \end{array}$

$\boxed{V_i \perp V_j}: \; V_i^T V_j = 0 = V_i \cdot V_j = 0$

$u_1 = V_1$ = eigen vector of $S\,(= X^T X)$ corr to largest eigen value $(= \lambda_1)$

max variance direction

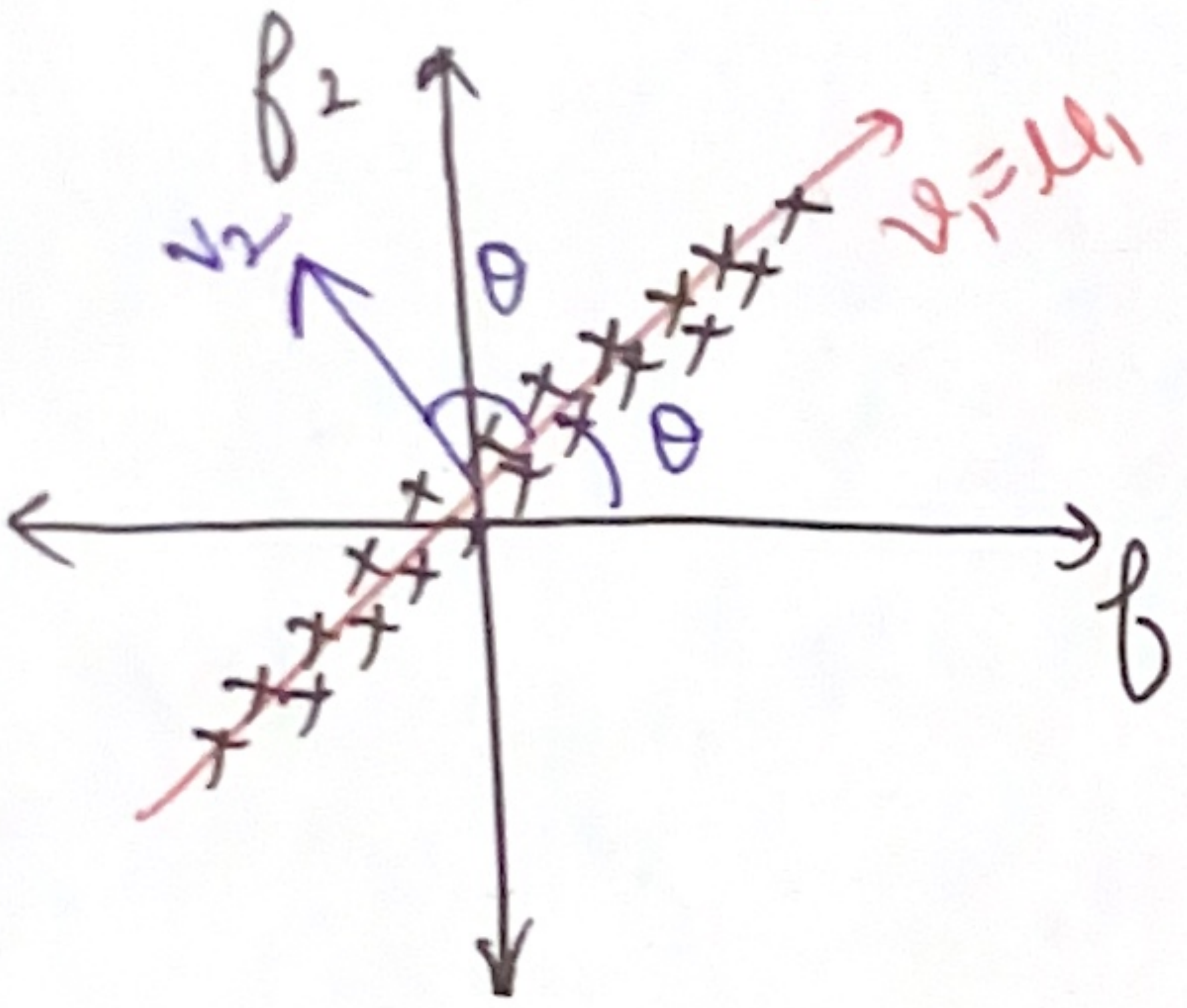
$X = \begin{bmatrix} \quad \end{bmatrix}$

