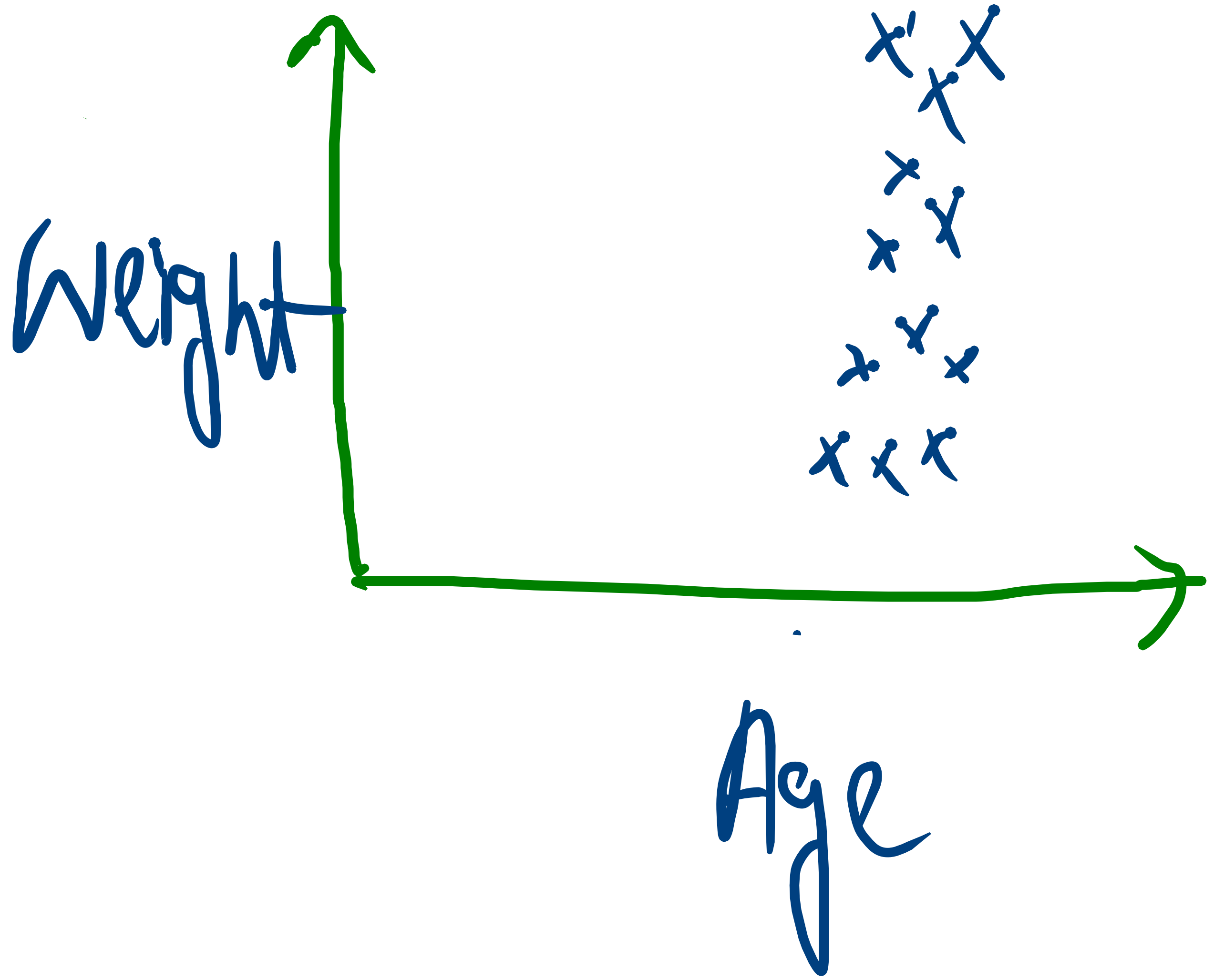
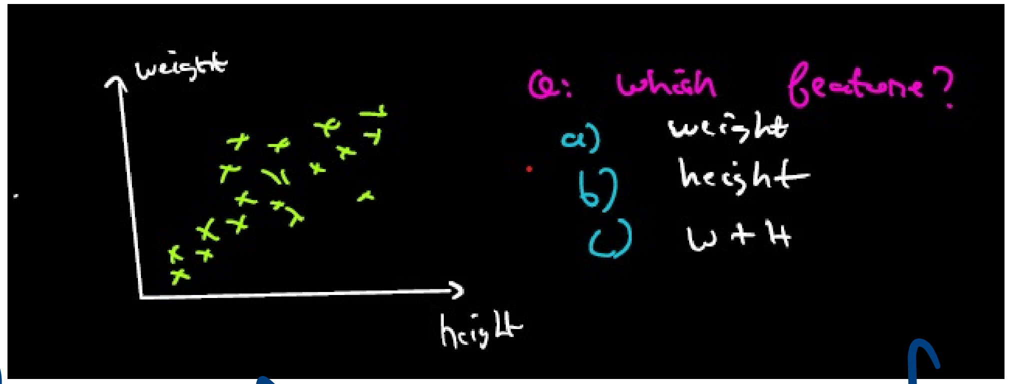
→ Class start at 9:05 pm

PCA (Principal Component Analysis)

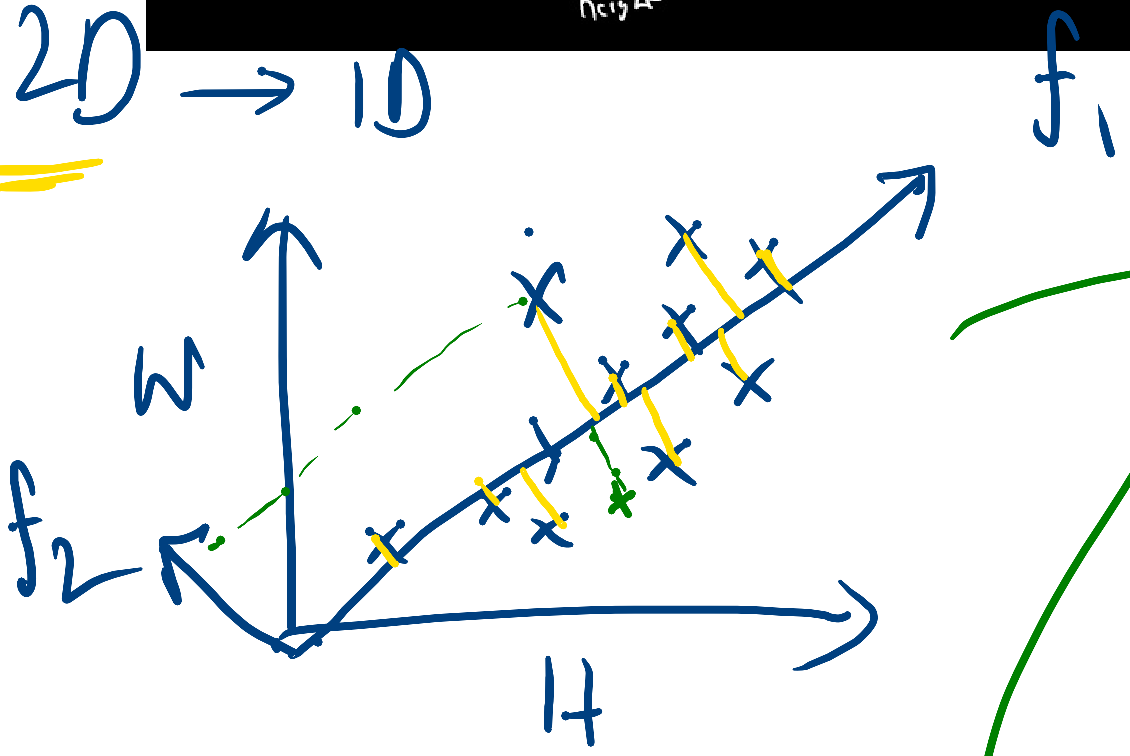
→ Unsuperivised ML algorithm for dimensionality reduction

