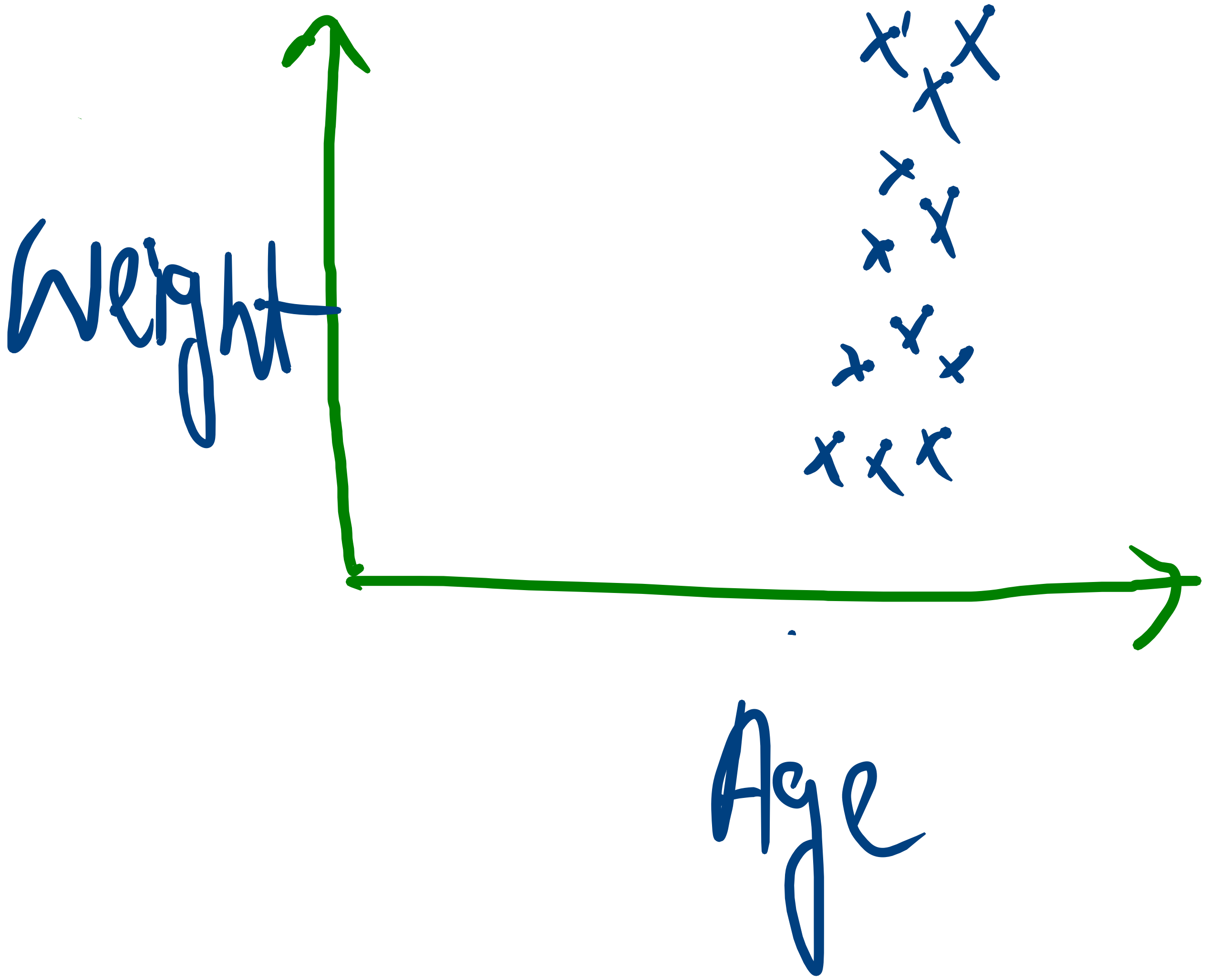
→ Class start at 9:05 pm

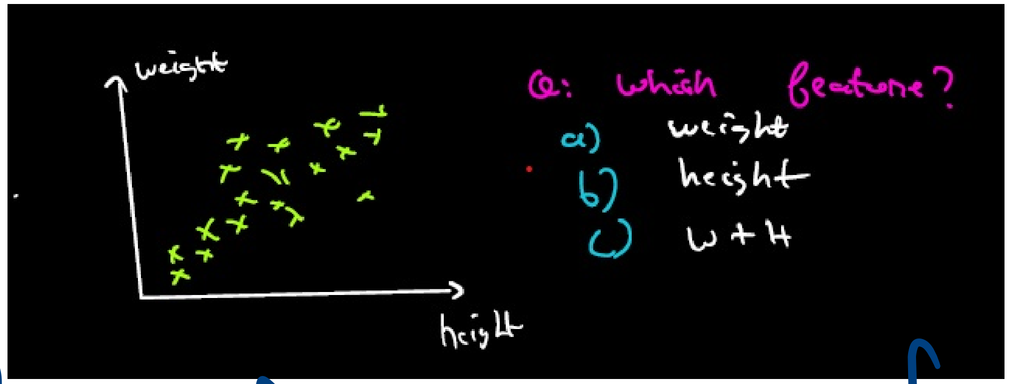
PCA (Principal Component Analysis)

