QDA 2023.06.13

PCA : dimensionality reduction

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \xrightarrow[\text{(rotation of the reference system)}]{\text{PCA}} \underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \\ \vdots \\ Z_p \end{bmatrix} \quad \} \; k < p$$

MOST OF THE VARIABILITY OF THE ORIGINAL p-variate PROCESS IS CAPTURED in the first $k < p$ PCs

"LATENT STRUCTURE" ←

COMMENT : THIS IS NOT A VARIATE SELECTION PROCEDURE

Hp : I have been centering in the $\underline{\mu}$

$\rightarrow \underline{X} - \underline{\mu} = \underline{X}'$ → this will be our new $\underline{X}$ to start PC's

original

(LOADINGS)

$\underline{\Sigma} = Var(\underline{X}) \rightarrow$

	λ_1	$\underline{\alpha}_1$	$\rightarrow z_1 = \underline{\alpha}_1' \underline{X}$
largest to the smallest ↓	λ_2	$\underline{\alpha}_2$	$\rightarrow z_2 = \underline{\alpha}_2' \underline{X}$
	$\vdots$	$\vdots$	$\vdots$
	λ_p	$\underline{\alpha}_p$	
	EIGENVALUES	EIGENVECTORS	

$Var(z_1) = \lambda_1$

$Var(z_2) = \lambda_2$

$\vdots$

$\lambda_1 / \sum_i \lambda_i \cdot 100 \rightarrow$ % VAR. EXPLAINED by z_1

$\vdots$

$Var(z_i) = \lambda_i$

$\lambda_i / \sum_i \lambda_i \cdot 100 \rightarrow$ % VARIAB. EXPLAINED by the i-th PC's

Keeping all the z's
$\ni$ we explain 80%
SUCH THAT
of the overall variability

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} > 0.8$$

$$\frac{\lambda_1}{\sum \lambda_i} + \frac{\lambda_2}{\sum \lambda_i} + \dots \frac{\lambda_k}{\sum \lambda_i} > 0.8$$

$$\underline{X} \longrightarrow \underline{z}' = [z_1 \; z_2 \, .. \, z_k] \longrightarrow \hat{\underline{X}} = f(\underline{z})$$

p-variate $\quad$ k-variate

$$X_1 \longrightarrow Y_1 = \frac{X_1 - \mu_1}{\sigma_1}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \longrightarrow \underline{R} = Var(\underline{Y})$$

$$X_p \longrightarrow Y_p = \frac{X_p - \mu_p}{\sigma_p}$$

STANDARDIZATION

→ When X's have very different order of magnitude is better to go with PCA on the $\underline{R}$ = correlation matrix

t	$x_1 \dots x_p$	z_1 — z_k
1	$\cdot \quad \cdot \quad \cdots$	
$\vdots$		
t	x_{1_t} - - x_{p_t}	z_{k_t}
m	$\cdot \quad \cdot \quad \cdots$	

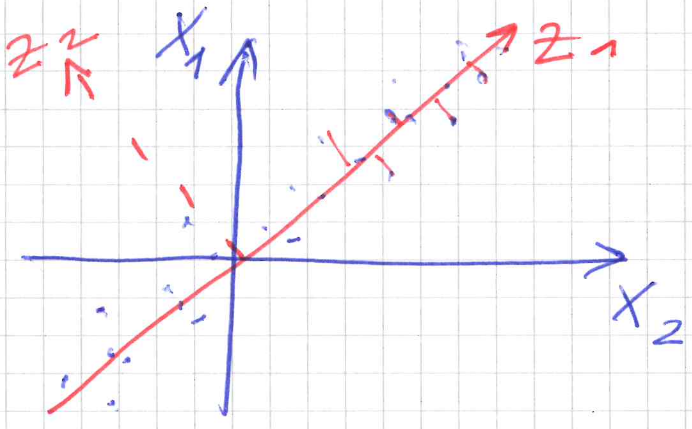
SCORE

$$z_{k_t} = \underline{\alpha}_k' \cdot \underline{x}_t$$

k^{th}

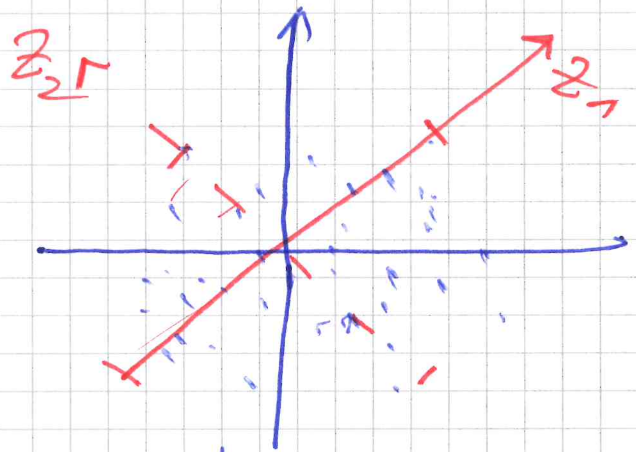
LOADINGS (eigenvector)

