Dimensionality Reduction

— d-dimension →

x_1	x_2	x_3	x_4	...	x_d	y

1000's ⇒ 10

Features Features

Reasons:

1. Reduce computational Complexity
2. Visualize data

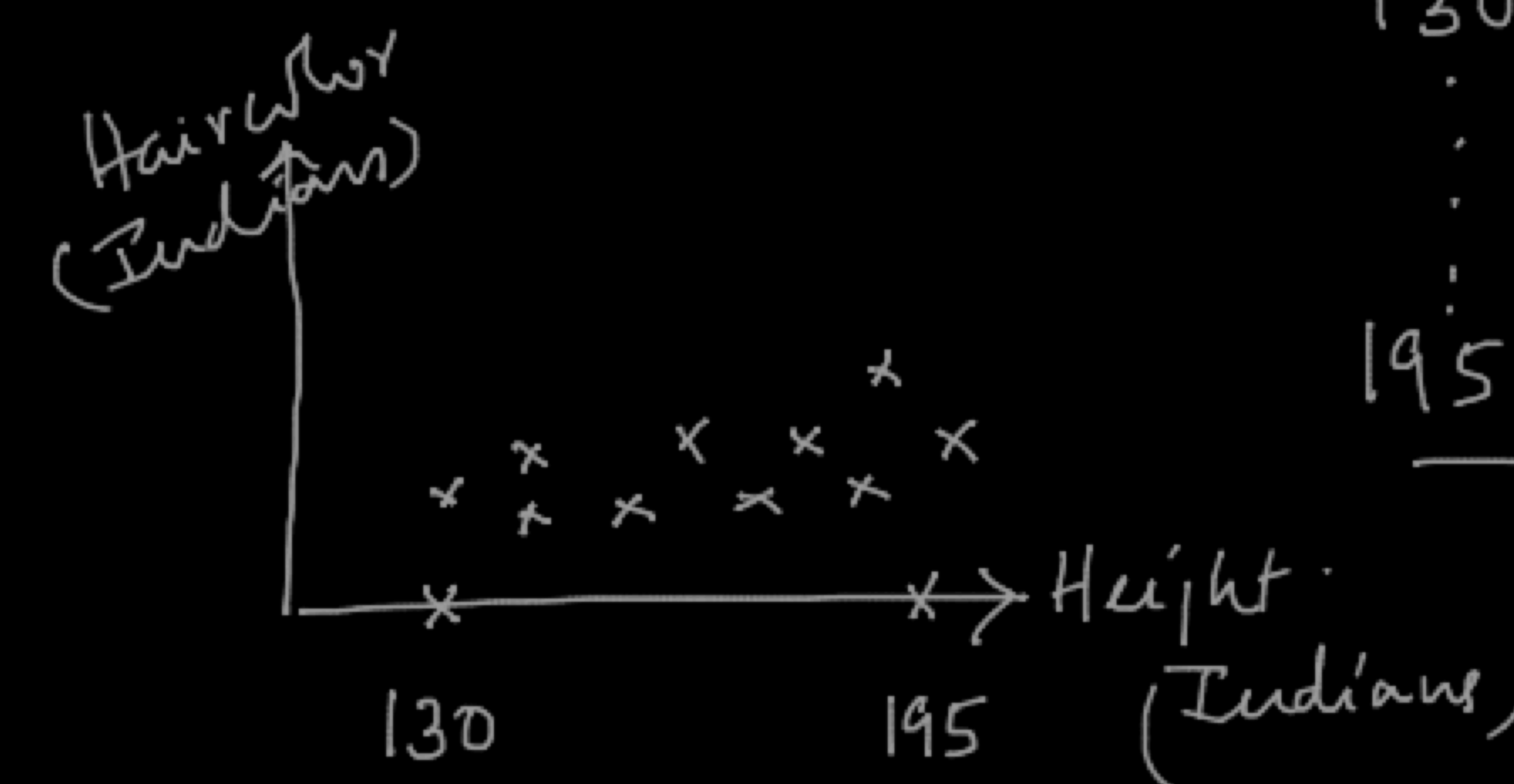
Better ←

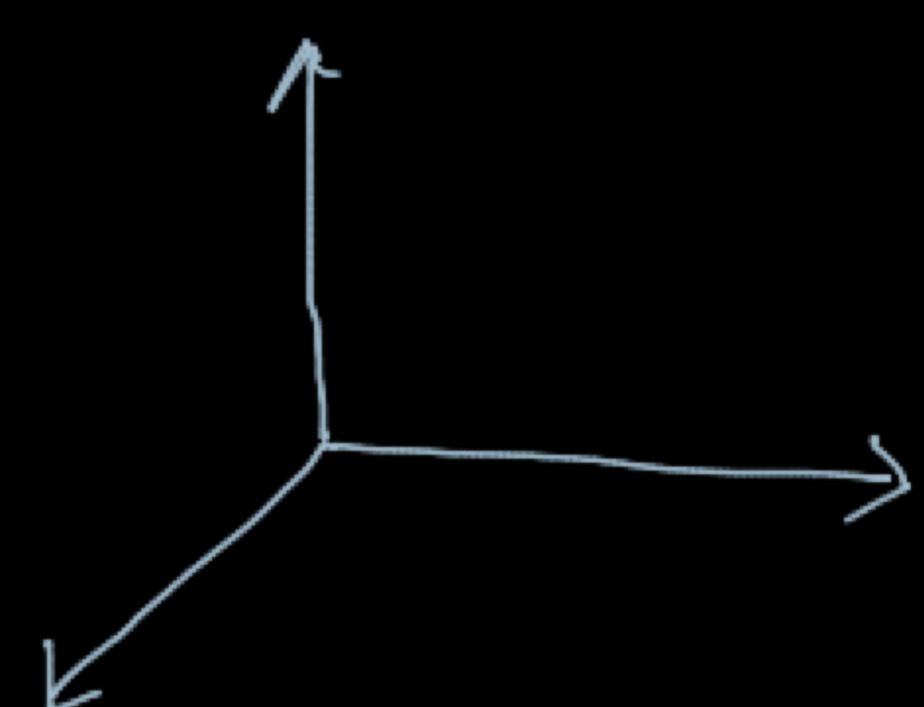
→ very little info

(X_1) Height	(X_2) Haircolor	Person
130	0.02	P_1 ←
.	0.08	P_2
.	0.03	P_3
195 cm	0.00	