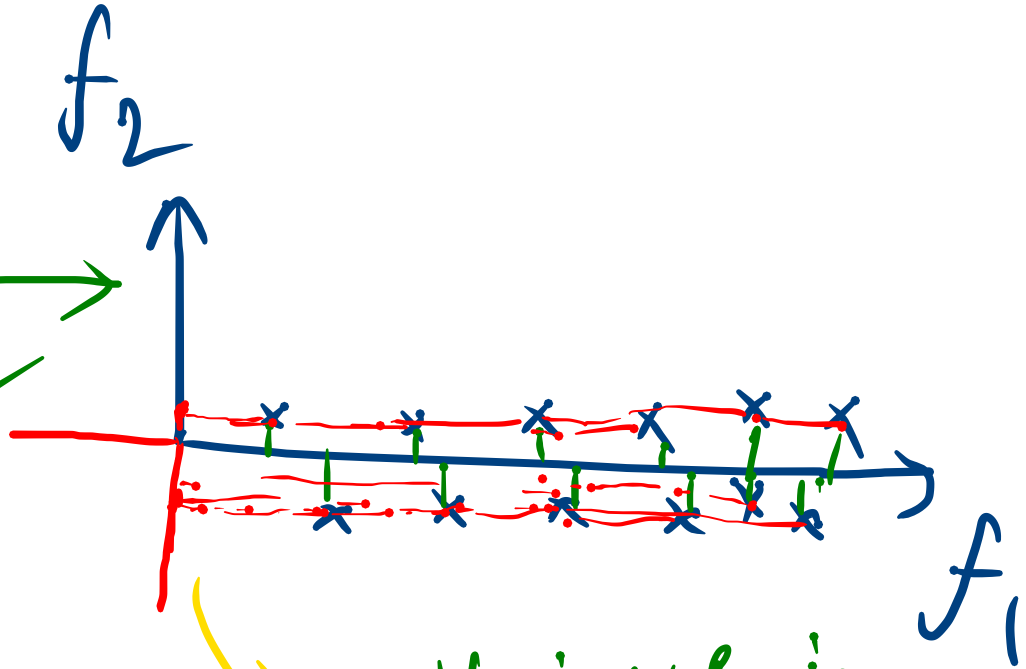
W	A	D
		N
		Y
		N
		Y
		N



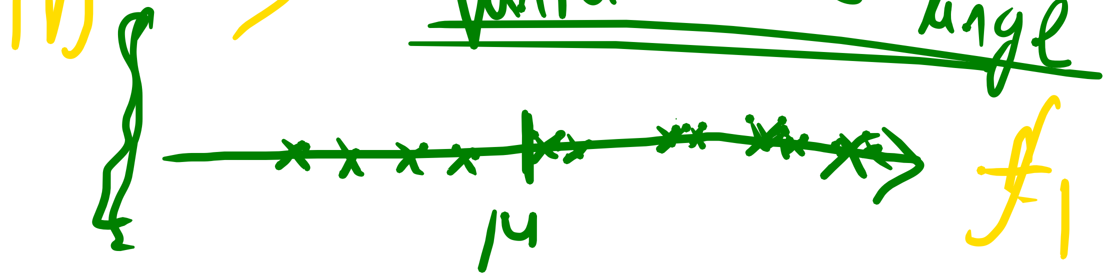


2D → 2D → 1D

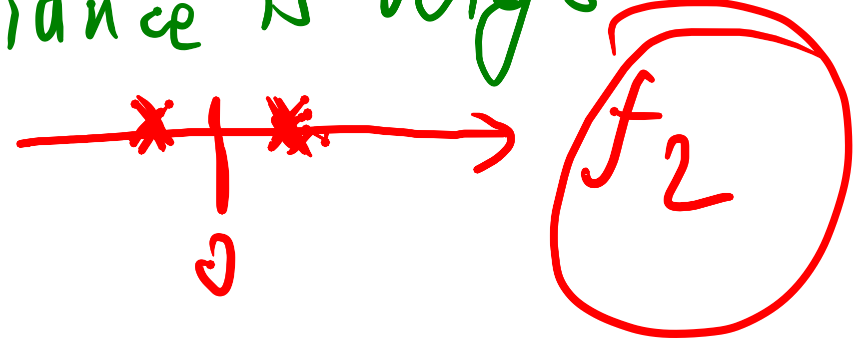




Variance is large



Variance is very small

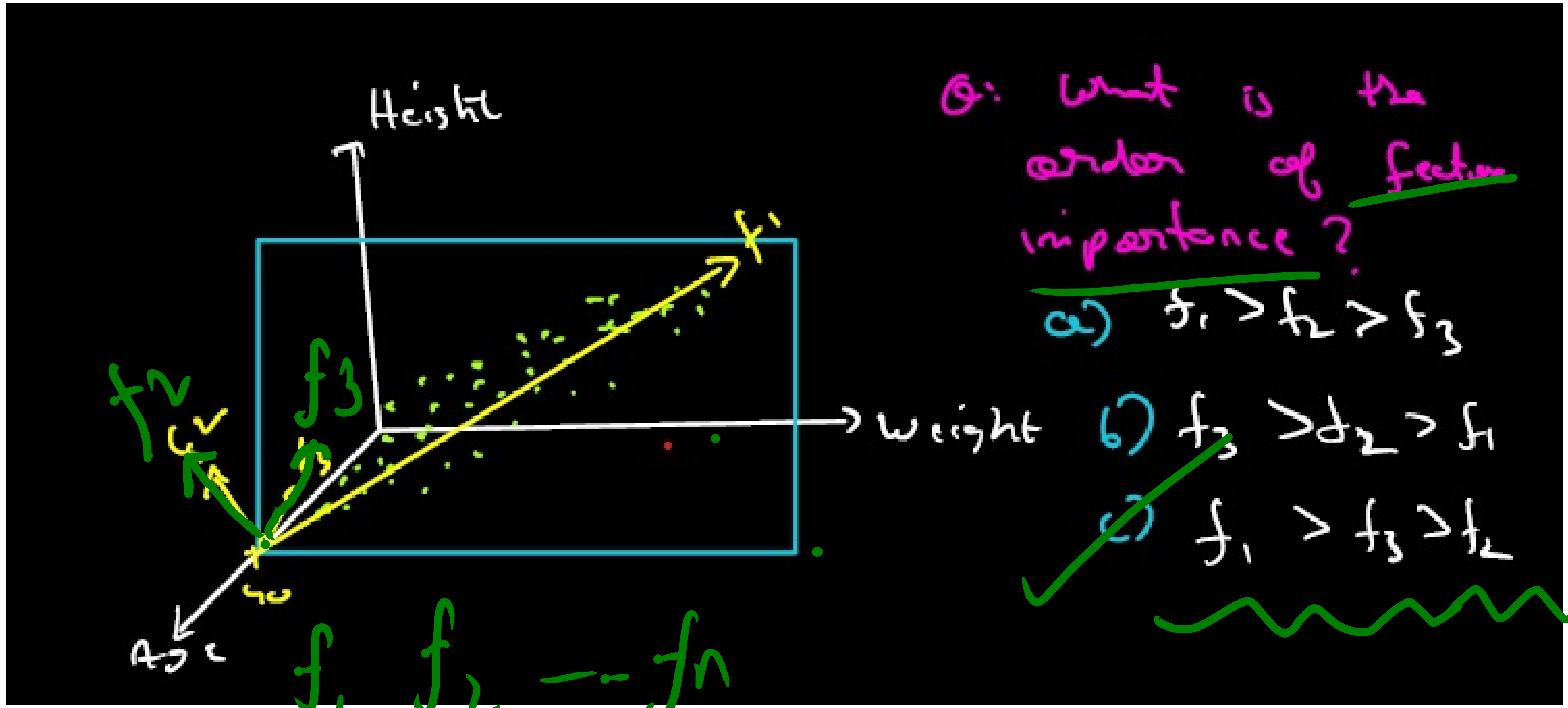


4D → 1D 70%

→ 2D 95%

nD → kD

k < n



$$X \rightarrow \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \begin{array}{c} f_1 \quad f_2 \quad \cdots \quad f_n \\ \left[\begin{array}{c} \\ \\ \\ \\ \end{array} \right] \end{array}$$

→ Max variance

$$\left\{ \begin{array}{l} f_1 \rightarrow 70\% \\ f_3 \rightarrow 20\% \\ f_2 \rightarrow 10\% \end{array} \right.$$