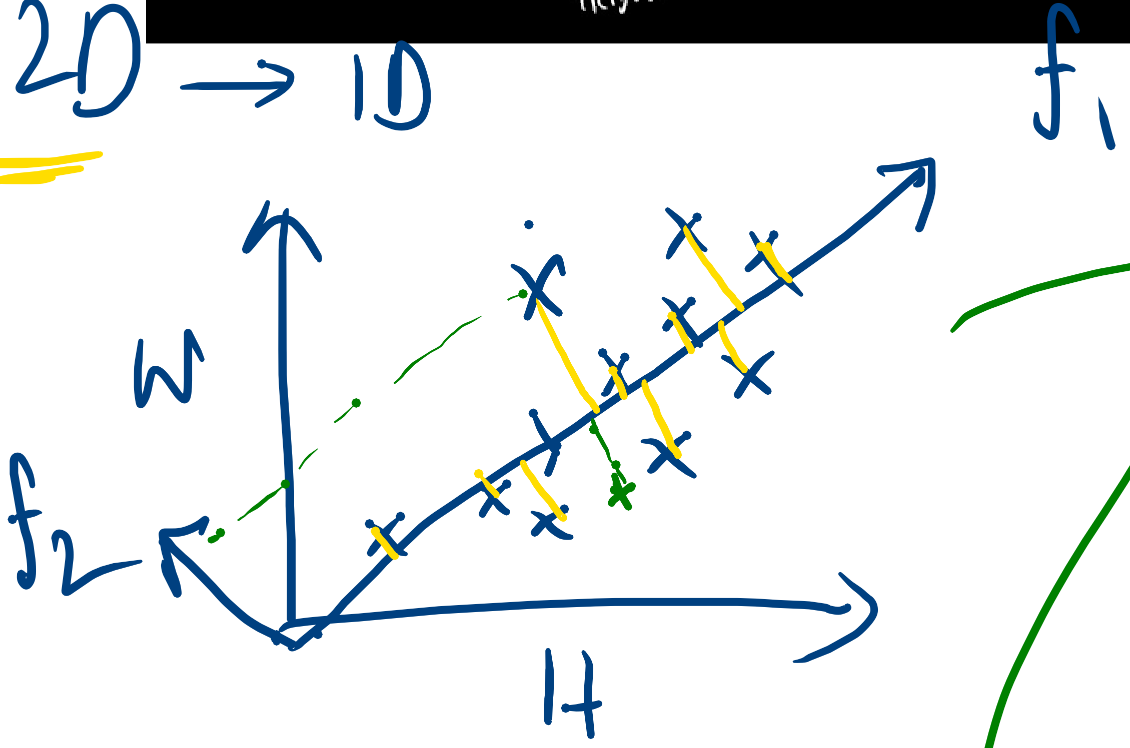
→ Unsuperivised ML algorithm for dimensionality reduction