Steps

1. Colmn std of X is done
2. $S = X^T X$
3. eigen values & vectors of S

   $\lambda_1 \geq \lambda_2 \text{ --- } \lambda_d$

   $v_1, v_2 \text{ ---- } v_d$
4. $u_1 = v_1$



2-dim

$d = 2$

$\lambda_1 \geq \lambda_2$

$v_1 \perp v_2$

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

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

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

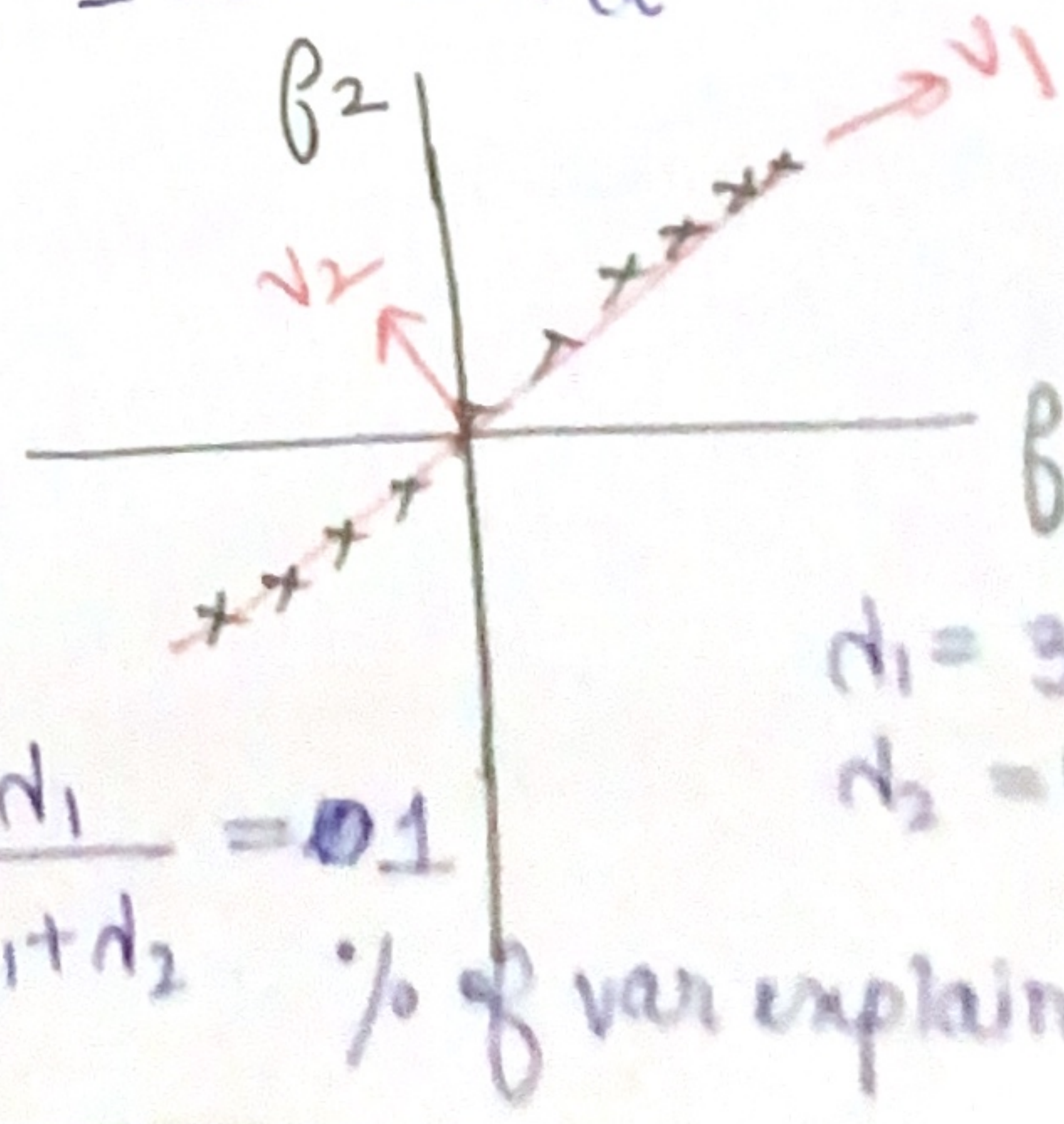
v_1: direction with max var

v_2: 2nd direc with most variance

v_3: 3rd maximal direction

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \text{ ---- } \geq \lambda_d$