$$f_1 + f_3 \rightarrow 90\%$$

PCA

→ Visualization, EDA

→ Faster ML training + inference

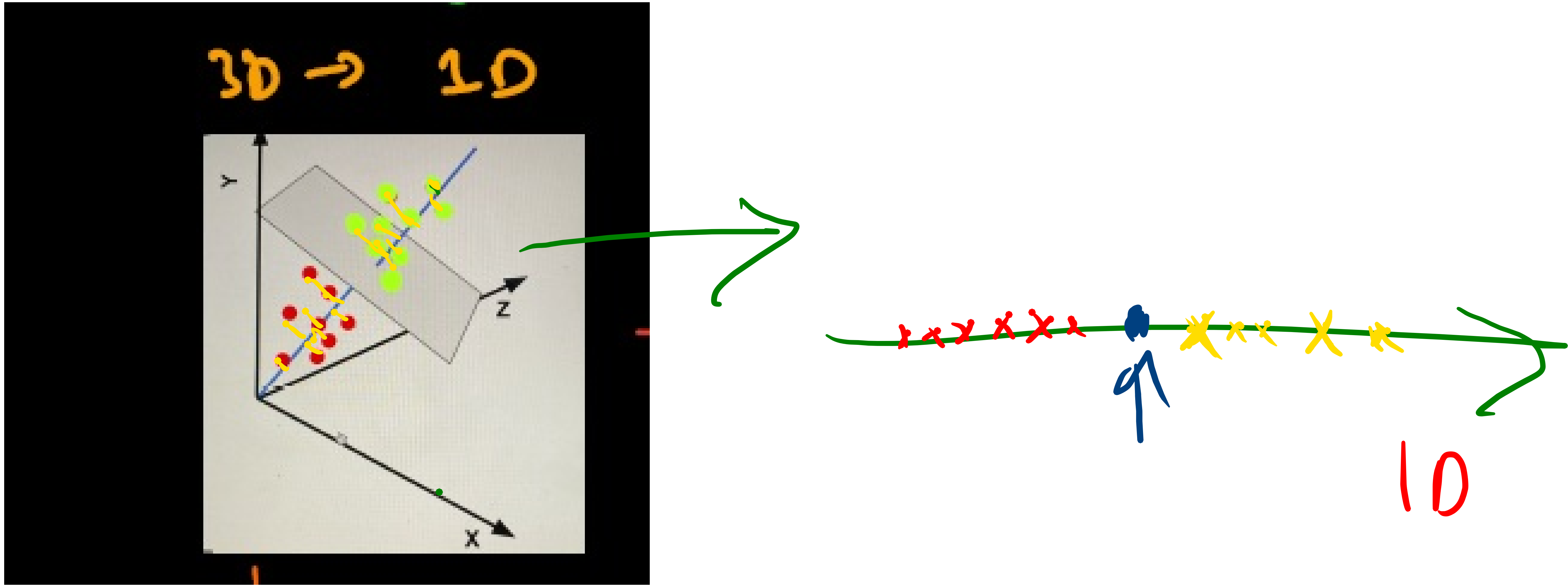
→ Compressing data on image

