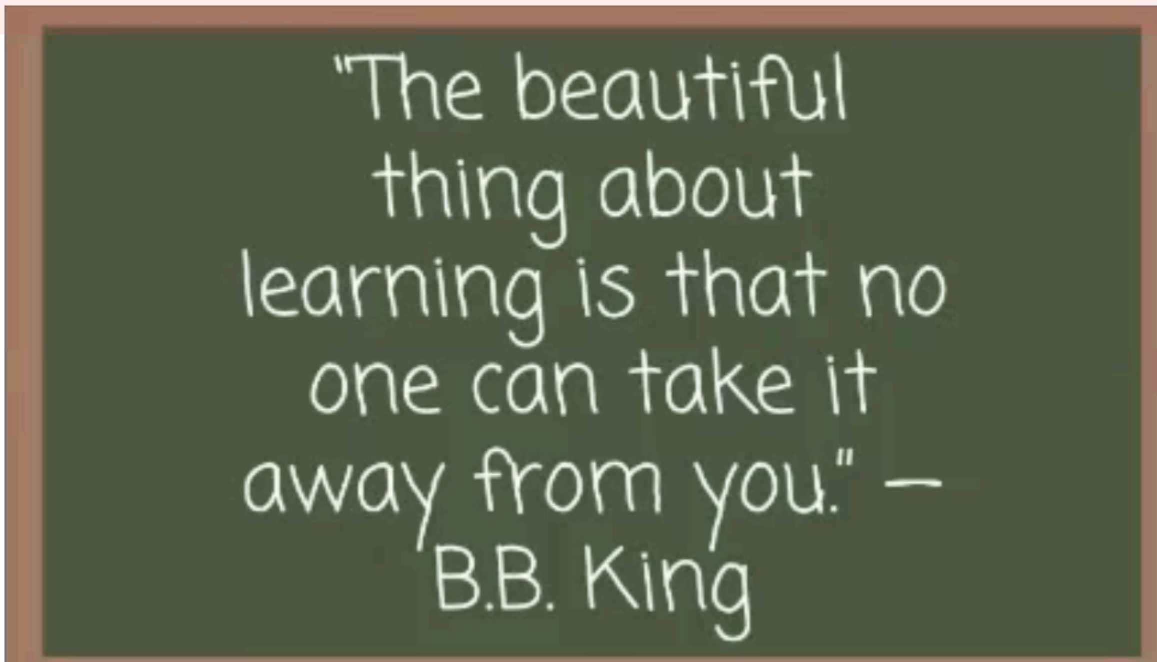
PCA cont. t-SNE



Agenda :-

→ PCA Recap

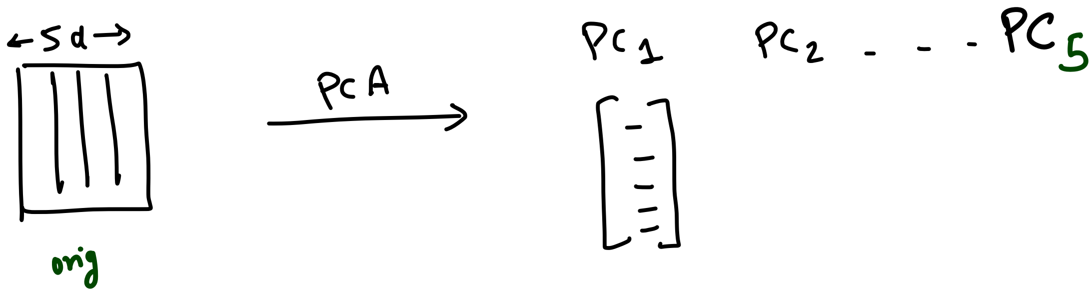
→ Maths

→ Limitations

→ t-SNE Intuition

→ t-SNE Code

How many PCs?