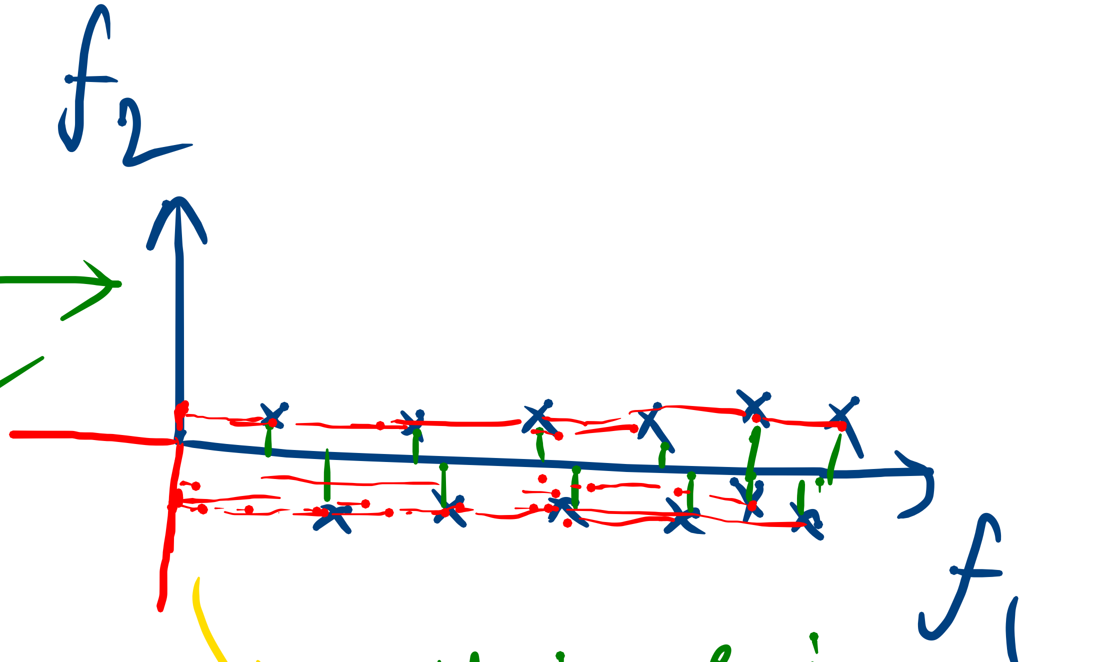
W	A	D
⋮	⋮	N
		Y
⋮	⋮	N
		Y
		N



$2D \rightarrow 2D \rightarrow 1D$

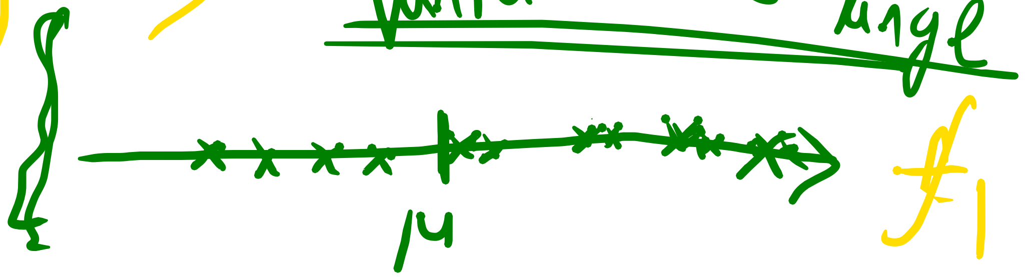




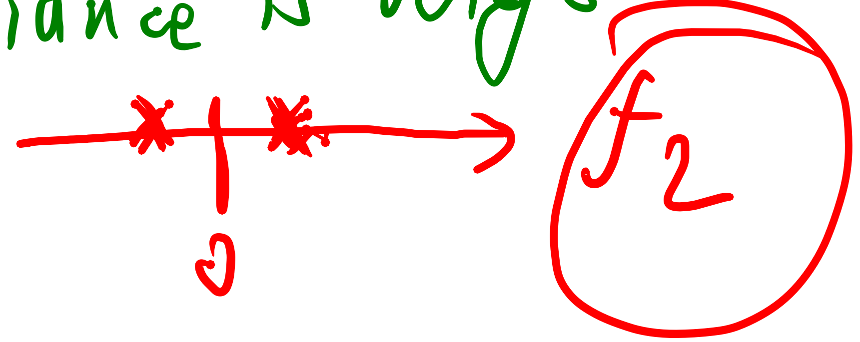


Variance is large

1D



Variance is very small

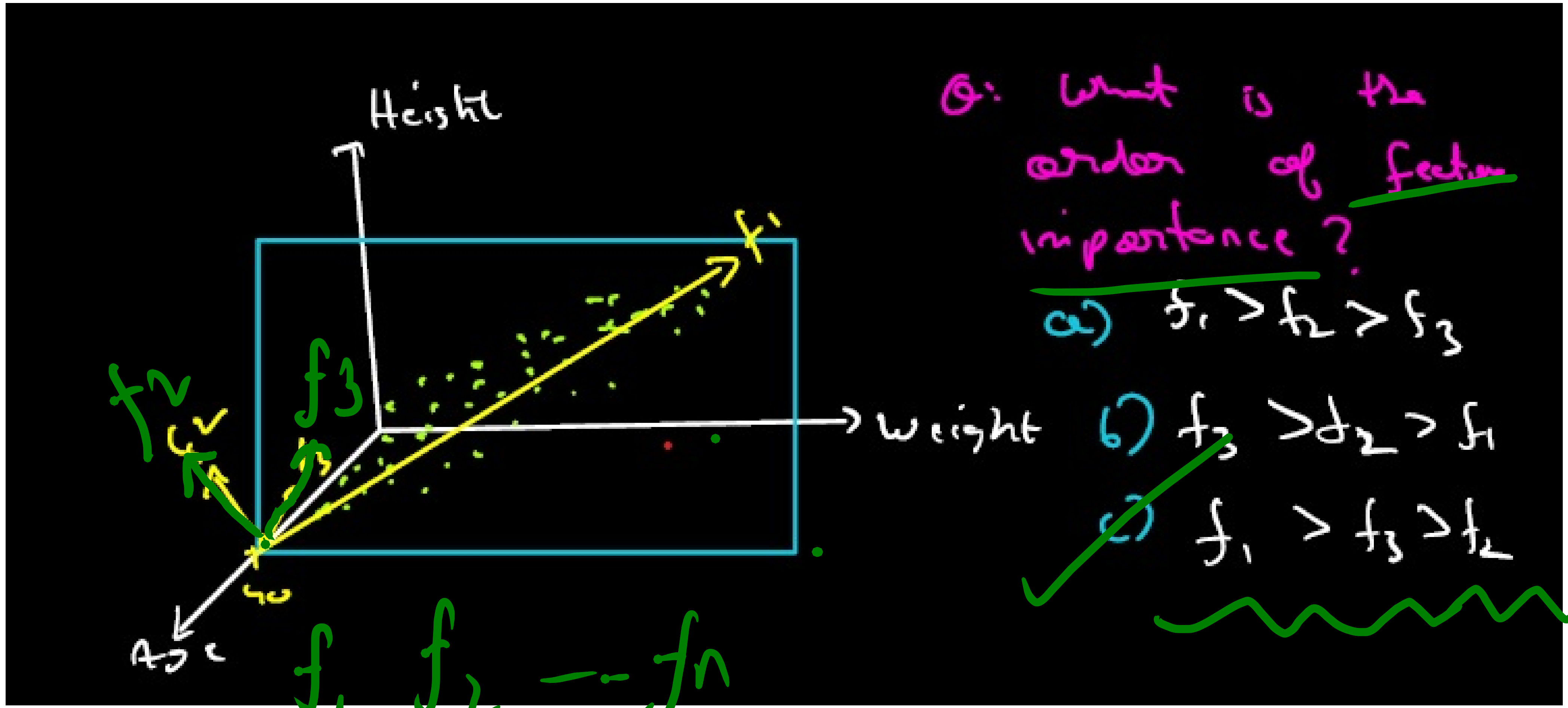


$4D \rightarrow 1D$ 70%

$\rightarrow 2D$ 95%

$nD \longrightarrow kD$

$k < n$



$$X \rightarrow \begin{array}{c|c} & f_1 \quad f_2 \quad \cdots \quad f_n \\ \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix} & \begin{bmatrix} \quad & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array}$$

$\rightarrow$ Max variance

$$\begin{cases} f_1 \rightarrow 70\% \\ f_3 \rightarrow 20\% \\ f_2 \rightarrow 10\% \end{cases}$$