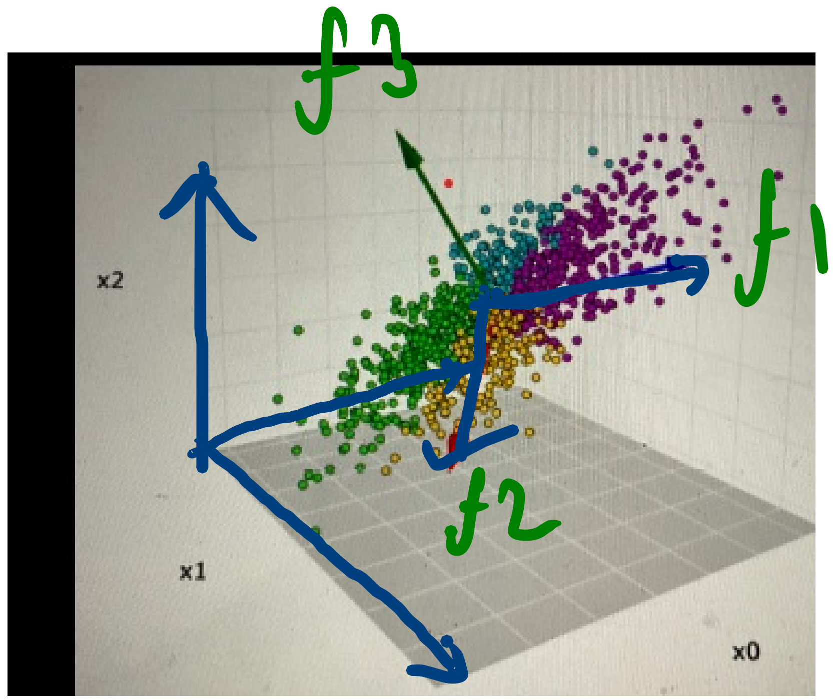
Many redundant info

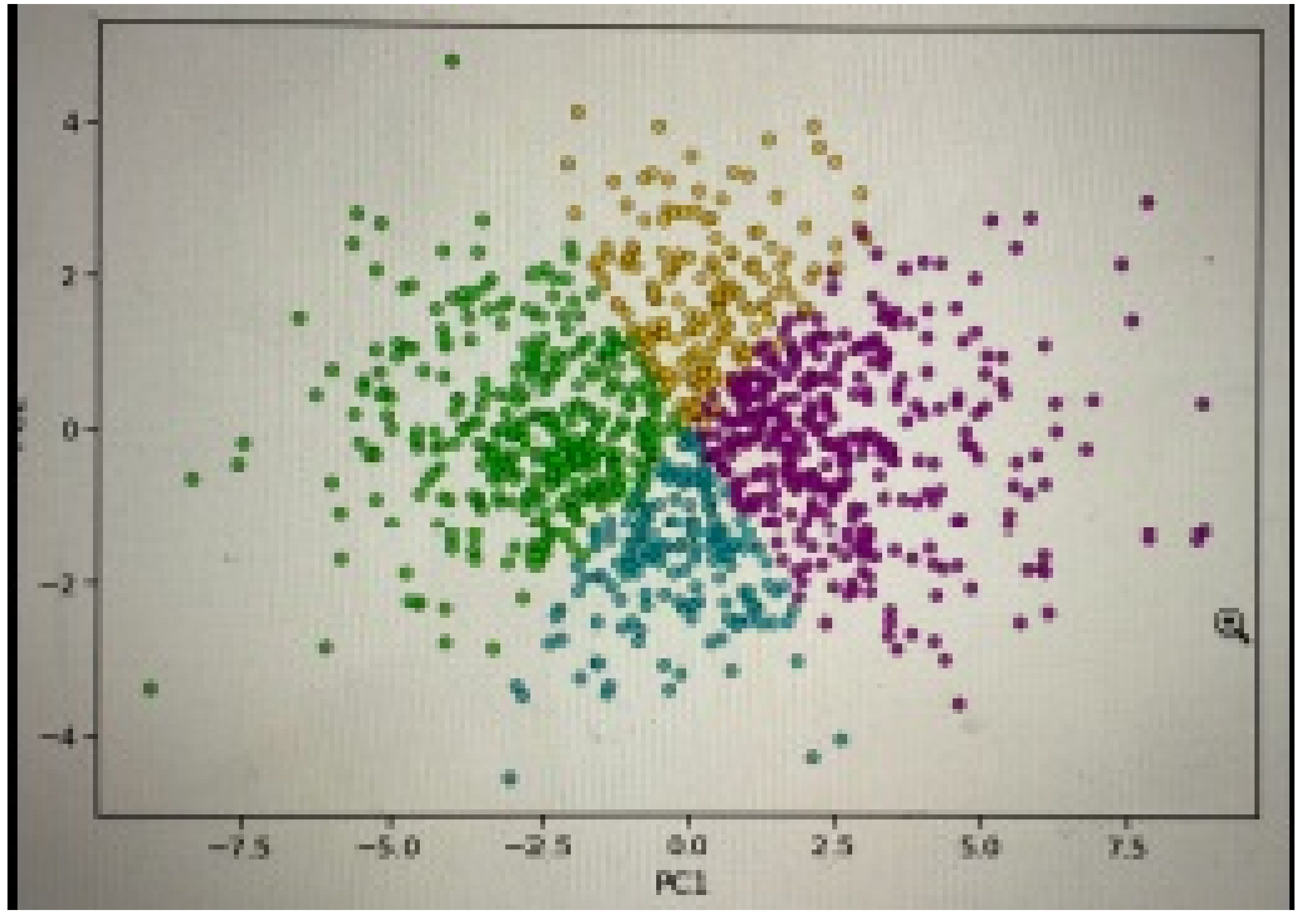
100D → 25D → lot of useful information

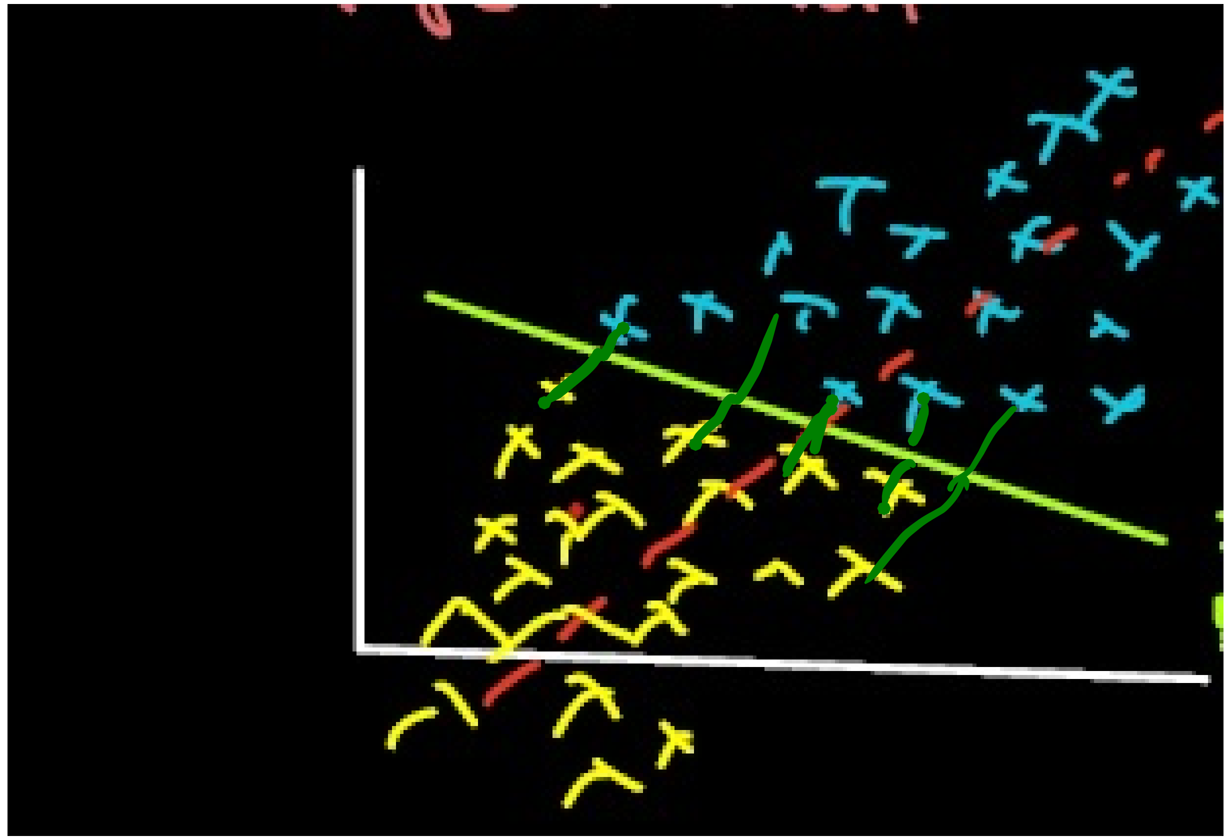
(95%)