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

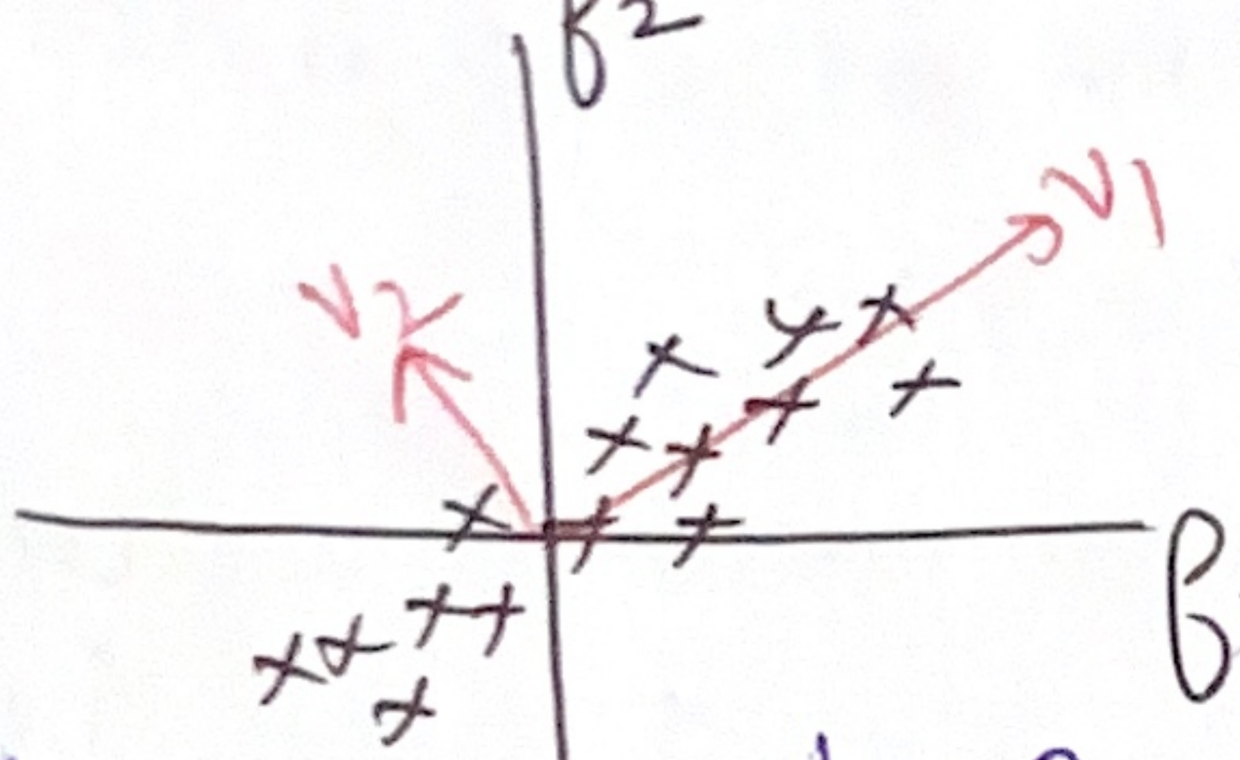


$\lambda_1 = 3$

$\lambda_2 = 0$

$\frac{\lambda_1}{\lambda_1 + \lambda_2} = 0.1$

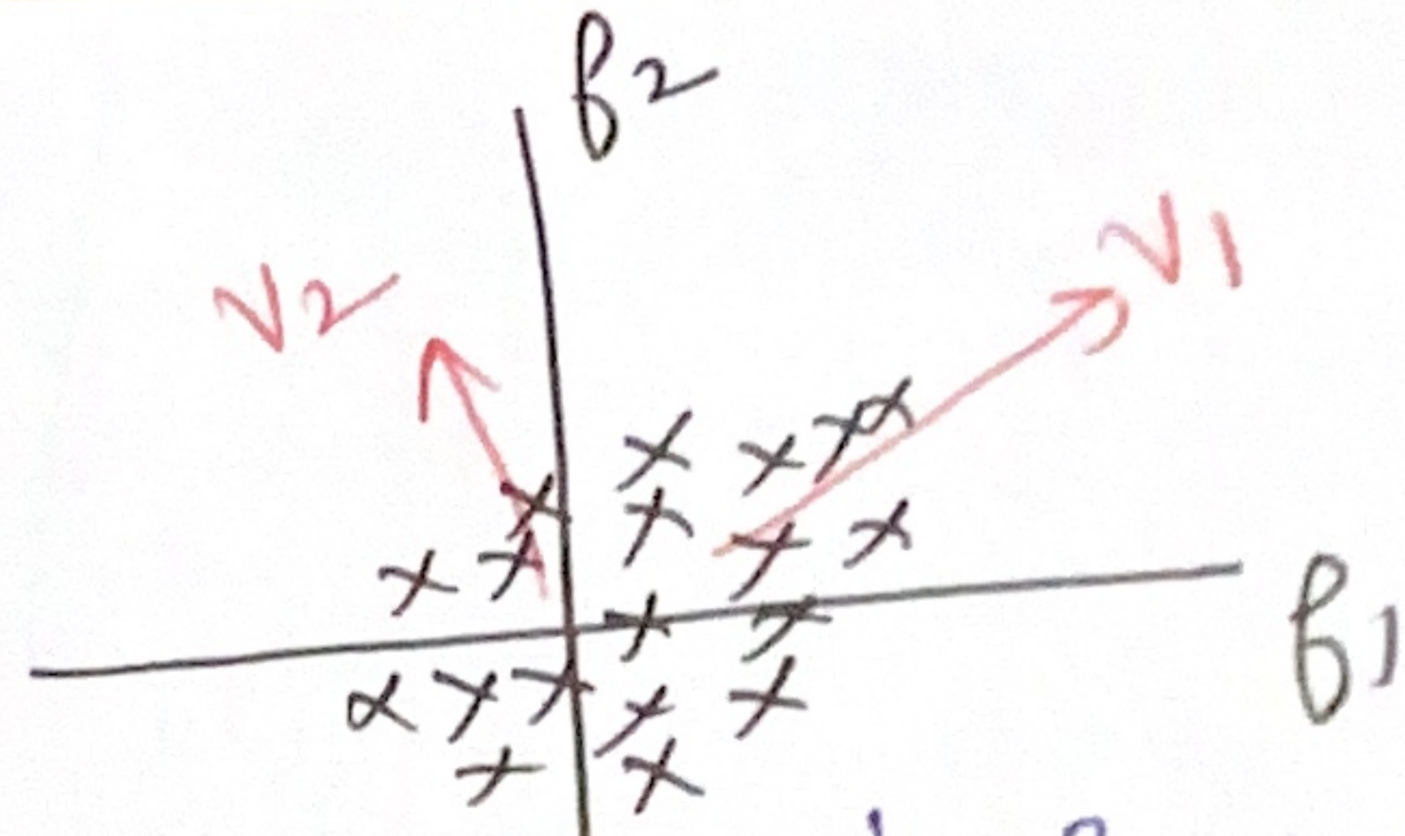
% of var explained



$\frac{d_1}{d_1+d_2} = \frac{3}{4}$

$d_1 = 3$

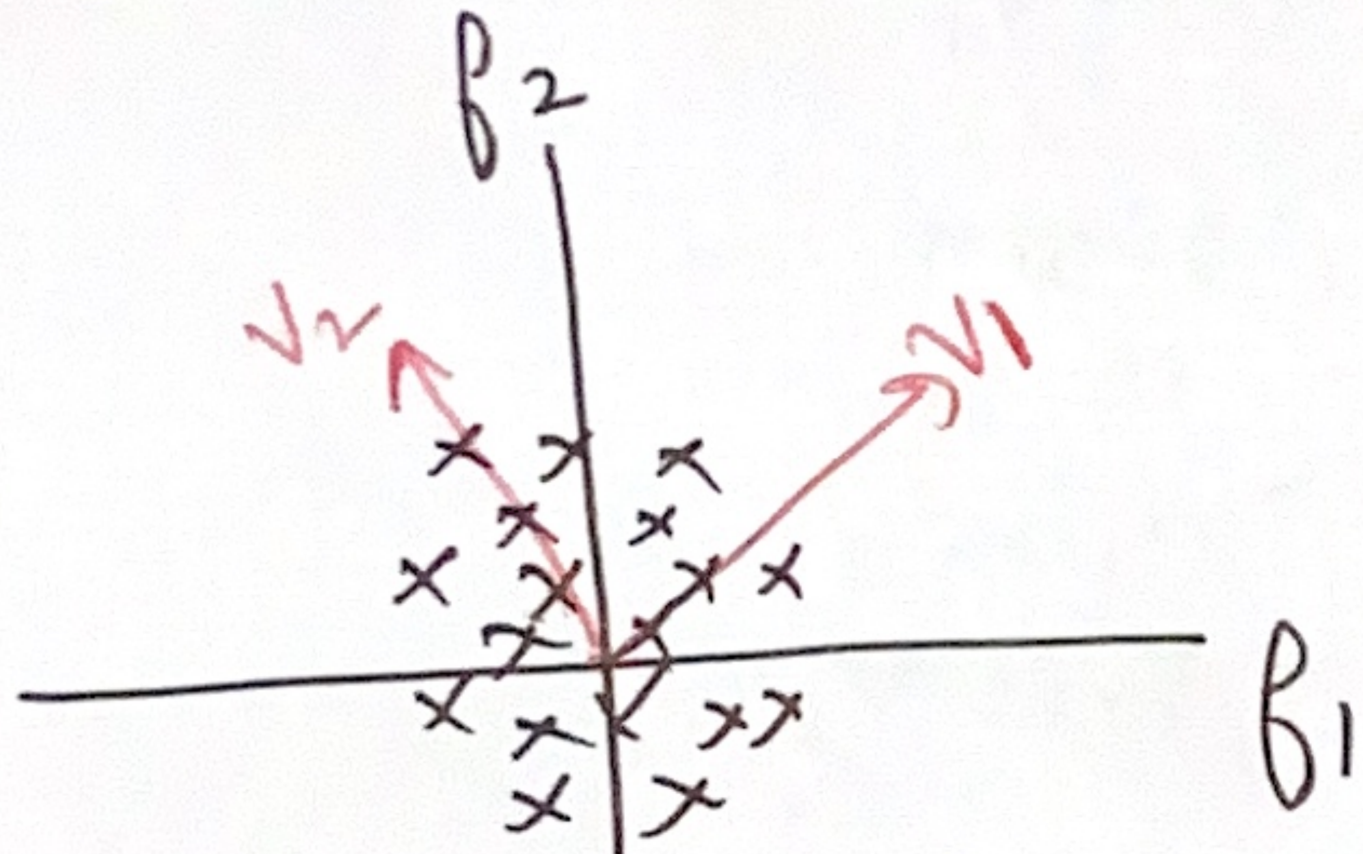
$d_2 = 1$



$\frac{d_1}{d_1+d_2} = \frac{3}{5}$

$d_1 = 3$

$d_2 = 2$

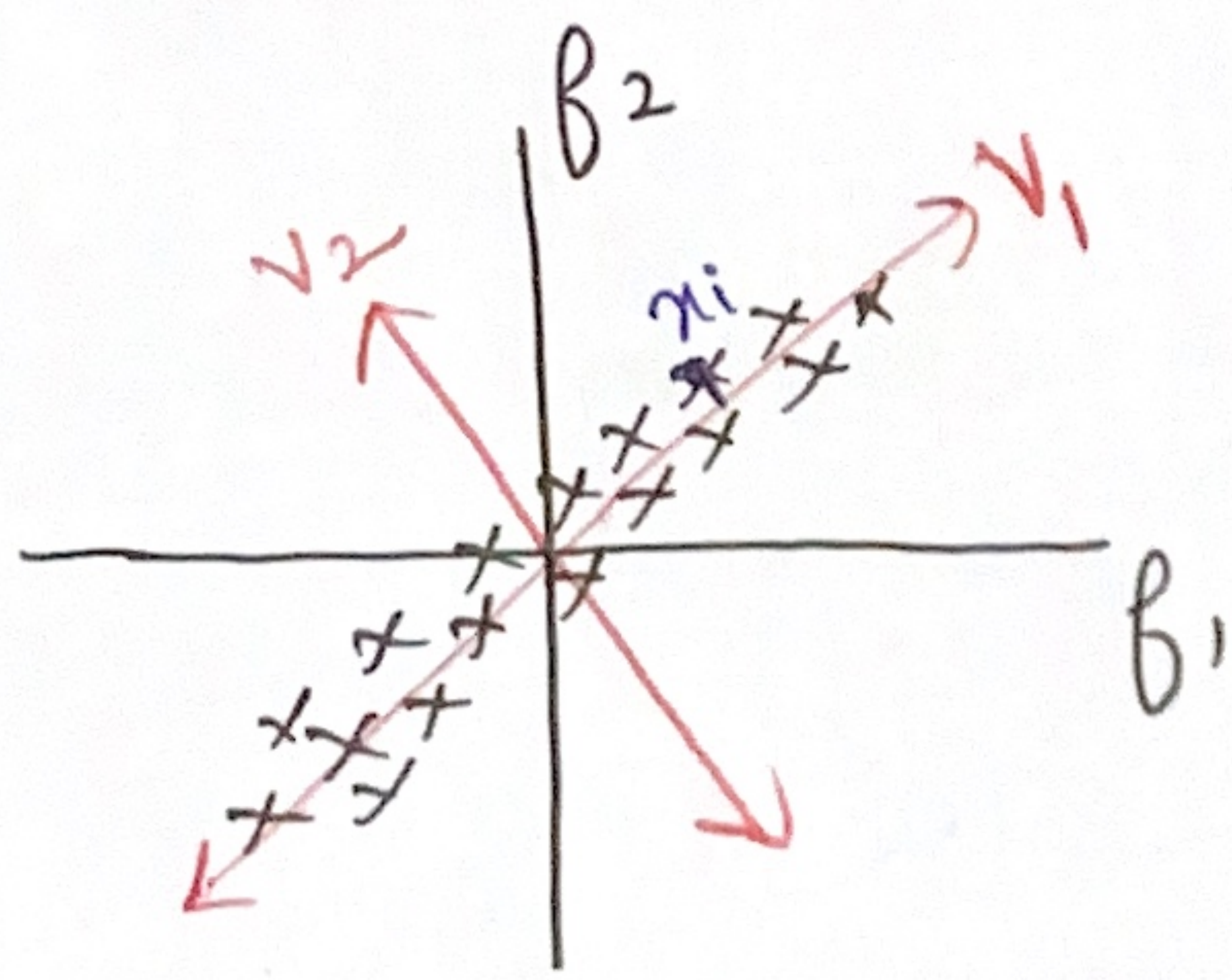


$\frac{d_1}{d_1+d_3} = \frac{3}{6} = \frac{1}{2}$

$d_1 = 3$

$d_2 = 3$

PCA for dim-reduction



2-D

$$X = \begin{array}{c|c} & f_1 \quad f_2 \\ \begin{matrix}1\\2\\ \vdots\\ n\end{matrix} & \begin{bmatrix} \\ \longleftarrow x_1^T \longrightarrow \\ \\ \end{bmatrix} \end{array}$$

max Varia Method (PCA)

$$S = X^T X$$

1-D

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

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

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

$S = X^T X =$

$\text{eigen}(S) = \lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_{10}$

$\downarrow \quad \downarrow \quad \downarrow \cdots \downarrow$

$v_1 \quad v_2 \quad v_3 \quad v_{10}$

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

$v_1, \; v_2$

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

$\underset{100 \times 100}{S = X^T X}$ $\quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_{100}$