↓

Information contained





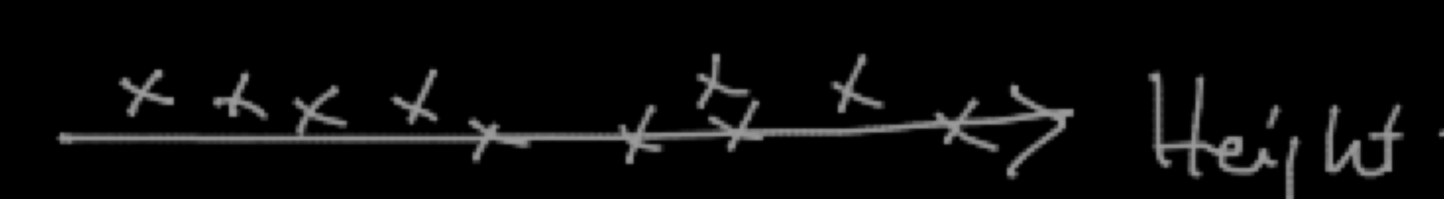
← 2d →

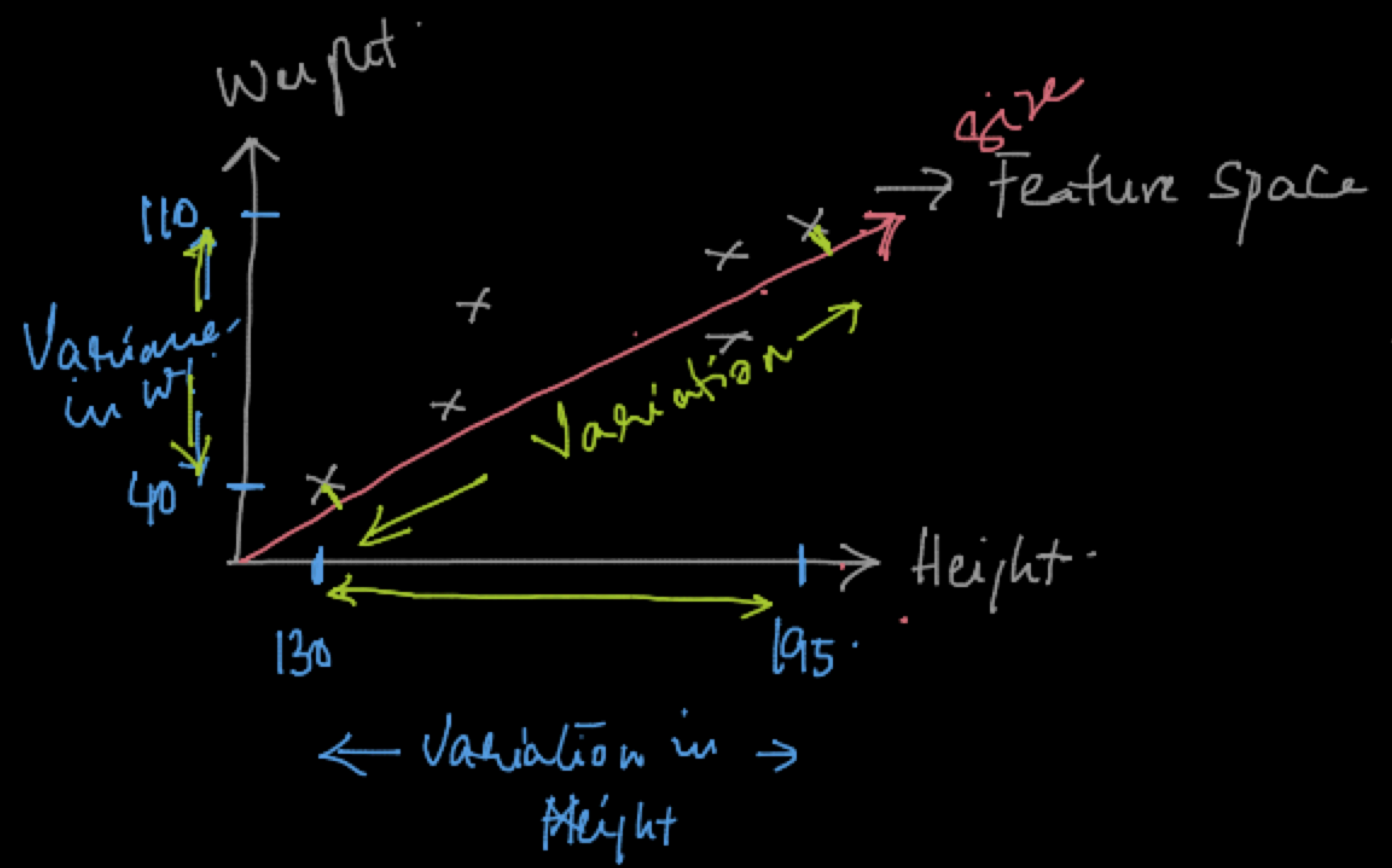
X_1	X_2	Person
.	.	.

→

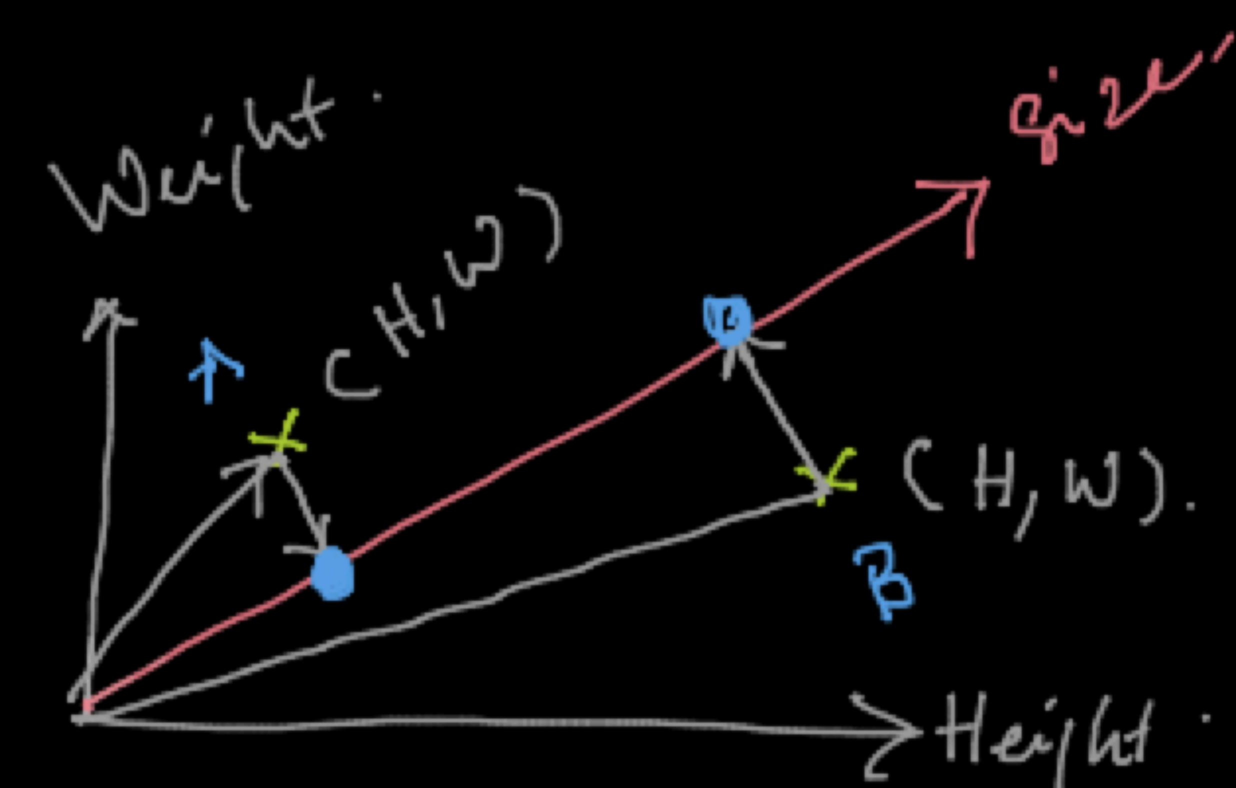
← 1d →

X_1	Person





2d → 1d.