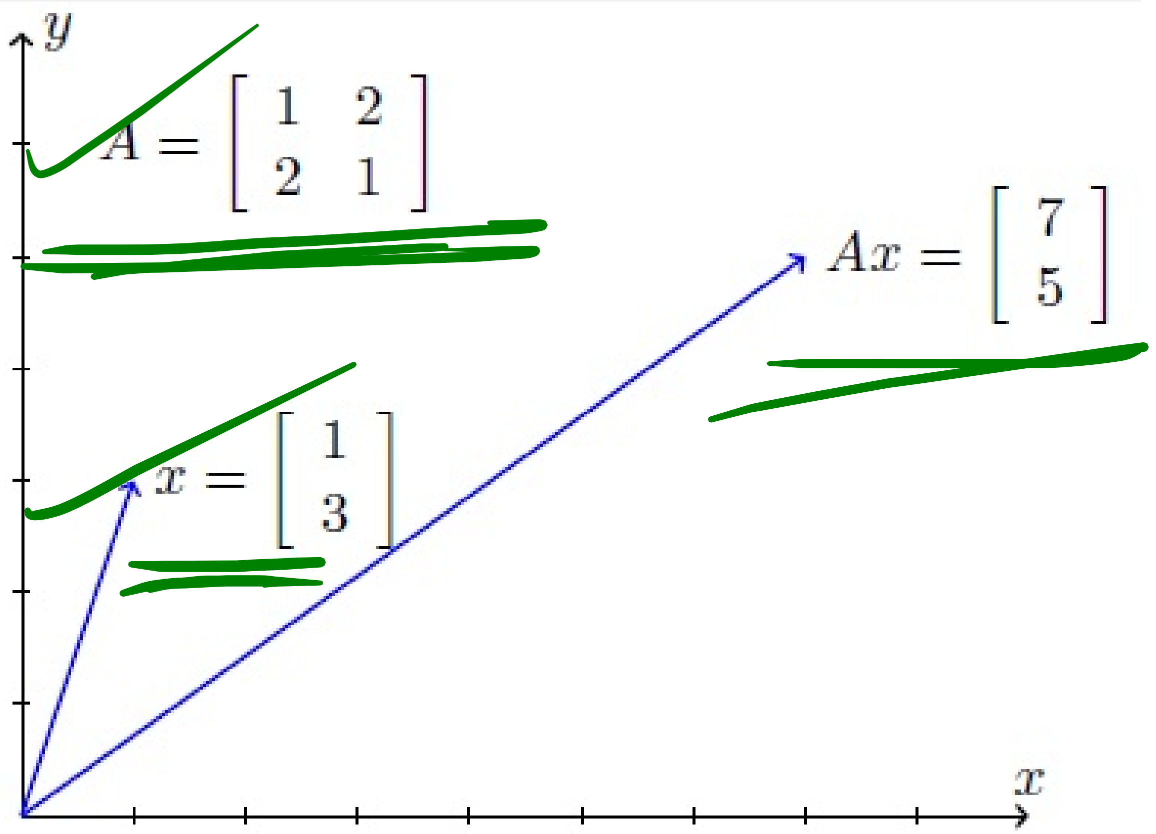
5%


3D → 1D
Y
Z
X
1D


f3
f1
x2
f2
x1
x0


4
2
0
−2
−4
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
PC1

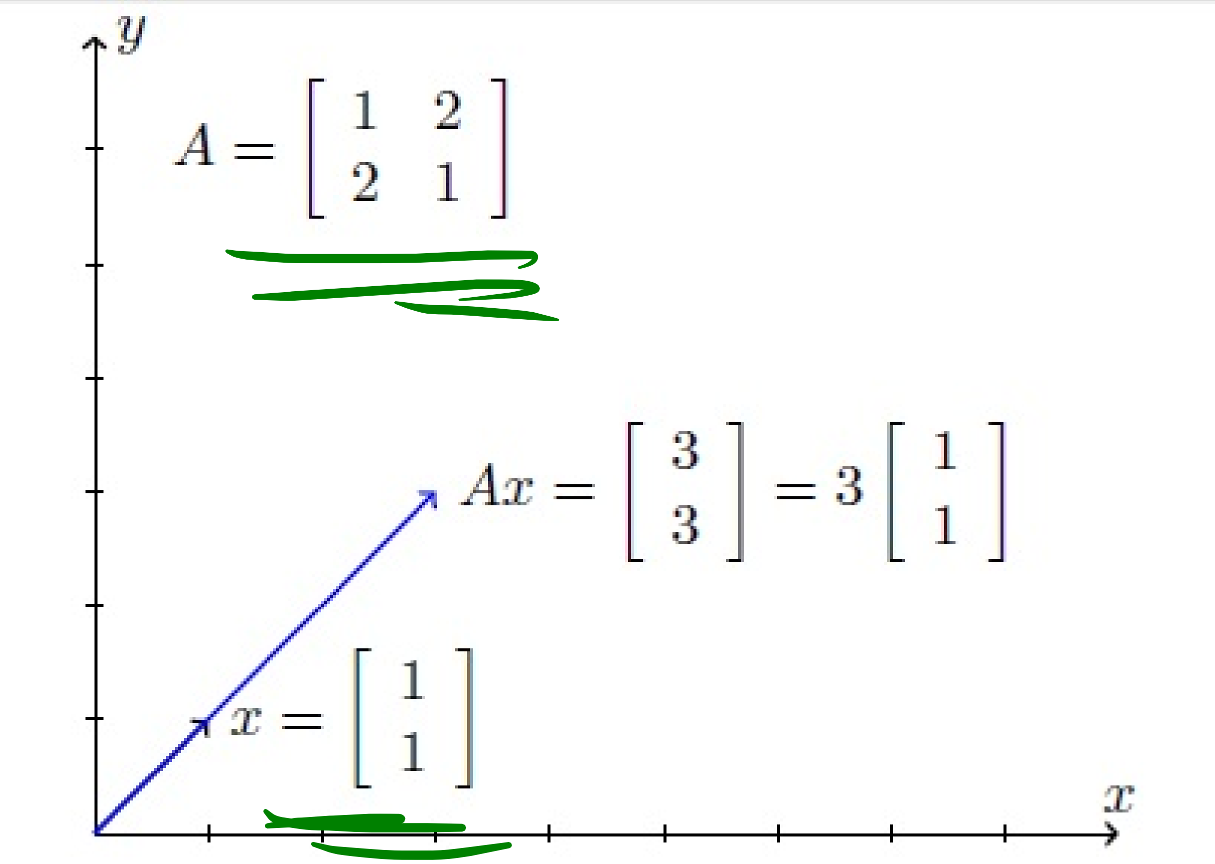




→ What happens when vector is multiplied with Matrix ?

- Scaling
- Change of direction (rotation)

Ax



What happen when we multiply A with x ?

-> Only scaling

-> No rotation

$$Ax = \lambda x$$

Matrix

Eigen vector

Eigen value

A is a square matrix

$u_1 \cdots u_n$ are eigen vector of A

$\lambda_1 \cdots \lambda_n$ are eigen values of A

$Ax = \lambda x$

$\boxed{Au_1 = \lambda_1 u_1}$

Matrix

$$U = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$AU = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ Au_1 & Au_2 & \cdots & Au_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Matrix Matrix

$$= \underbrace{\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}}\underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & \lambda_n \end{bmatrix}}$$

Diagonal matrix

$$= U\Lambda$$

$$\underline{AU = U\Lambda} \rightarrow \boxed{A = U\Lambda U^{-1}} \rightarrow EVD$$