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

PCA

→ Visualization, EDA

→ Faster ML training + inference

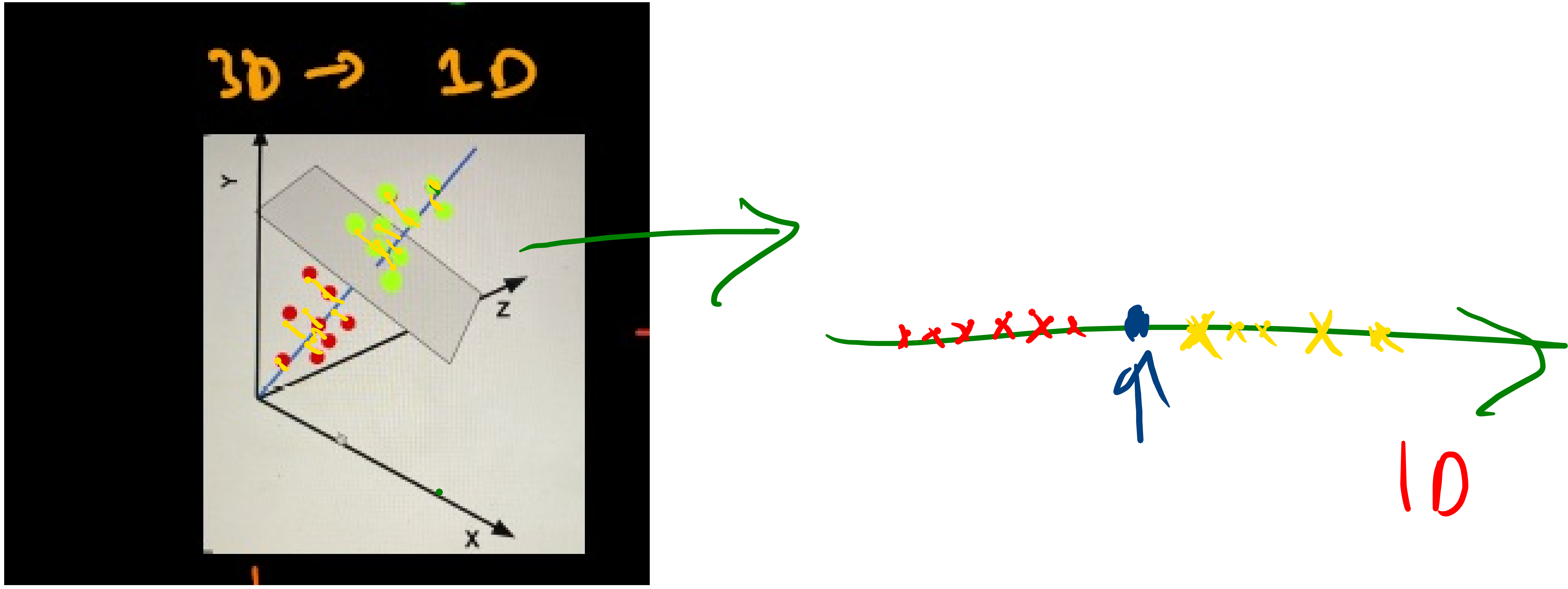
→ Compressing data on image

Many redundant info

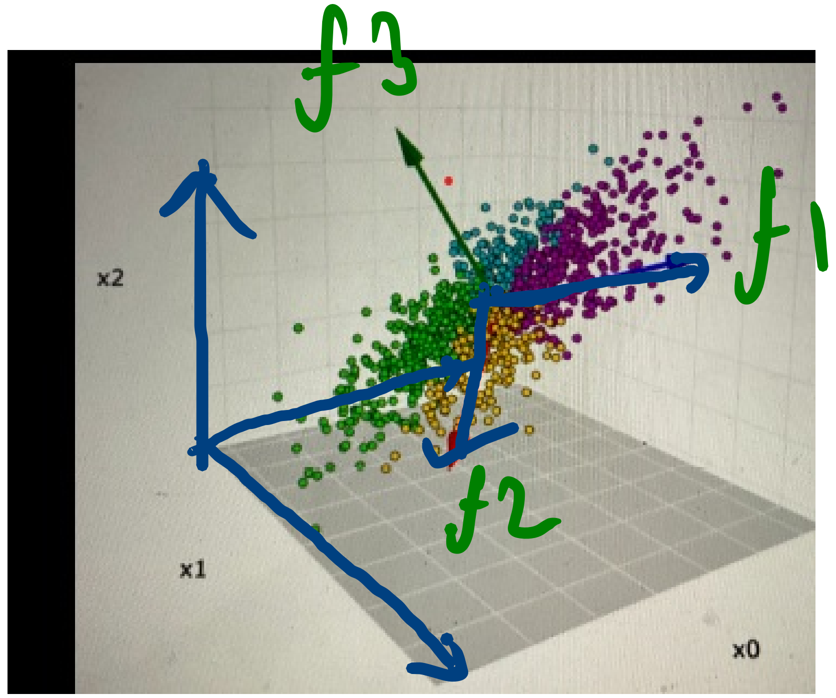
100D → 15D → lot of useful information

(95%)