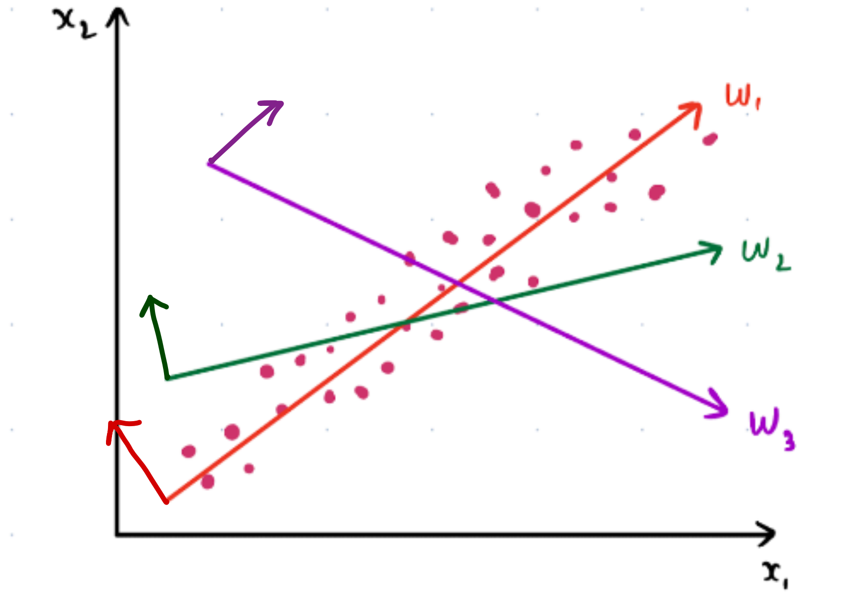


Q → angle b/w PC_3, PC_5 ?

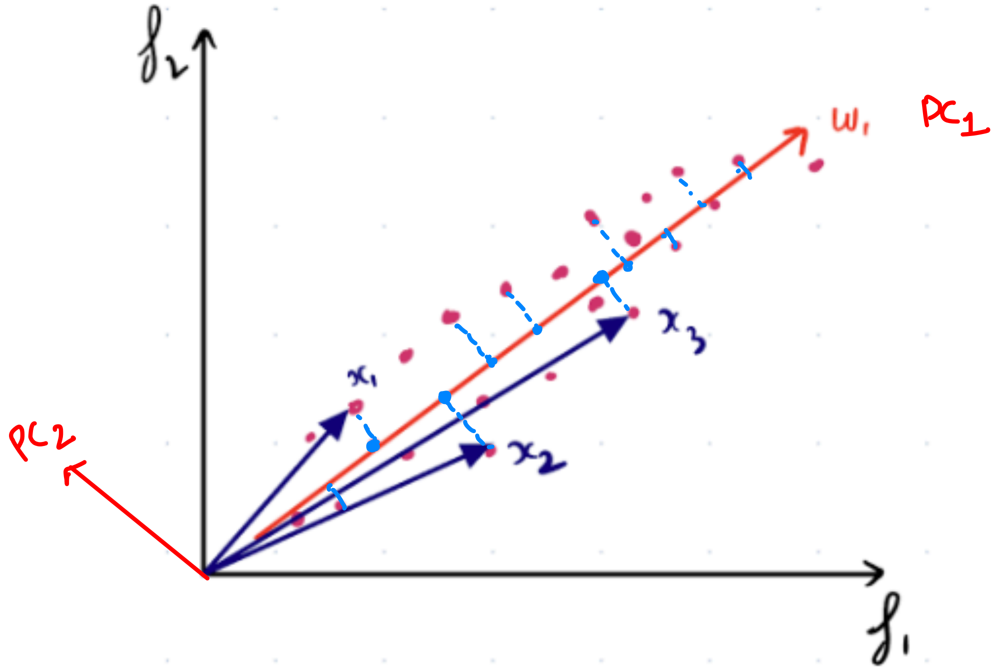
A → 90° orthogonal.

Q → M.C exists after PCA?

A → yes.