$AUU^{-1} = U\Lambda U^{-1}$

$A = U\Lambda U^{-1}$

$\because$ U^{-1} will exist

↳ linearly independent eigen vector of A

↳ A has all distint eigen values

$$U = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\underline{|U| = 0}$$

$$U^{-1} = \frac{1}{|U|} \text{Adj } U$$

$$\boxed{A = U \Lambda U^{-1}} \boxed{= U \Lambda U^{T}}$$

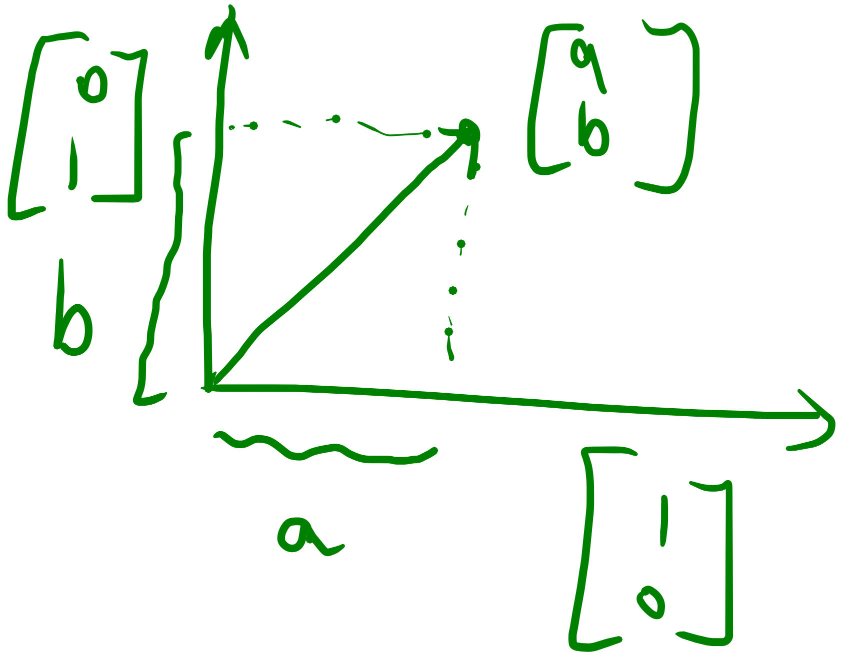
Eigen Value decomposition

$$\boxed{U^{-1} = U^{T}}$$

if A is symetric

→ Break until 10:18 PM

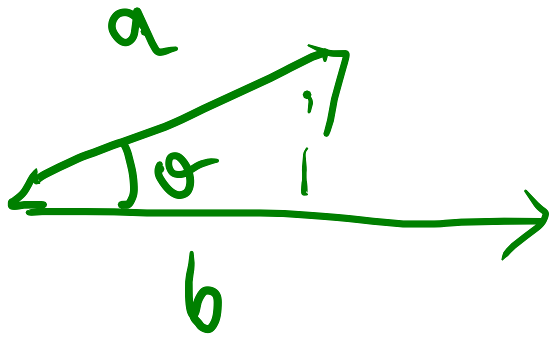
Representation of a point



$$\begin{bmatrix} a \\ b \end{bmatrix} = a\begin{bmatrix} 1 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\frac{\begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{1} = a$$

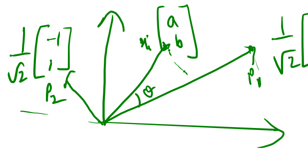
Projection of vector



Project of a on b $= |a| \cos\theta$

$$= |a| \frac{a \cdot b}{|a|\,|b|}$$

$$= \frac{a \cdot b}{|b|}$$



Project x_i on single vector p_1

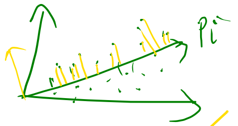
$x_i \cdot p_1 = \frac{a \cdot 1}{\sqrt{2}} + \frac{b \cdot 1}{\sqrt{2}} = \frac{a+b}{\sqrt{2}}$

$x_i \cdot p_2 = a \times \frac{-1}{\sqrt{2}} + b \times \frac{1}{\sqrt{2}}$

$= \frac{-a+b}{\sqrt{2}}$

$x_i \cdot p_i$

$$\underset{m\times d}{X} = \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{d} \; m$$



Zero mean

$\boxed{\hat{x}_i = X p_i}$

$m\times d \quad d\times 1 = m \times 1$

$Var(\hat{x}_i) = Var(X p_i)$

$\frac{1}{m}\sum x_i^2$

$= \frac{1}{m}$

$\frac{\hat{x}_i^T \hat{x}_i}{m}$

$= \frac{(X p_i)^T (X p_i)}{m}$

$A = X^T X$

$= \frac{p_i^T (X^T X) p_i}{m} = \frac{p_i^T A p_i}{m}$

$\max Var(\hat{x}_i) = \max \frac{p_i^T A p_i}{m}$

$= \max p_i^T A p_i$

such $\|p_i\| = 1$

$= \max p_i^T A p_i - \lambda(\|p_i\| - 1)$

$L = p_i^T A p_i - \lambda(p_i^T p_i - 1)$

$\frac{\partial L}{\partial p_i} = 2 A p_i - \lambda(2 p_i) = 0$

$\boxed{A p_i = \lambda_i p_i}$ → Characteristics Equation

$(X^T X)$

p_i is eigen vector of A

$\parallel$

$X^T X$

$\max(Var(\hat{x}_i)) = p_i^T A p_i = p_i^T \lambda_i p_i = \lambda_i p_i^T p_i$

$= \lambda_i \quad \because p_i^T p_i = 1$

Inference

$$\max(\mathrm{Var}(x_i)) = \max \lambda_i$$

Orthogonal

$$\begin{cases} p_i p_j = 0 & i \neq j \\ = 1 & i = j \end{cases}$$

where p_i is eigen vector of $X^T X$ $\quad \|p_i\| = 1$

λ_i eigen value of $X^T X$

$$L = p^T A p - \lambda (p^T p - 1)$$

$$\frac{\partial L}{\partial p} = Ap + (p^T A)^T - \lambda (p + (p^T)^T)$$

$$\frac{\partial}{\partial x}(x^T a) = a$$

$$\frac{\partial}{\partial x}(a^T x) = a$$

$$= Ap + A^T p - \lambda (p + p)$$

$$= (A + A^T) p - 2p$$

$$\boxed{A^T = A}$$

$$= 2Ap - 2p$$

$A = X^T X \rightarrow$ Is A symetric ?

$X = m \times d$

$X^T \quad d \times m$

$X^T X = d \times d$

1) Yes

2) No

$$A^T = (X^T X)^T = X^T X = A$$

$P_1 \sim 70\%$

$$\underbrace{\overset{m\times 1}{\hat{X}_i}}_{\text{vect}} = \overset{m\times d}{X}\, \overset{d\times 1}{p_i}$$

$$\underset{m\times d}{\hat{X}} = \underset{m\times d}{X}\, \overset{d\times d}{P}$$

$$\hat{X} = XP$$

$$\underbrace{Cov(\hat{X})} = \frac{1}{m}\hat{X}^T\hat{X} = \frac{1}{m}(XP)^T(XP)$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \\ & \sigma_2^2 & & \\ & & \sigma_2^2 & \end{bmatrix}$$

$$= \frac{1}{m} P^T(X^TX)P$$

$$= \frac{1}{m} \underbrace{P^TAP}$$

$$= D = \Lambda$$

$$\begin{bmatrix} \ddots & 0 \\ 0 & \ddots \end{bmatrix}$$

$$\underline{A = U\Lambda U^T}$$

$$= \underline{\underset{\text{eige vector}}{P}\,\Lambda\, P^T}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

$$\underline{P^TAP = \Lambda}$$

$$A = U\Lambda U^T$$

$$= U\Lambda U^T$$

Conclusion

→ Data has high variance on that dimension

1st proof

Eigen vector of $X^T X$

$X_{m \times d}$

→ Dimensions should uncorrelated

→ 2nd proof

$P^T P$

$p_1 \cdot p_2 = 0$

P is eigen vect

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ p_1 & p_2 & \cdots & p_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\underbrace{\max(\mathrm{var}(\hat{X}_i))} = \max \underbrace{p_i^T X^T X p_i}$$