2d → 1d.

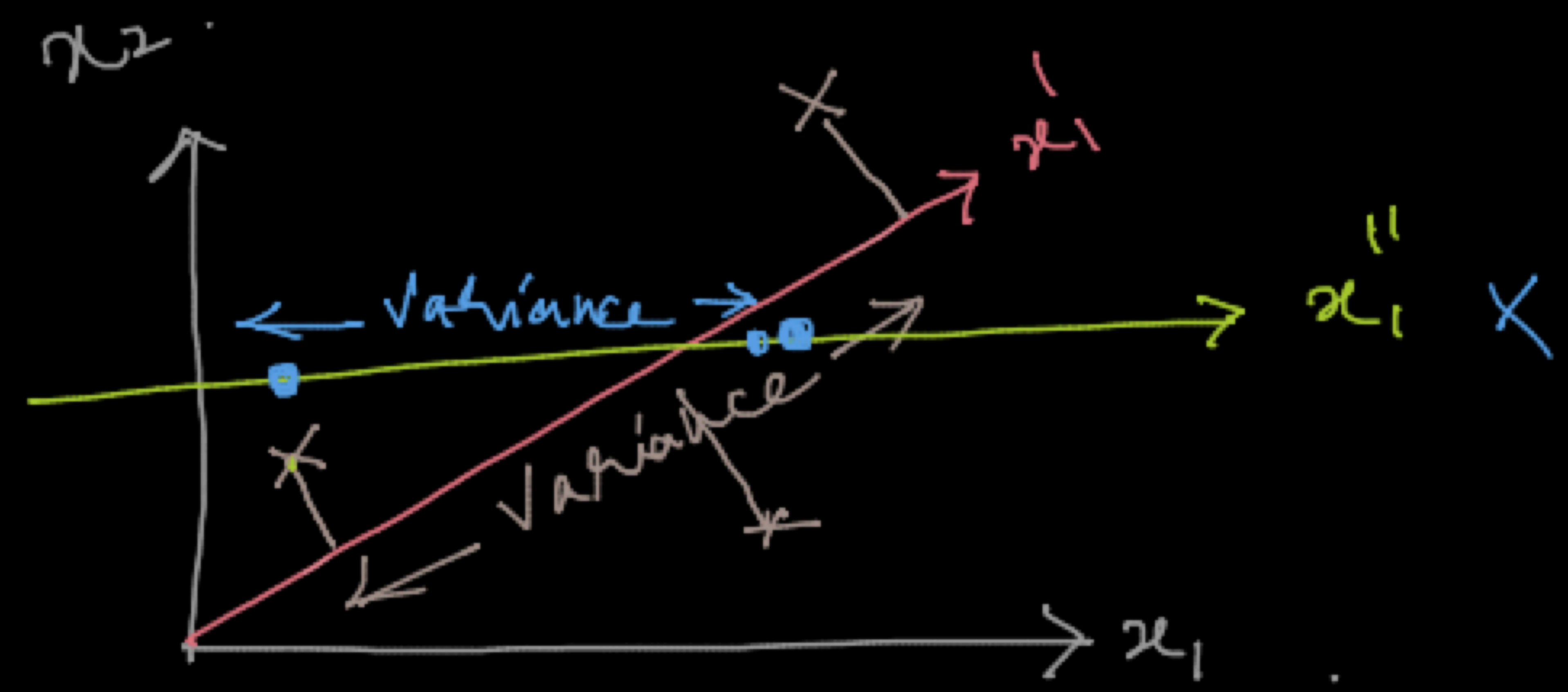
Height Weight ⇒ 1 Feature size

$(a_1 \text{Height} + a_2 \text{Weight})$

Dimensionality Reduction:

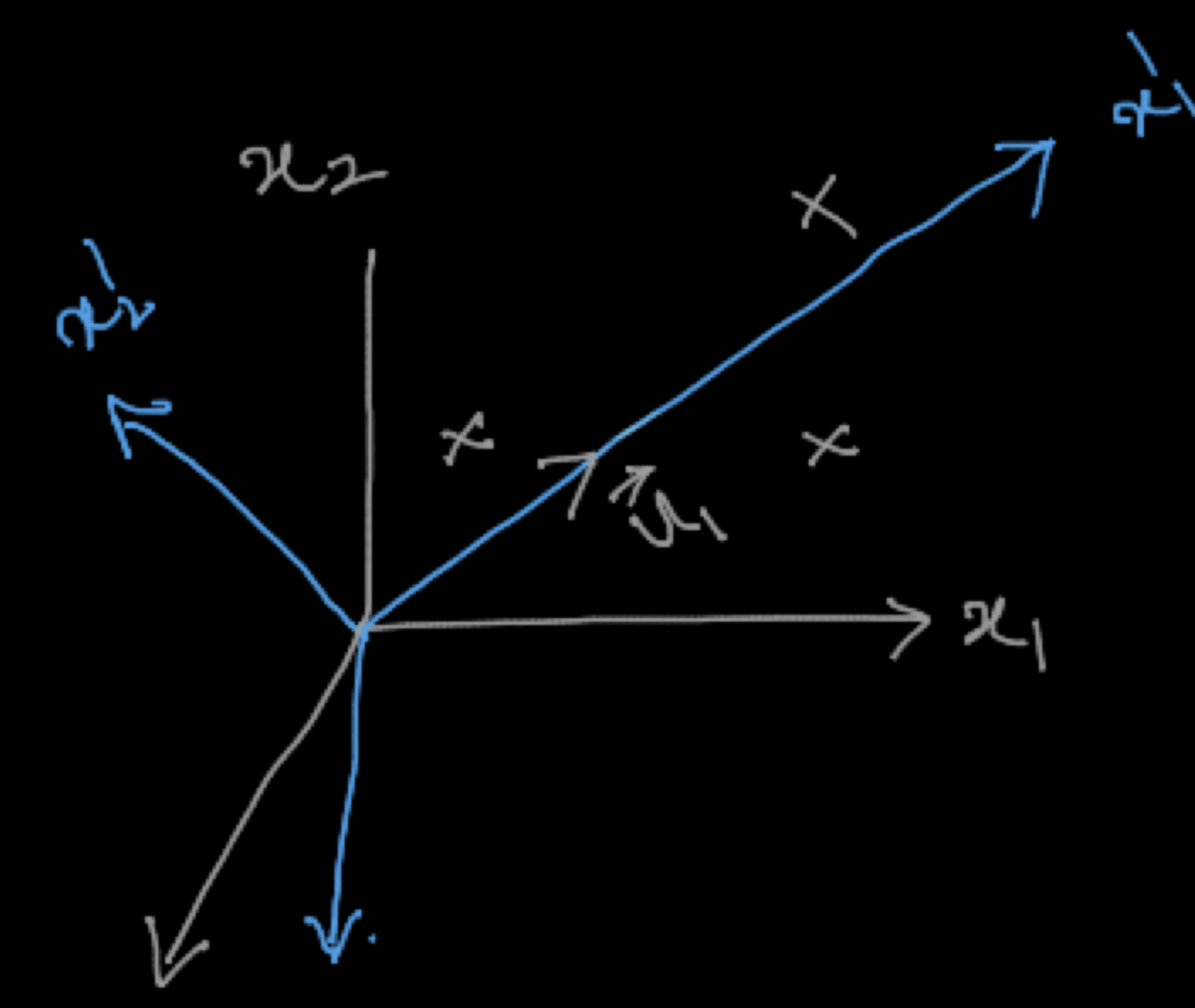
Higher dimension → Lower dimension

Without loss of information.