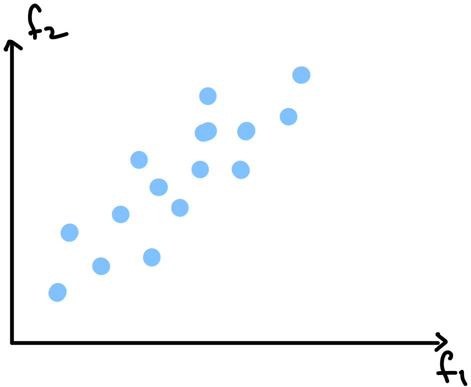


x2
x1
w1
w2
w3



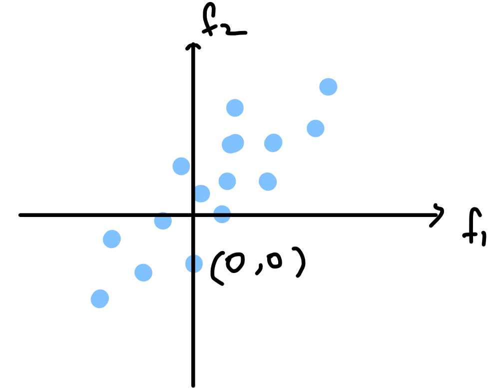
f2
f1
w1
PC1
PC2
x1
x2
x3

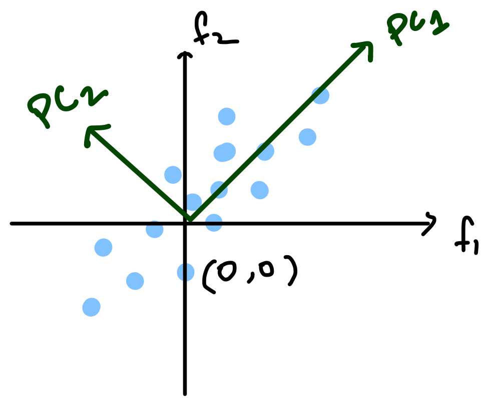
Maths [Heavy Maths Ahead]



feature scaling

$\left[\begin{array}{l}\text{mean centering}\\ \text{unit var.}\end{array}\right.$





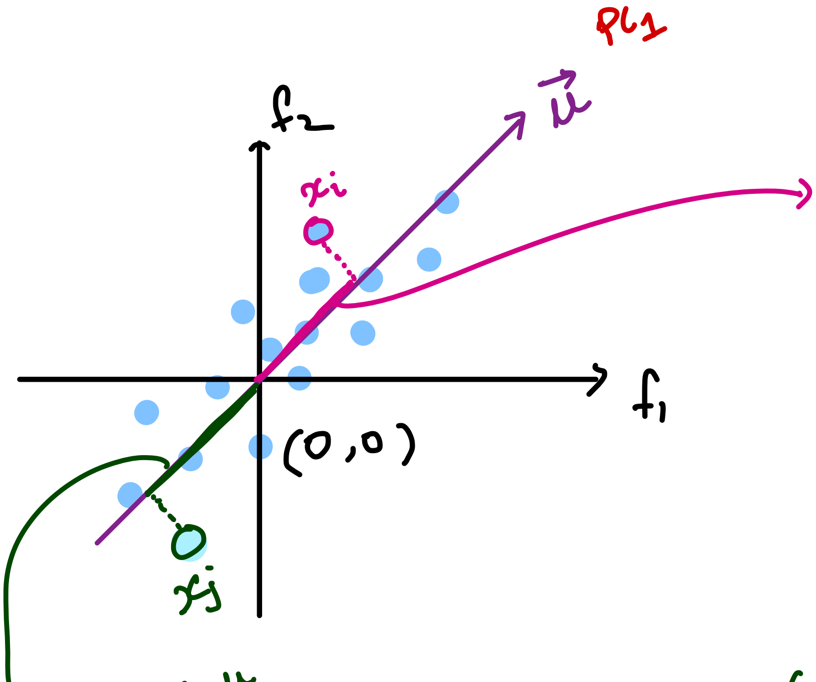
$\vec{u}$: $w^T x + w_0 = 0$

$[w_0 = 0]$

$w^T x = 0$

$$\boxed{u^T x = 0}$$

$$u = \begin{bmatrix} 2.5 \\ 3.1 \end{bmatrix}$$



$$Proj_{\vec{u}}^{x_i} = \frac{x_i \cdot u}{||u||}$$

Calculate Proj of All data Points onto $\vec{u}$

$$\max\left[\frac{1}{n}\sum_{i=1}^{n} \frac{x_i \cdot u}{||u||}\right]$$ → avg length of Proj. of all data Points

$$Proj_{\vec{u}}^{x_j} = \frac{x_j \cdot u}{||u||}$$

$$\max_{\vec{u}} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i \cdot u}{\|u\|} \right)^2 \right]$$