$\downarrow$

$v_1 \quad v_2 \quad v_3 \cdots v_{100}$

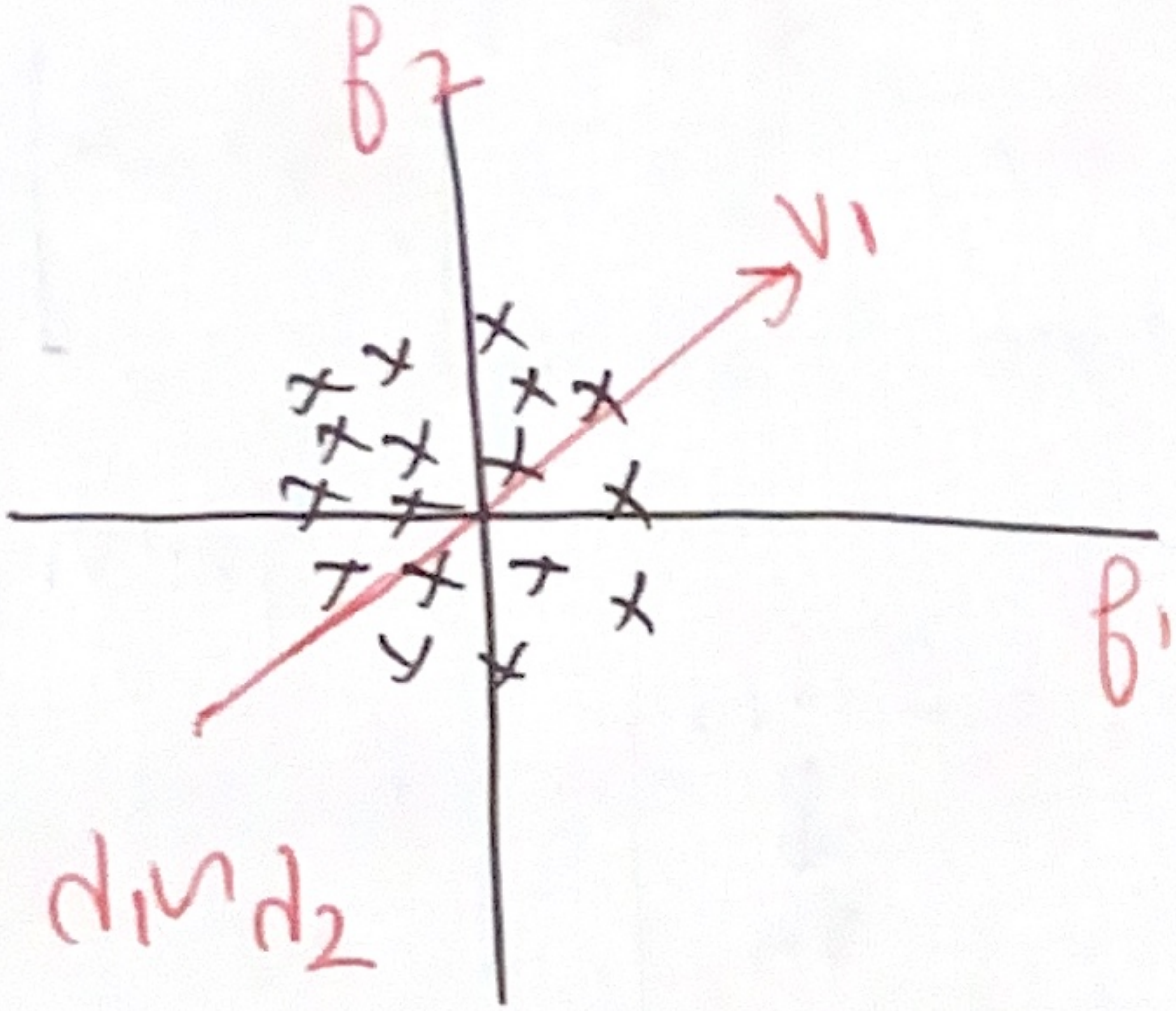
* $x_i \in \mathbb{R}^{100}$; $x_i \in \mathbb{R}^{d'}$ $[d' < 100]$