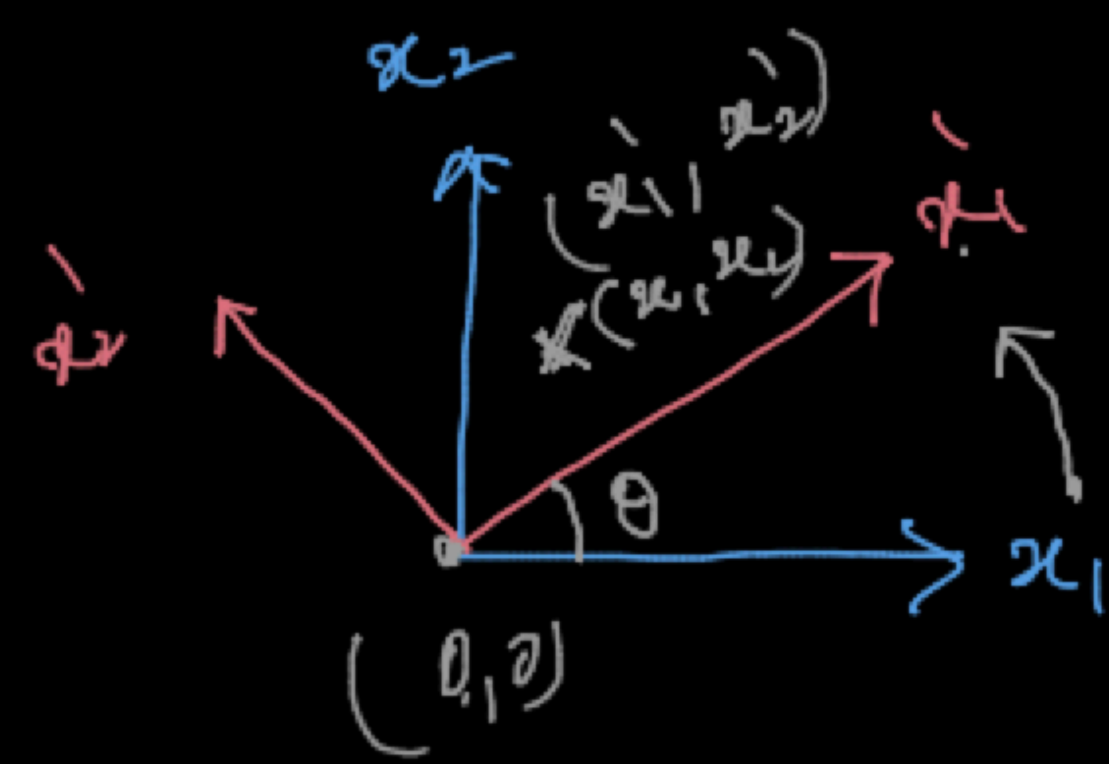
$x_1 \quad x_2 \Rightarrow x_1'$

$(y = mx + c)$



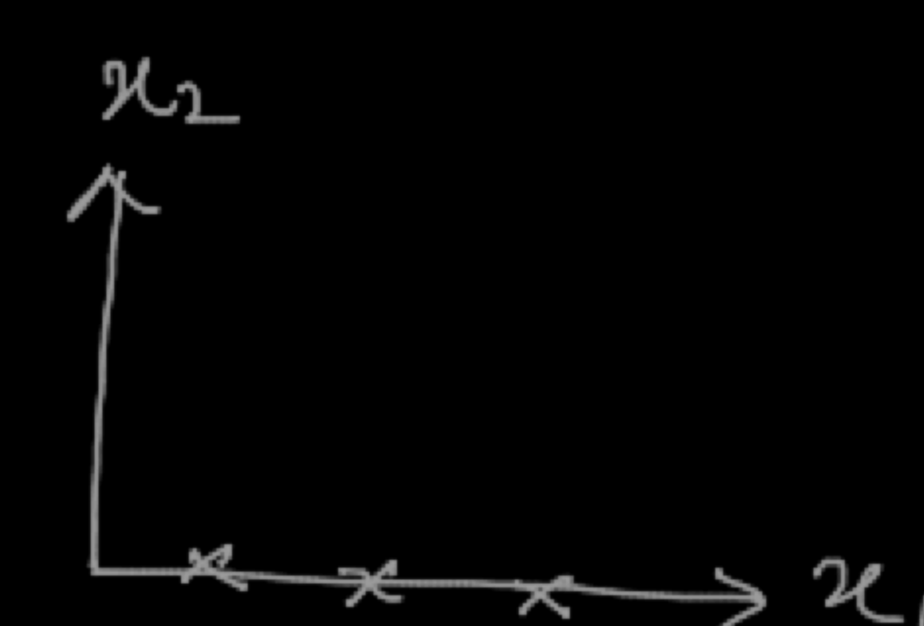
$x_1 \perp x_2 \perp x_3$

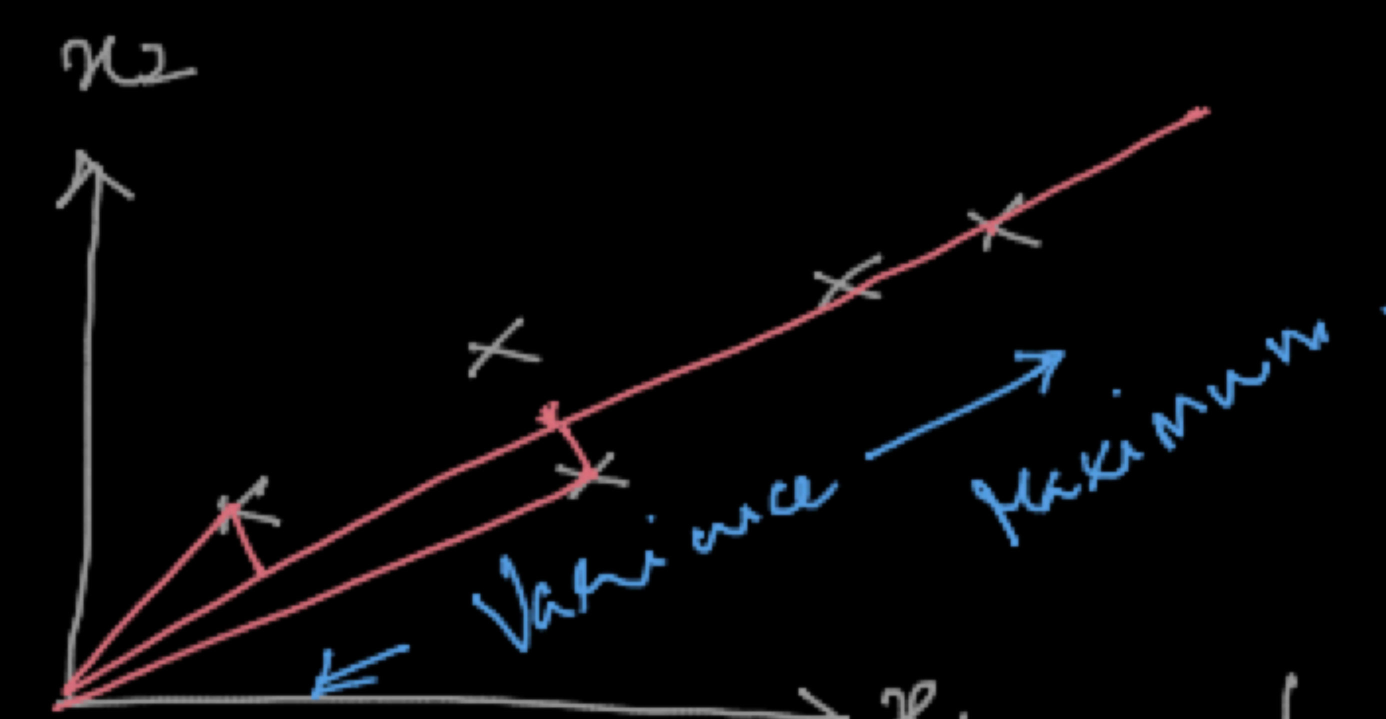
$x_1' \perp x_2' \perp x_3'$