→ -ve & +ve Proj length

→ diff

Constraint: $\|u\| = 1$

Constraint optimization

↓

Unconstraint optimization

↓

$$\max_{u,\lambda} \frac{1}{n}\sum_{i=1}^{n}(x_i \cdot u)^2 + \lambda(\|u\|^2 - 1)$$

$A^2 = A^T A = \|A\|^2$

$$\Downarrow$$

$$\max_{u,\lambda} \frac{(X \cdot \vec{u})^2}{n} + \lambda(\|u\|^2 - 1)$$

$$\Downarrow$$

$$\frac{(X \cdot u)^T \cdot (X \cdot u)}{n} + \lambda(u^T u - 1)$$

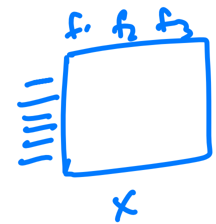
$(AB)^T = B^T A^T$

$$\Downarrow$$

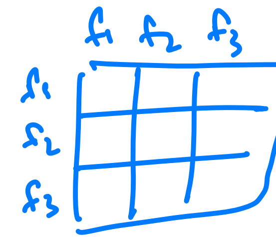
$$\max_{\lambda, u} \quad \frac{u^T X^T . X . u}{n} + \lambda \left(u^T u - 1 \right)$$

Covar matrix (X)

$= V$



$Cov(X) =$



$$L = \max_{u, \lambda} \left[u^T V . u + \lambda u^T u - \lambda \right]$$

$$L = \max_{u,\lambda} \left[u^T V . u + \lambda u^T u - \lambda \right]$$

$$a = a^T a = a^2$$

$$\frac{d a^T a}{d a} = 2a$$

$$\frac{d a^2}{d a} = 2a$$

$$\frac{\partial L}{\partial u} = 2Vu + 2u\lambda = 0$$

Scalar derivative $f(x) \rightarrow \frac{df}{dx}$			Vector derivative $f(\mathbf{x}) \rightarrow \frac{df}{d\mathbf{x}}$		
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{B}$	$\rightarrow$	$\mathbf{B}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{b}$	$\rightarrow$	$\mathbf{b}$
x^2	$\rightarrow$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\rightarrow$	$2\mathbf{x}$
bx^2	$\rightarrow$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\rightarrow$	$2\mathbf{B}\mathbf{x}$

$$\frac{\partial L}{\partial \lambda} = 0 + u^T u - 1 = 0$$

Set $= 0$

$[u^T u = 1]$

$$2Vu + 2u\lambda = 0$$

$$\cancel{2}Vu = -\cancel{2}u\lambda$$

$$Vu = \lambda' u$$

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$[1, 2, 3] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} =$$

$$\boxed{\lambda' = -\lambda}$$

$$\boxed{V.\vec{u} = \lambda \vec{u}}$$

Matrix (dxd)

vector (dx1)

Scalar

vector (dx1)

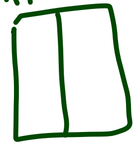
just find $\vec{u}$ using '<u>eig</u>'

Conclusion / summary

$X \xrightarrow{PCA} X\text{-new}$

$(n,5)$ $\quad$ $(n,2)$

f_1' f_2'



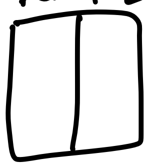
1. Scaling
2. Covar matrix $V = \frac{X^T X}{n}$ → $(d \times d)$
3. find $\vec{u}$ & λ for 'V'

evec , eval = np.linalg.eig(V)