$$\underbrace{\|P_i\| = 1}$$

$$= \underbrace{\max \lambda_i}$$

PCA → Original data → n dimension

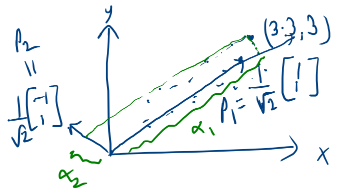
$\downarrow$

$\underbrace{\text{K dimension}}$

→ Select the dimension (basis) consisting of top-K eigen vector

→ The n-K dimension contribute very little to reconstruction error

↳ Variance is minimum



$$x_i = \begin{bmatrix}3.3\\3\end{bmatrix} = \alpha_1 p_1 + \alpha_2 p_2 \qquad p_1^{-1} = p_1^T$$

$$= \frac{\alpha_1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} + \alpha_2\begin{bmatrix}-1\\1\end{bmatrix}$$

$$\alpha_1 = x_i^T p_1 = x_i p_1^T = \begin{bmatrix}3.3 & 3\end{bmatrix}\begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \frac{6.3}{\sqrt{2}}$$

$$\alpha_2 = x_i^T p_2 = x_i p_2^T$$

$$= \begin{bmatrix}3.3 & 3\end{bmatrix}\begin{bmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$

$$= \frac{-0.3}{\sqrt{2}}$$

$$\alpha_1 = \frac{6.3}{\sqrt{2}} \qquad \alpha_2 = \frac{-0.3}{\sqrt{2}}$$

$$x_i = \alpha_1 p_1 + \alpha_2 p_2$$

$$\begin{bmatrix}3.3\\3\end{bmatrix} = \underbrace{\frac{6.3}{\sqrt{2}}\begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}}\ \underbrace{\frac{-0.3}{\sqrt{2}}\begin{bmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}}$$

Initially it was represented in x & y.
After transformation p_1 & p_2.

2D $\longrightarrow$ 2D

2D $\longrightarrow$ 2D $\longrightarrow$ 1D

Only 1D

$$x_{reduced} = \alpha_1 p_1 = \frac{6.3}{\sqrt{2}}\begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \begin{bmatrix}6.3/2\\6.3/2\end{bmatrix} = \begin{bmatrix}3.15\\3.15\end{bmatrix}$$

$$x = \begin{bmatrix}3.3\\3\end{bmatrix} \qquad x_{reduce}$$

$$x = \begin{bmatrix} 3.3 \\ 3 \end{bmatrix} \qquad x_{reduced} = \begin{bmatrix} 3.15 \\ 3.15 \end{bmatrix}$$

$$\sum (x - x_{reduced})^2 = (3.3 - 3.15)^2 + (3 - 3.15)^2$$

$$= (0.15)^2 + (0.15)^2$$

If P_2 is used

$$x_{reduced} = \frac{-0.3}{\sqrt{2}} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.3/2 \\ -0.3/2 \end{bmatrix} = \begin{bmatrix} 0.15 \\ -0.15 \end{bmatrix}$$

$$\text{Error} = \sum (x - x_{reduced})^2 = \left(\begin{bmatrix} 3.3 \\ 3 \end{bmatrix} - \begin{bmatrix} 0.15 \\ -0.15 \end{bmatrix} \right)^2$$

→ Consider input data is 5 dim

5 eigen value

$$\left.\begin{matrix} 1.0 \\ 0.7 \\ 0.6 \end{matrix}\right\} K = 3$$

$$\left\{\begin{matrix} 0.3 \\ 0.1 \end{matrix}\right\} n-K = \text{Error}$$

$$\text{Error} = \frac{0.3 + 0.1}{1 + 0.7 + 0.6 + 0.3 + 0.1}$$

$$\text{Explained Variance ratio} = \frac{1}{2.7}, \frac{0.7}{2.7}, \frac{0.6}{2.7}, \frac{0.3}{2.7}, \frac{0.1}{2.7}$$

$K=1$ 37%

$K=2$ 62.9%

$K=3$ 85%

$K=4$ 96.2%

$K=5 = n$ 100%

$$X \rightarrow m \Big\{ \overset{d}{\longleftrightarrow}$$

$$\text{Cov} \longrightarrow \underbrace{d \times d} \qquad m \times m$$

$$\text{Cov} = \frac{(X^T X)}{m} \longrightarrow d \times d$$

$$U = \text{Eigen vector of } (X^T X)$$

$$\max(\text{variance } \hat{x}^T \hat{x}) = \max(\lambda_i)$$

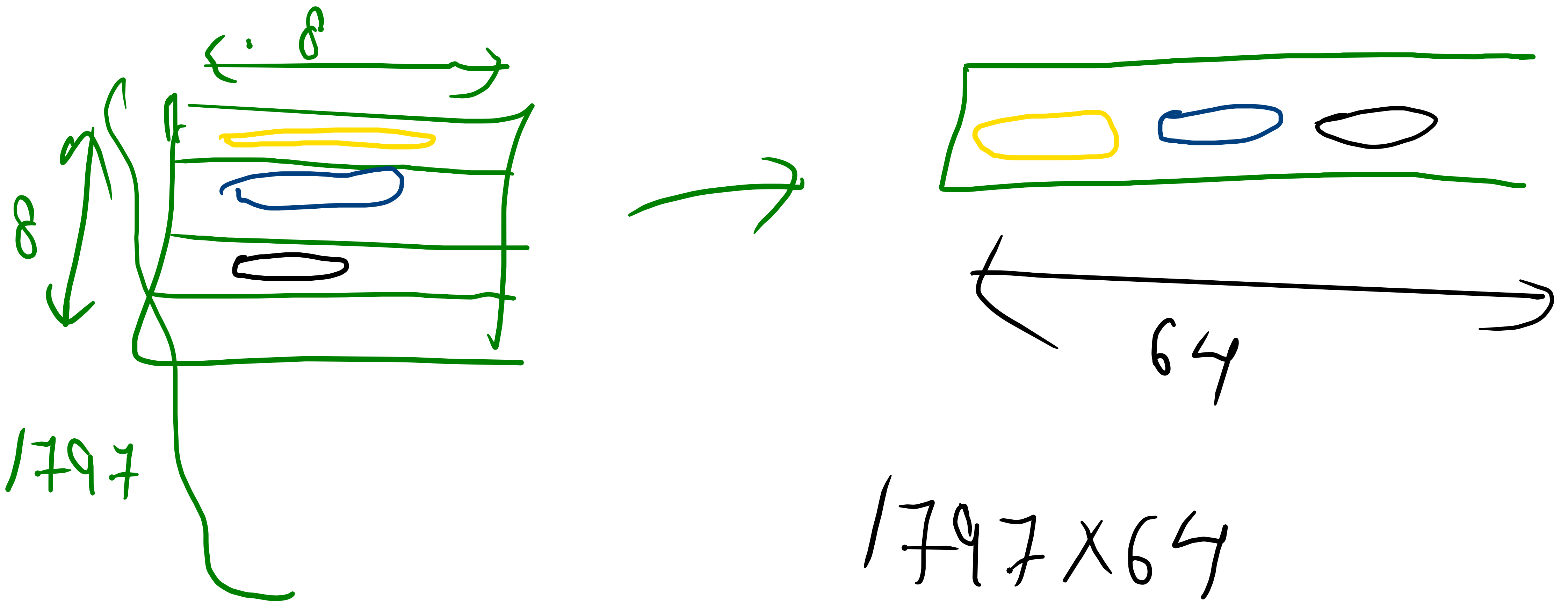
$$X = \alpha_i p_1 + \alpha_2 p_2 \;--\; \alpha_n p_n$$