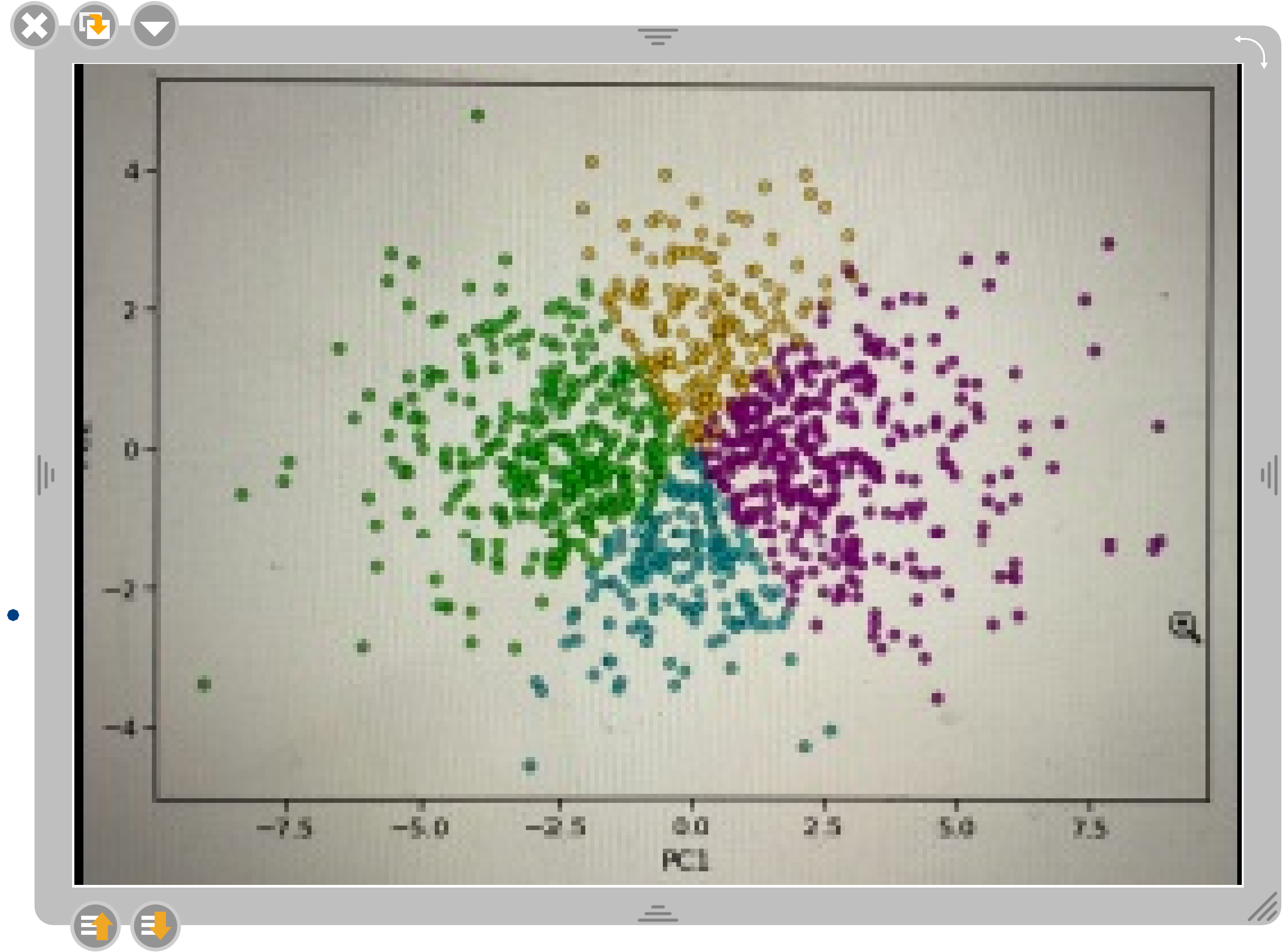
5%



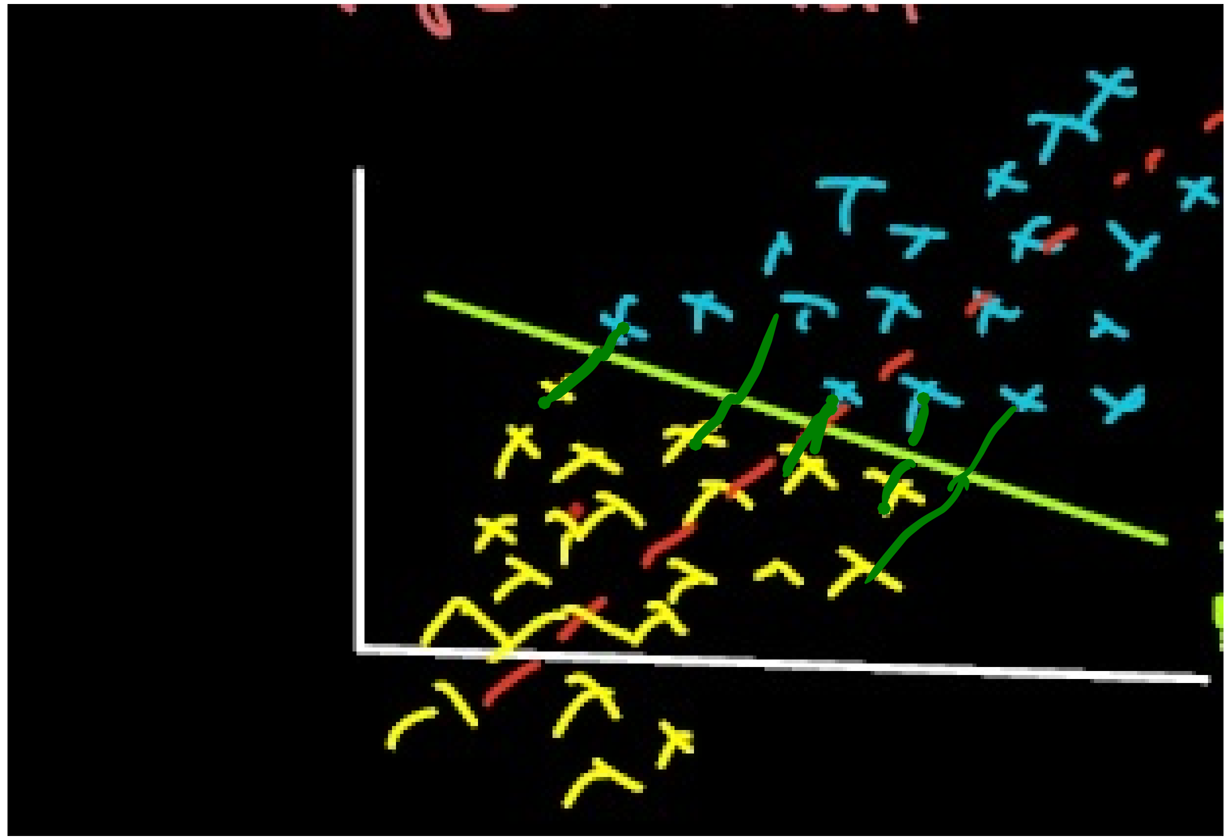
3D → 1D
Y
Z
X
1D

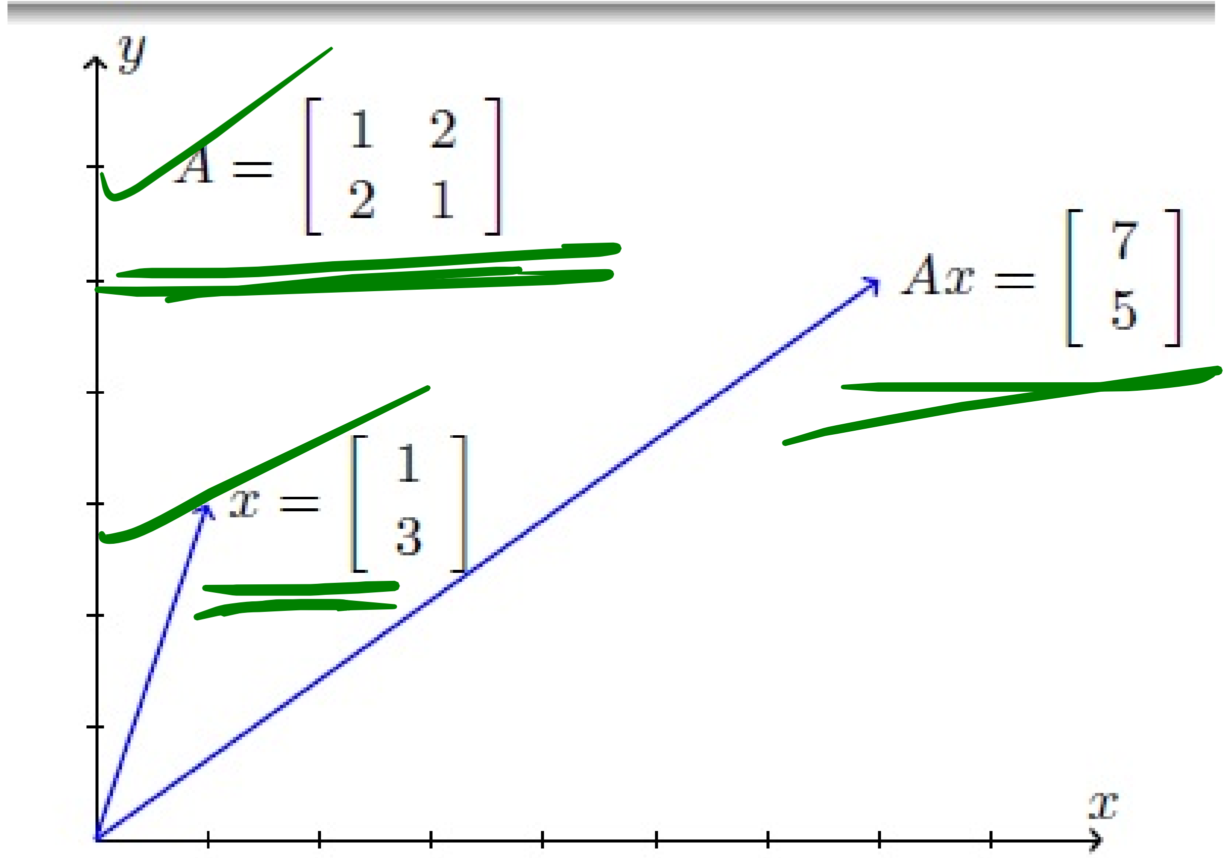