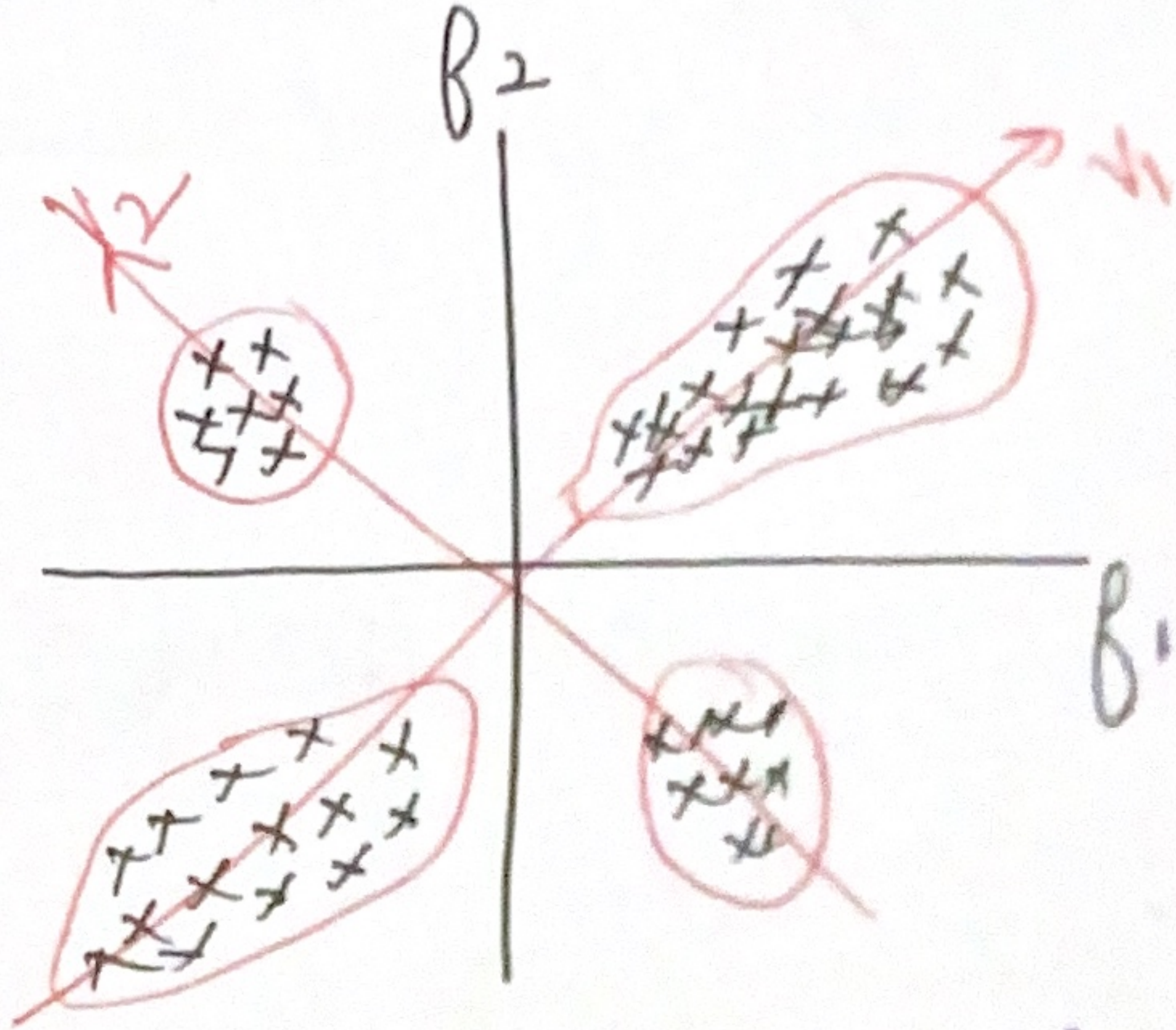
preserve 99% of variance

let $\dfrac{\lambda_1 + \lambda_2 \cdots \lambda_{51}}{\sum_{i=1}^{100} \lambda_i} = 0.99$