→ orthogonal transformation.

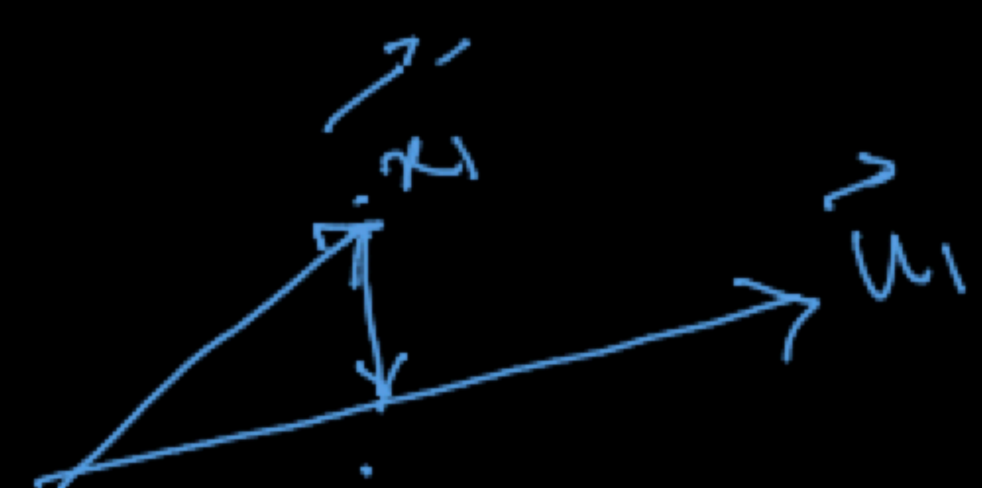
$\sqrt{a}$

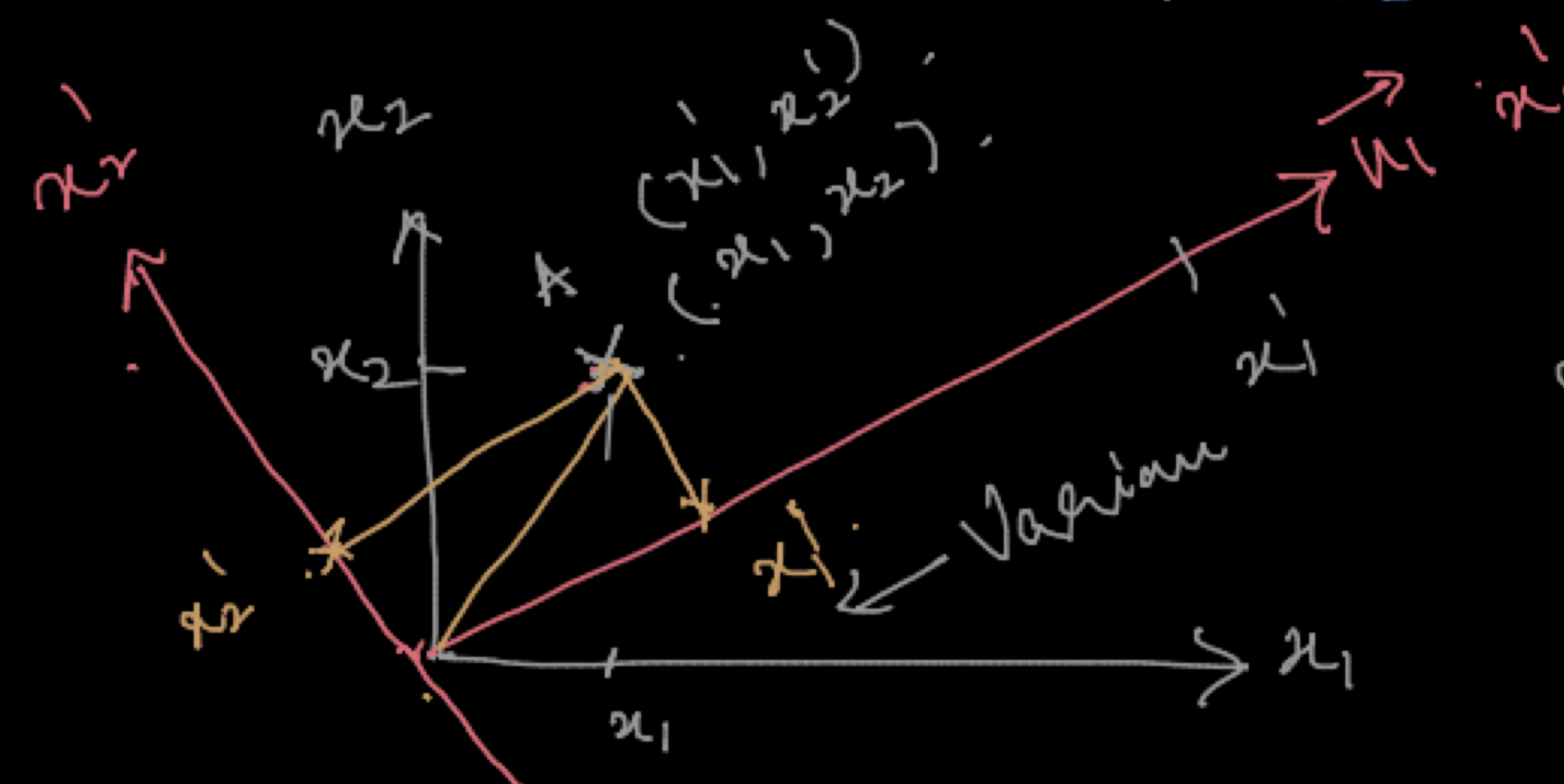




Find that direction ($\vec{u}_1$) that maximizes the Variance in the data.