f3
f1
x2
f2
x1
x0



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

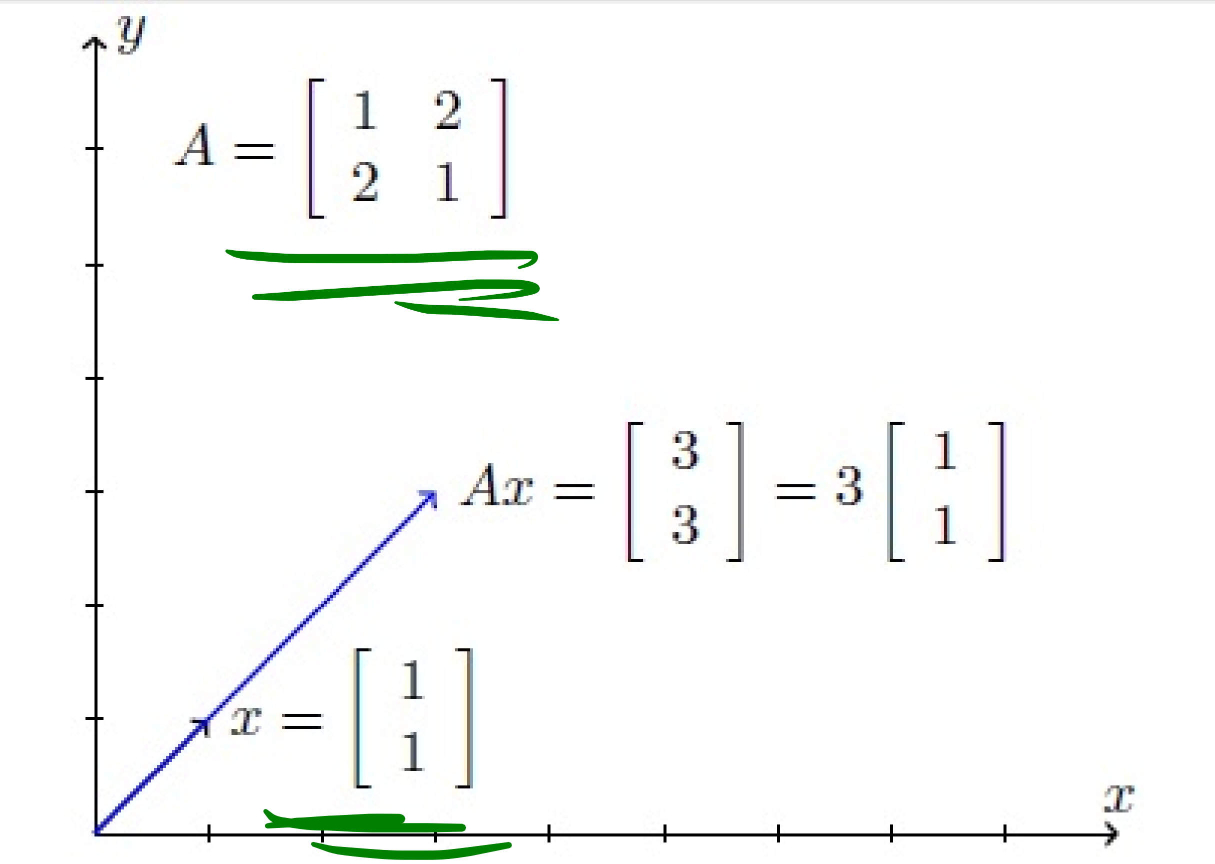




→ What happens when vector is multiplied with Matrix ?

- Scaling
- Change of direction (rotation)

Ax



What happen when we multiply A with x ?