$d' = 51$

Limitations of PCA

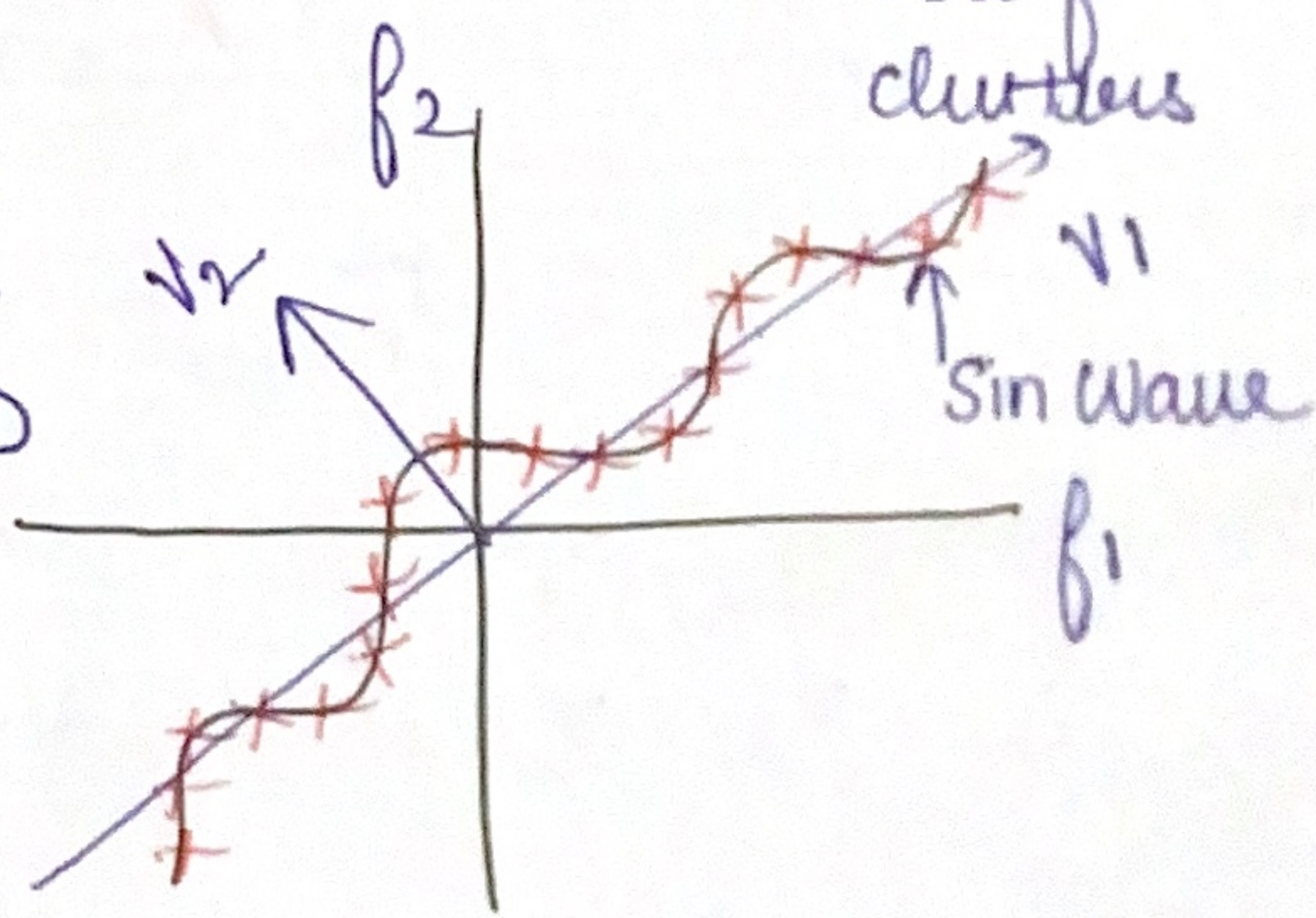




* (info lost is very high)

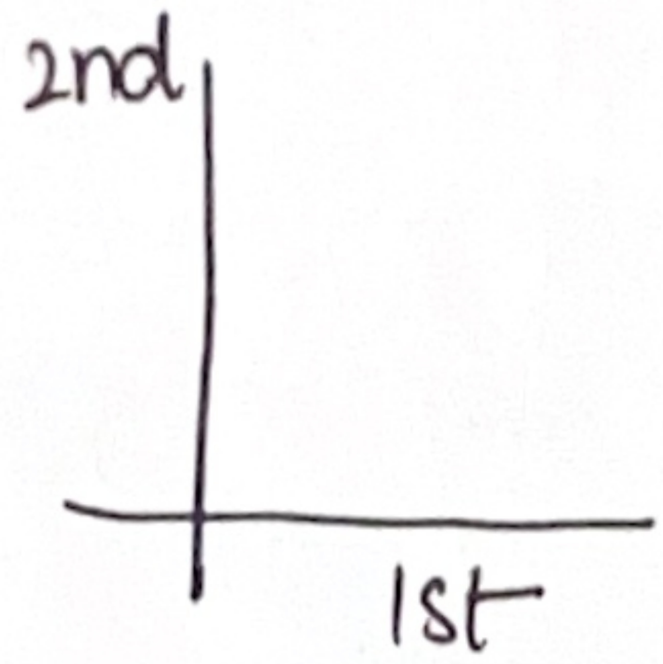
* unable to distinguish data from different clusters

* 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

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

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ V_1 & V_2 & \cdots & 200 \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}_{784 \times 200}$$

$S = X^T X$

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

$V_i \quad i: 1 \rightarrow 784$

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

$784 \rightarrow 10$
? 20
50
100
200

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

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