$$x_1' = \text{Proj}_{\vec{u}_1} \vec{x}_1 = \frac{\vec{u}_1^T \cdot \vec{x}_1}{\|\vec{u}_1\| \Rightarrow 1} = \vec{u}_1^T \vec{x}_1$$





Step 1: Standardize the data

$$\frac{x-\mu}{\sigma} \Rightarrow \begin{array}{l} \bar{x} = 0 \\ \sigma = 1 \end{array}$$

$$\text{Variance} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

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

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

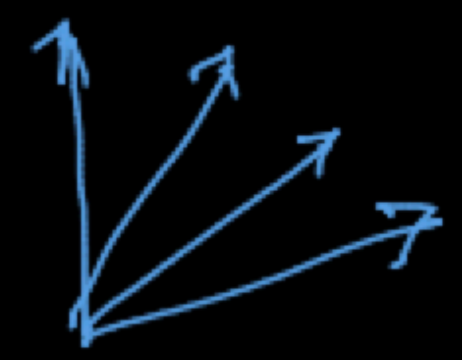
$$x_i' = \vec{u}_1^T \cdot x_i$$

$$\bar{x}' = \vec{u}_1^T \underbrace{\bar{x}}_{\to 0}$$

$$\bar{x}' = 0$$

Task :-

Objective fn $\Rightarrow$ Variance in the new co-ordinate System -



$$Var = \frac{1}{n-1} \sum_{i=1}^{n} \left(\vec{u_i}^T \cdot x_i \right)^2$$

$\rightarrow$ Maximize this

— Find that direction ($\vec{u_i}$) along which we get maximum Variance

Covariance matrix of X.   X - original.

$$X = \begin{array}{c c} & \begin{array}{ccccc} x_1 & x_2 & x_3 & \cdots & x_d \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} & \left[\begin{array}{ccccc} \; & \; & \; & \; & \; \\ \; & \; & \; & \; & \; \\ \; & \; & \; & \; & \; \\ \; & \; & \; & \; & \; \end{array} \right] \end{array}$$

$x \quad y$

$$\text{Covariance} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

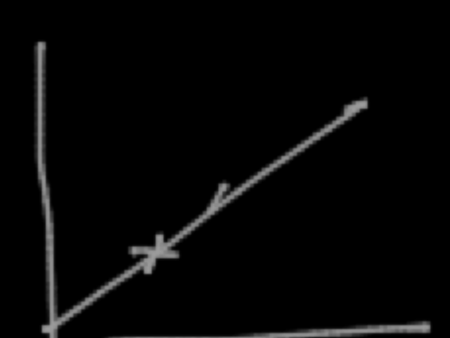
→ How x & y vary w.r.t. each other.

$$X \quad \begin{array}{c|cccccc} & x_1 & x_2 & x_3 & \cdots & x_d \\ \hline 1 & & & & & \\ 2 & & & & & \\ 3 & & & & & \\ \vdots & & & & & \\ n & & & & & \end{array} \Rightarrow \begin{array}{c} S \\ \text{Covariance Matrix of } X \end{array} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & \cdots & x_d \end{array} \\ \begin{bmatrix} V_{x_1} & & & & \\ S_{21} & V_{x_2} & & & \\ S_{31} & S_{32} & V_{x_3} & & \\ & & & & \\ S_{d1} & S_{d2} & S_{d3} & \cdots & V_{x_d} \end{bmatrix} \end{array} \; d \times d$$

↳ Square Matrix

→ Variance of the features along the diagonal.

$$\mathrm{Cov}(x_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)$$



$$\mathrm{Cov}(x_1, x_1) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \Rightarrow \text{Variance of } x_1$$

$$=$$

Eigen Values & Vectors $\rightarrow$

$$S = \begin{bmatrix} & \\ & \end{bmatrix}_{d \times d} \Rightarrow$$

d- Eigen vectors -(directions).

d - Eigen Values (Variance along these direction).

$\vec{u_1}$

$$\vec{v_1} \quad \vec{v_2} \quad \boxed{\vec{v_3}} \; \cdots \; \vec{v_d} \rightarrow \text{d-eigen vectors}$$

$$\lambda_1 \quad \lambda_2 \quad \boxed{\lambda_3} \quad \lambda_d \rightarrow \text{d-eigen Values.}$$

$$S\vec{v} = \lambda \vec{v} \rightarrow \text{Eigen vector}$$

Matrix $\leftarrow$ (S)

λ $\rightarrow$ Eigen Value (Constant)

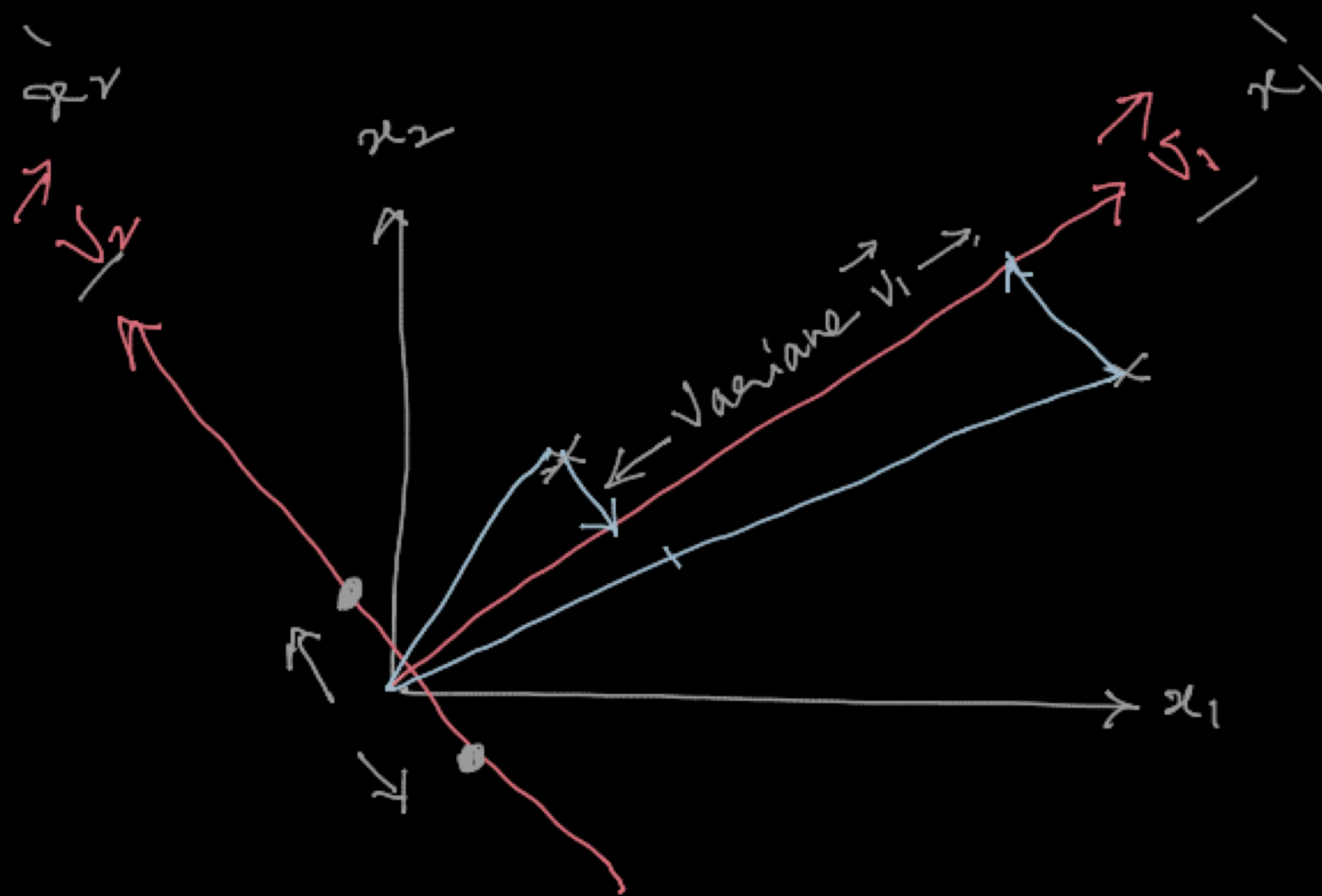
$X \rightarrow$ Standardize the data (Z_x) $\rightarrow$ Covariance matrix of (Z_x) (S) $\rightarrow$ Find the Eigen Values & Eigen Vectors of 'S'