-> Only scaling

-> No rotation

$$A x = \lambda x$$

Matrix — Eigen vector — Eigen value

A is a square matrix
$u_1 \dots u_n$ are eigen vector of A
$\lambda_1 \dots \lambda_n$ are eigen values of A

$Ax = \lambda x$

$\boxed{Au_1 = \lambda_1 u_1}$

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

Matrix

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

Matrix Matrix

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

$= U\Lambda$

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

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

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

$\because$ U^{-1} will exist
→ linearly independent eigen vector of A
→ A has all distint eigen values

$U = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$ $\quad |U| \neq 0$

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

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

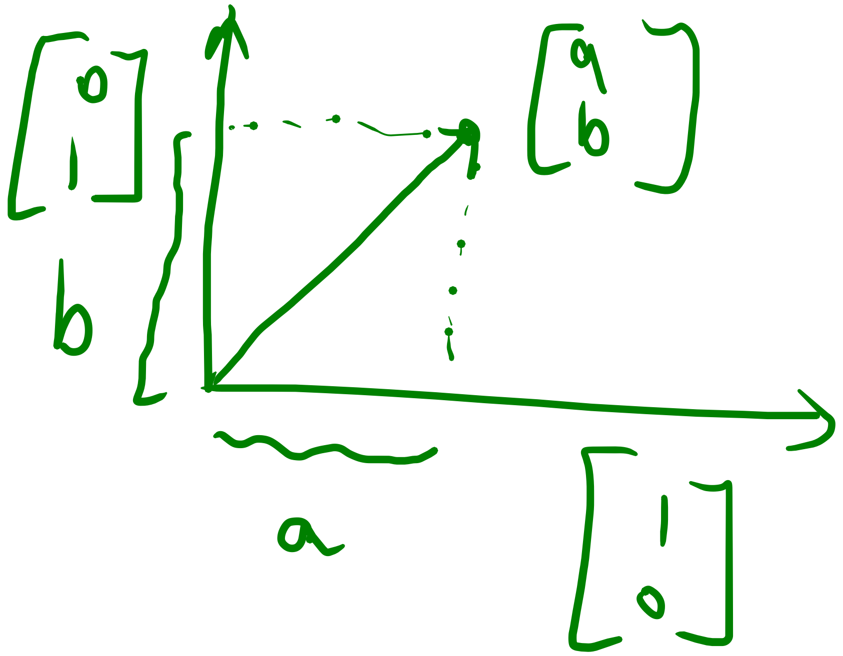
Eigen Value decomposition

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

if A is symetric

→ Break until 10:18 PM

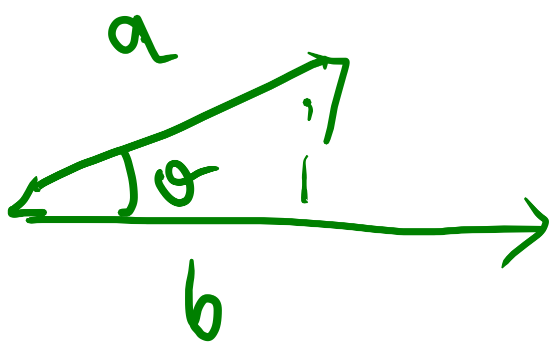
Representation of a point



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

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

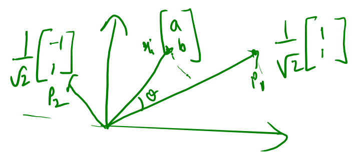
Projection of vector



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

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

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



$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$

Project x_i on singlevector p_1

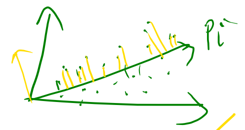
$$x_i \cdot p_1 = \frac{a \cdot 1}{\sqrt{2}} + \frac{b \cdot 1}{\sqrt{2}} = \frac{a+b}{\sqrt{2}}$$

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

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

$$x_i \cdot p_i$$

$$\underset{m \times d}{X} = \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \quad d \text{ (columns)},\ m \text{ (rows)}$$



Zero mean

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

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

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

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

$$= \frac{1}{m}$$

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

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

$$A = X^T X$$

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

$$\max \text{Var}(x_i) = \max \frac{p_i^T A p_i}{m}$$

$$= \max p_i^T A p_i$$

such $\|p_i\| = 1$

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

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

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

$$\boxed{A p_i = \lambda_i p_i} \rightarrow \text{Characteristics Equation}$$

$(X^T X)$

p_i is eigen vector of A

$(X^T X)$

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

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

Inference

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

where p_i is eigen vetor of $X^T X$

λ_i eigen value of $X^T X$

Orthogonal

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

$$\|p_i\|^2 = 1$$

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

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

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

$$- \lambda (p + (p^T)^T)$$

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

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

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

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

$$= 2Ap - 2p$$

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

$X = m \times d$

$X^T \quad d \times m$

$X^T X = d \times d$

1) Yes

2) No

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

$P_1 \sim 70\%$

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

$$\hat{X} = X P \qquad \hat{X} \to m \times d,\quad X \to m \times d,\quad P \to d \times d$$

$$\hat{X} = XP$$

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

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

$$= \frac{1}{m}\underbrace{P^T A P}_{}$$

$$= D = \Lambda$$

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

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

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

$$= P\Lambda P^T$$

P → eige vector

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

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

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

$$= U\Lambda U^T$$

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

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