t^{th} observation



1^{st} process

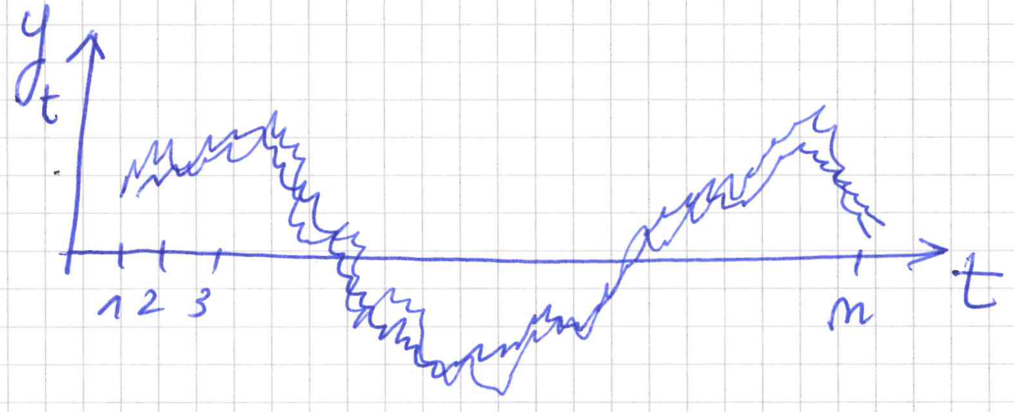
PROCESS A



2^{nd} process

PROCESS B

~~ex~~: SIGNAL or PROFILE DATA or FUNCTIONAL DATA



1^{st} SIGNAL $\underline{y}_1' = [y_{11} \dots y_{1m}]$

signal point in the signal

j^{th} SIGNAL $\underline{y}_j' = [y_{j1} \dots y_{jm}]$

= $j = 1 \dots 100$ SIGNALS each SIGNAL has $748 = m$ POINTS =

SIGNAL = p-variate vector (6)

$$\underline{y}_j = [\, y_{j1} \;\; y_{j2} \;\; y_{j3} \;\; \dots \;\; y_{jp} \,]$$

p-variate random variable

$p = 768$

$j = 1 \dots J \qquad J = 100 \rightarrow$ I have 100 SIGNALS
100 OBSERVATIONS of a p-variate process with $p = 768$

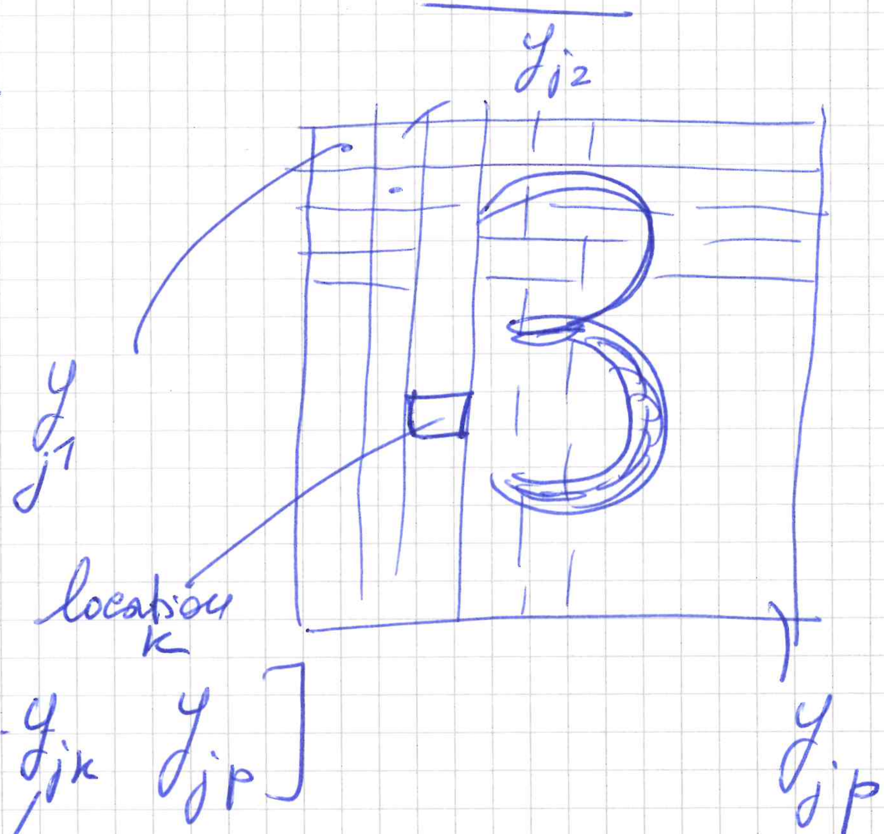
$\underline{y} = \Sigma_y \rightarrow$ PC's $\rightarrow$ Keep a subset of PC's say $k=3$

$\downarrow$ y (ORIGINAL) DIMENSIONAL REDUCTION via PCA:

$$768 = p \longrightarrow \underbrace{z_1 \;\; z_2 \;\; z_3}_{k=3}$$

$p = 768$

Ex IMAGES



$$\underline{y}_j' = [y_{j1} \ldots y_{jk} \; y_{jp}]$$

pixel intensity observed for the j^{th} image at location k