Arrange them in desc. order of Eigen values.

$$\vec{V_1} \perp \vec{V_2} \perp \vec{V_3}$$

$$\lambda_1 > \lambda_2 > \lambda_3 \cdots \lambda_d.$$

$$\vec{V_1} \quad \vec{V_2} \quad \vec{V_3} \; \cdots \; \vec{V_d} \rightarrow$$



⇒ Standardized X.

↳ Find Cov. Matrix.

↳ Eigen Values & Eigen Vectors.

→ $\lambda_1 > \lambda_2 > \lambda_3 \dots$

Principal Components. → $\vec{V_1}$ $\vec{V_2}$ $\vec{V_3}$

x_1, x_2, x_3 $\quad$ x_1, x_2, x_3 $\quad$ x_1, x_2, x_3

x_1 x_2

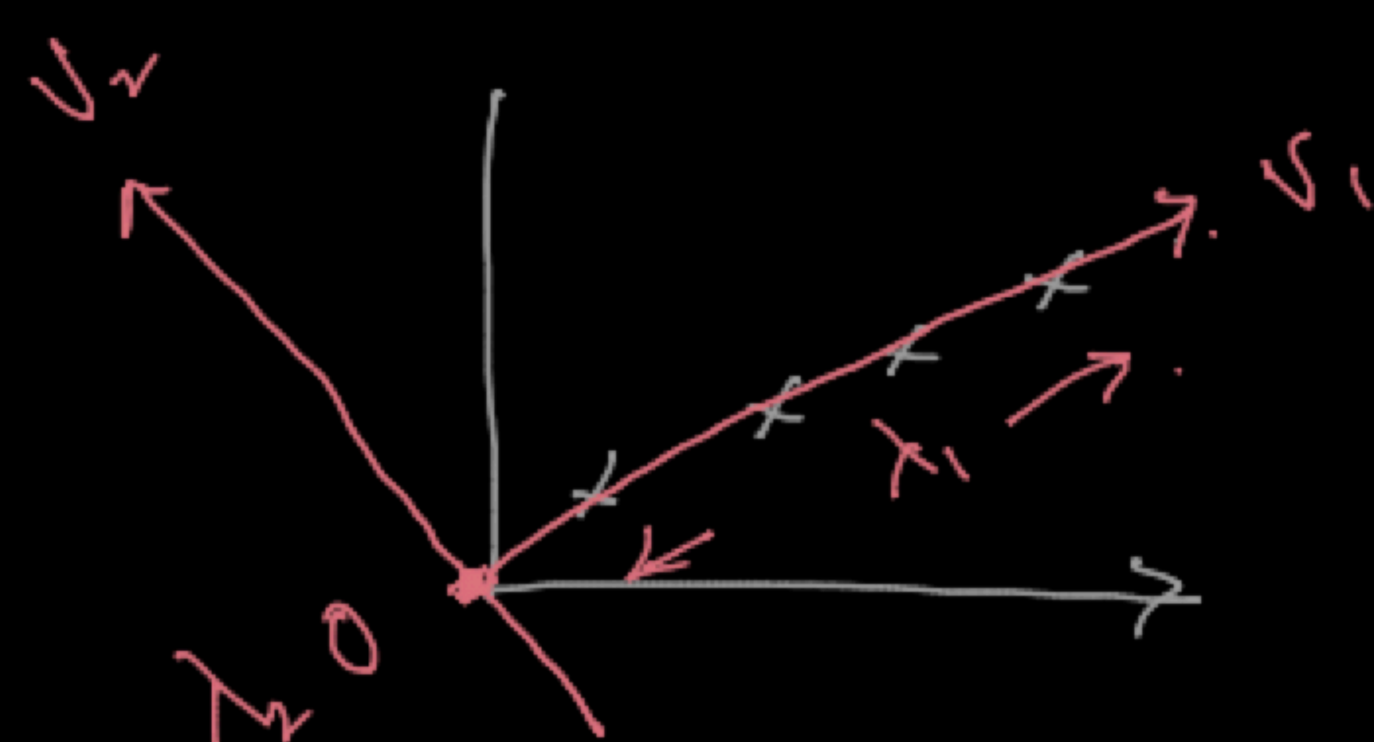
$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ ⇒ Height $\vec{V_1}$

max $\begin{bmatrix} - \\ - \end{bmatrix}$

$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_d \Rightarrow V_1 \quad V_2 \quad V_3 \quad \ldots \quad V_d$