$$= \alpha_i p_1 + \alpha_2 p_2 -- \alpha_k p_k \Big\} \text{ k-component}$$

$$\underline{[\alpha_1 \ \alpha_2 \ -- \ \alpha_k]} = X \cdot [p_1 -- p_k]^T$$

Break until 10:25pm

Image PCA



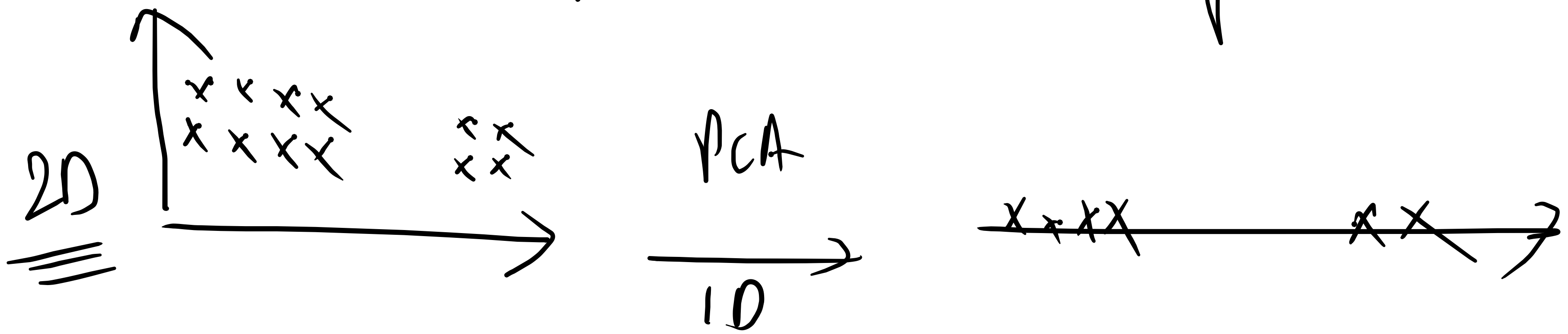
1797X64

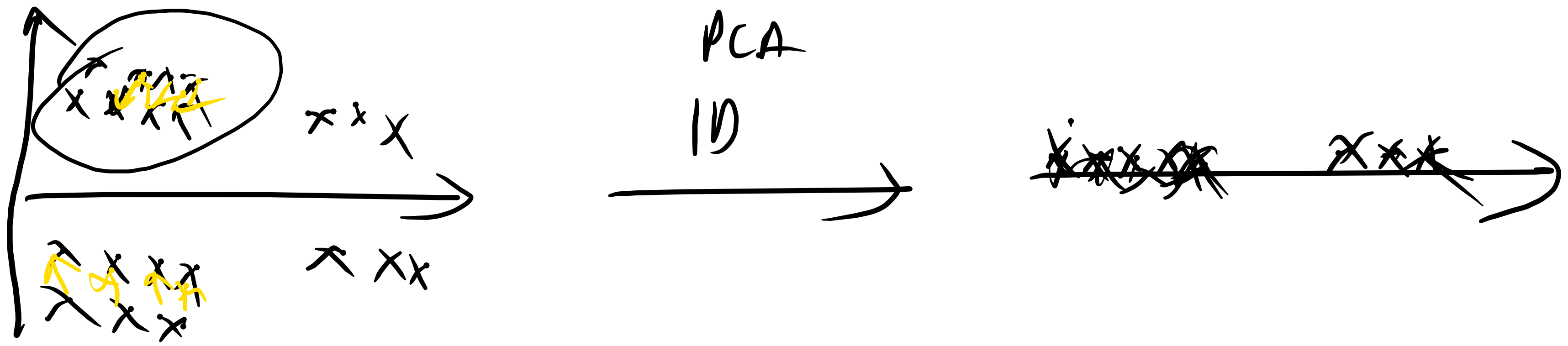
		Loss
64	100%	
32	96%	4%
16	84%	16%
2	28%	

Disadvantage

1. Only carry out linear mapping

2. Does not preserve local neighbours





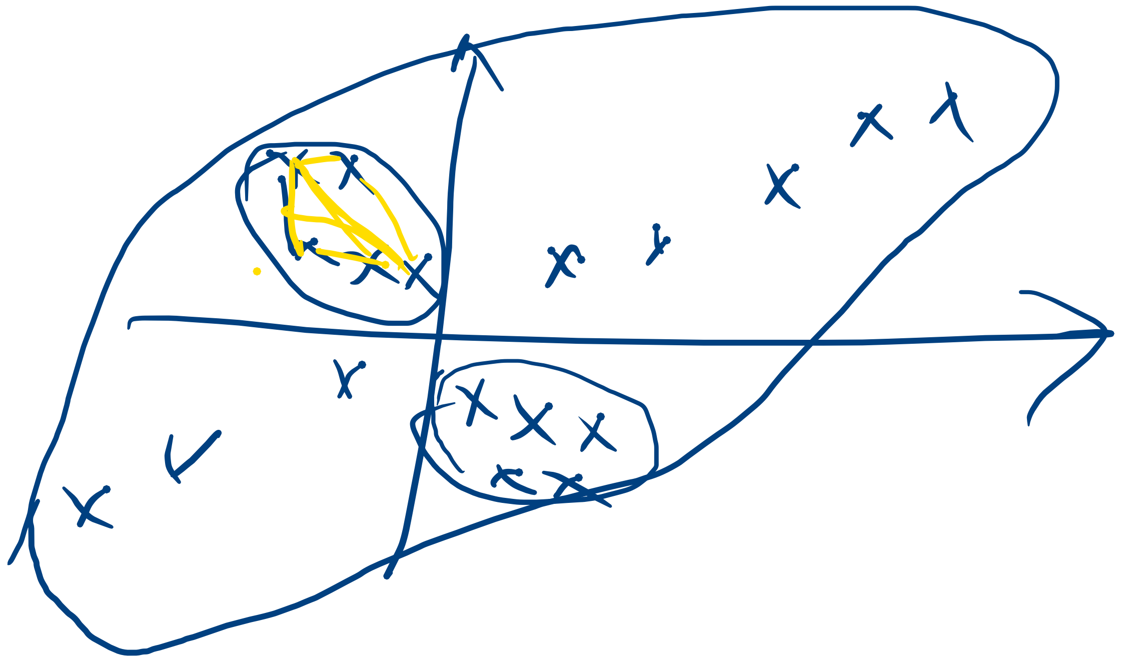
→ Points which are close in high dimension to be close in lower dimension.

t-SNE (t-dist Stochastic Neighborhood Embedding)

↳ Geoffrey Hinton

→ Try to preserve neighborhood information

→ Used for data visualization



→ Both cluster will be represented in the same region axis=1

$$A = \left[[1 \quad 2, 3], \; [4, 5, 6] \; [1, 3] \right]$$

axis=0

np.sum(A, axis=0)

(3, 2)

(1, 2, 3)