PC_1 & λ_1

PC_2 & λ_2

PC1 PC2



eigenvectors

all PC's (eigen vectors)

all λ's (eigen values)

4. Projection of all points (x) on 2 PC's

eigenvectors → $(d,2)$

$\boxed{X \cdot eVector}$

(n,d) $\quad$ $(d,2)$ $\Rightarrow (n,2)$

Calculate Importance of PC's

d features

$u_1 \quad u_2 \quad u_3 \quad \cdots \quad u_d$

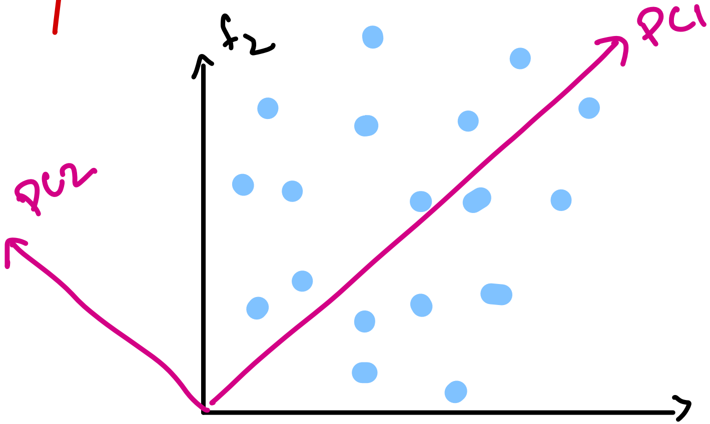
$\lambda_1 > \lambda_2 > \lambda_3 > \quad > \lambda_d$

$3.8 \quad 2.9 \quad 1.4 \quad 0.07$

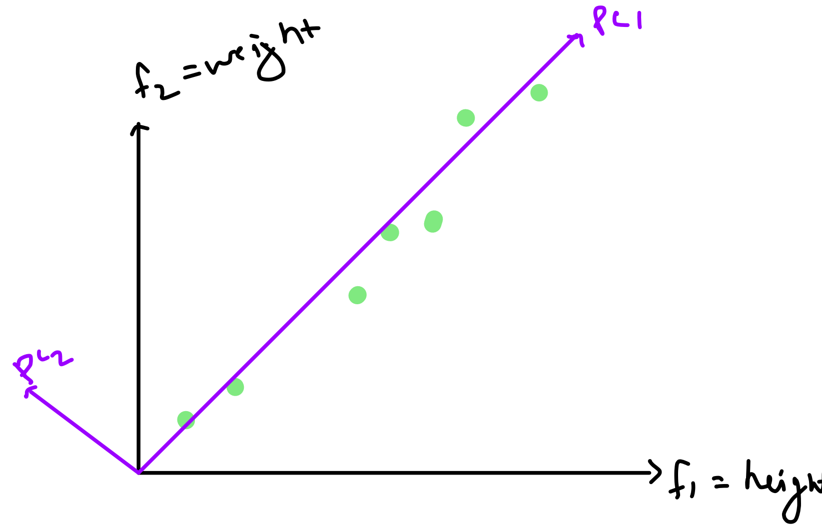
$$\text{Imp}(u_1) = \frac{3.8}{\text{sum all } \lambda\text{'s}} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots \lambda_d}$$

$$\underset{\text{Preserved}}{\text{Info}}(u_1, u_2) = \frac{\lambda_1 + \lambda_2}{\sum \lambda_i\text{'s}}$$