80% $\qquad$ 51.5%

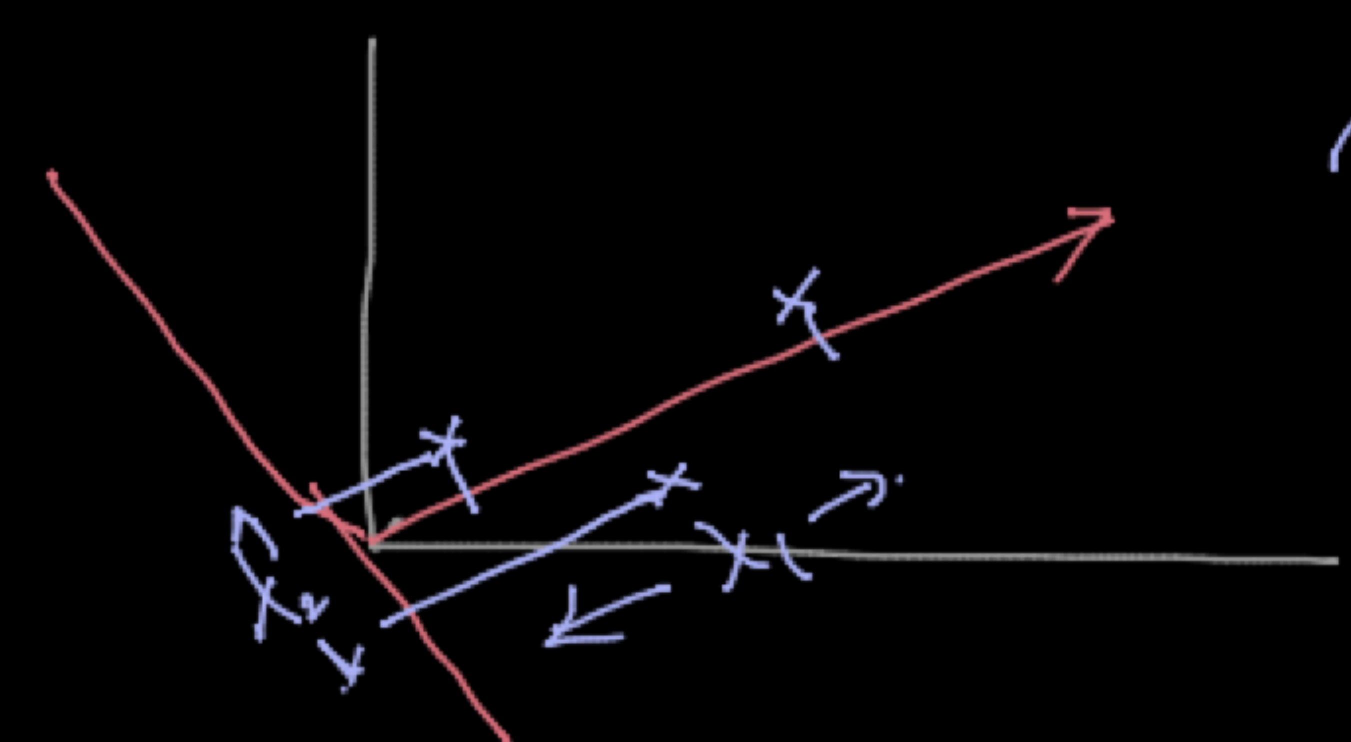
How much info should I retain? → 95%, 90%, 80%



$\lambda_1 = 3$

$\lambda_2 = 0$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_1} = \frac{3}{3+0} = 100\%$$

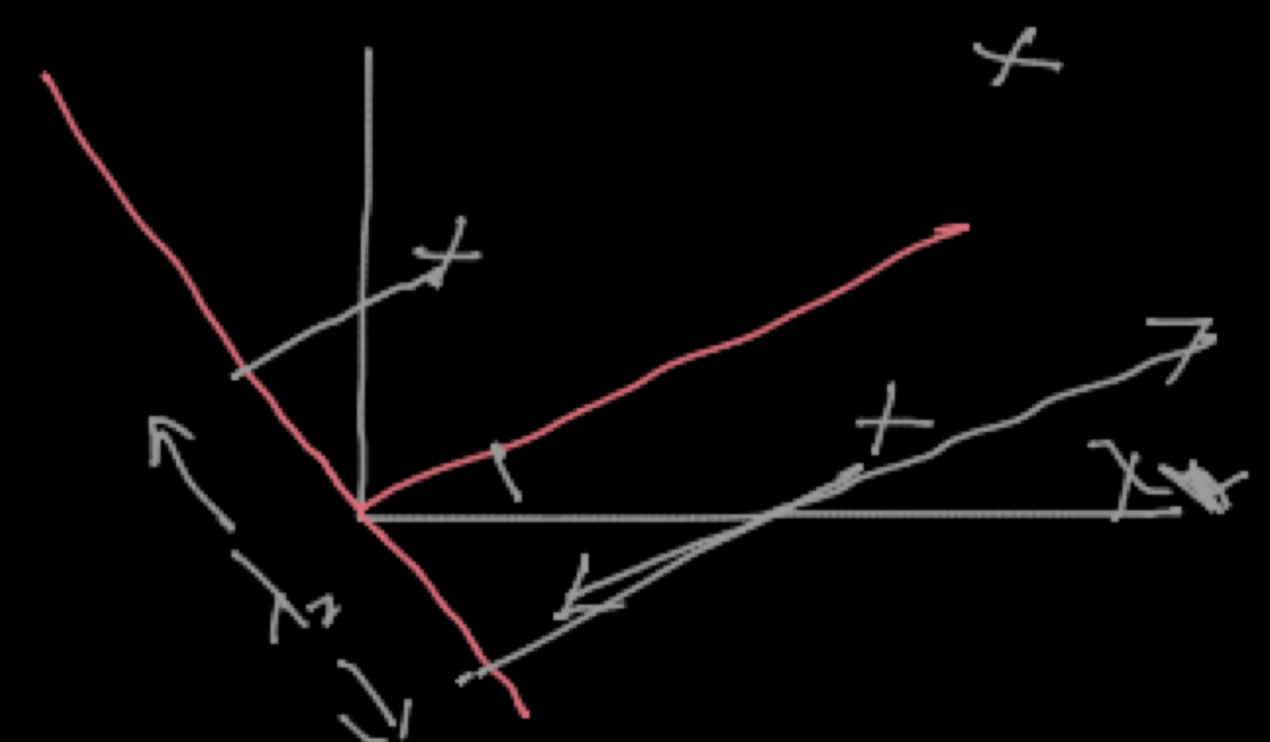


$\lambda_1 = 3$

$\lambda_2 = 1$

$$\Rightarrow \frac{3}{3+1} = 3/4 = 75\%$$

$$\Rightarrow \frac{1}{3+1} = 1/4 = 25\%$$



$\lambda_1 = 3$

$\lambda_2 = 2$

$$\Rightarrow \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{3}{5} = 60\%$$

$$\Rightarrow \frac{\lambda_2}{\lambda_1 + \lambda_2} = 2/5 = 40\%$$

$\vec{V_1}$ $\vec{V_2}$ $\vec{V_3}$ $\vec{V_3}$ $\vec{V_4}$

λ_1 λ_2 λ_3 λ_4 λ_5 λ_6 . . . .

80% 10% 5% 2% 1% 0.9% . . .

90% $\rightarrow$ $\vec{V_1}$ $\vec{V_2}$

95% $\rightarrow$ $\vec{V_1}$ $\vec{V_2}$ $\vec{V_3}$

Retaining 95% of info

d-dimension $\rightarrow$ 3-dimension

Height Weight

x_1 x_2 x_3 x_4 . . . x_d ✓

$\Rightarrow$

↗ ? $\vec{V_1}$	$\vec{V_2}$	$\vec{V_3}$	V_4	. . .	$\vec{V_d}$
PC_1	PC_2	PC_3			

Interpretability.

$$\vec{V_1} = \underline{a_{11}}x_1 + \underline{a_{12}}x_2 + a_{13}x_3 + \cdots \; a_{1d}x_d.$$

$$\vec{V_2} = a_{21}x_1 + \underline{a_{22}}x_2 + a_{23}x_3 + \cdots \; a_{2d}x_d$$

x_1	x_2	x_3	$\ldots$	x_d

X

$\Rightarrow$

PC1	PC2	PC3	$\ldots$	PCd
V_1	V_2	V_3	$\ldots$	V_d

↳ Build models.

PCA → Principal Component Analysis

- dimensionality reduction for building model.
- We control the amt. of information retained.
- We choose the no. of dimensions in the resultant dataset.

zoom.us Meeting View Edit Window Help

t-SNE.

→ t-distributed Stochastic Neighbourhood Embeddings.

→ 'Data Visualization'

–

12 / 12