Assumptions



PCA X



Limitations

1. Low Interpretability of PC's

f_1' $PC_1 = 0.8$ mileage $+ 1.4$ odo $+ \ldots$

$PC_2 = 4.9$ age $+ \ldots$

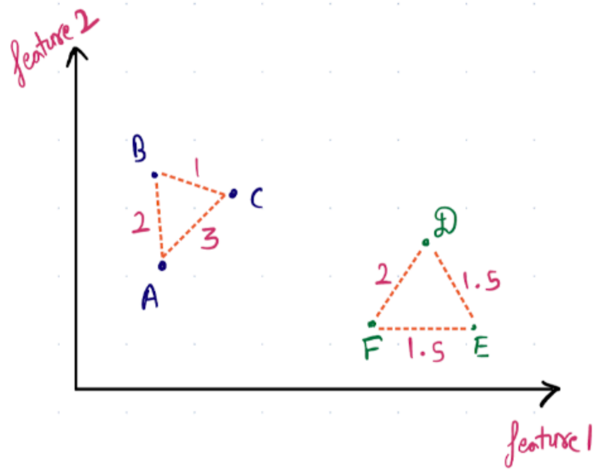
2. Trade off b/w info lost & dimⁿ reduction

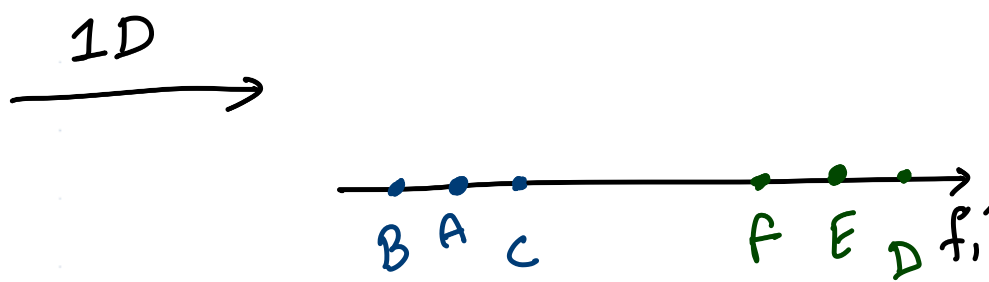
3. PCA is Not robust to Outliers !

t-SNE → 2008 Geoff Hinton

[only viz]
2D, 3D

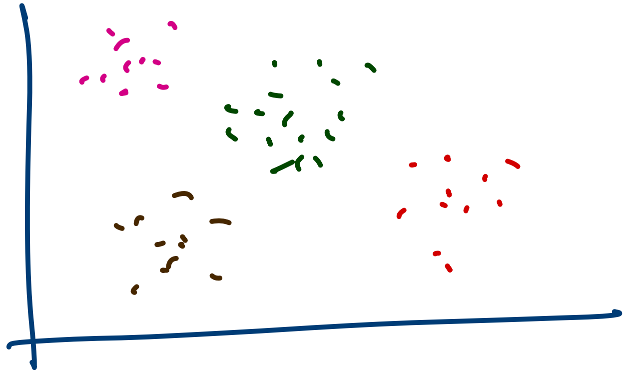
[t-distributed Stochastic Neighbourhood Embedding]



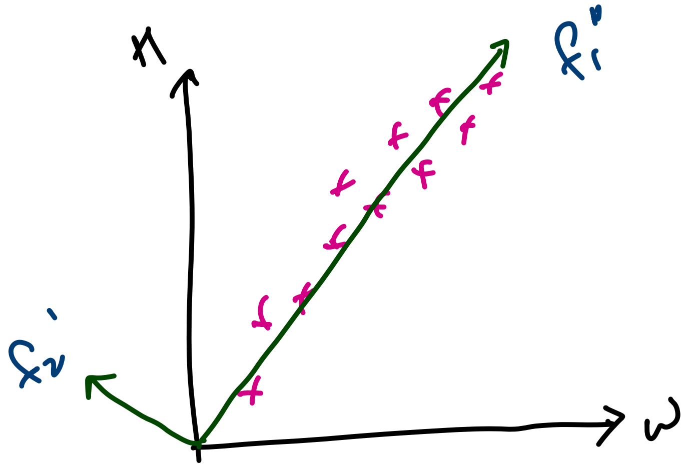


→ local neighbourhood

→ relative distances



$$\left[\underset{\text{High Dim}^n}{\text{Probality}_N(i,j)} \approx \underset{\text{Low Dim}^n}{\text{Probability}_N(i,j)} \right]$$



$$f_1' = 1.8H + 0.3W$$