

# TÉCNICAS DE REDUCCIÓN DE DIMENSIONALIDAD

(+ SELECCIÓN ATRIBUTOS + PROYECCIÓN DATOS EN DIMENSIÓN REDUCIDA)

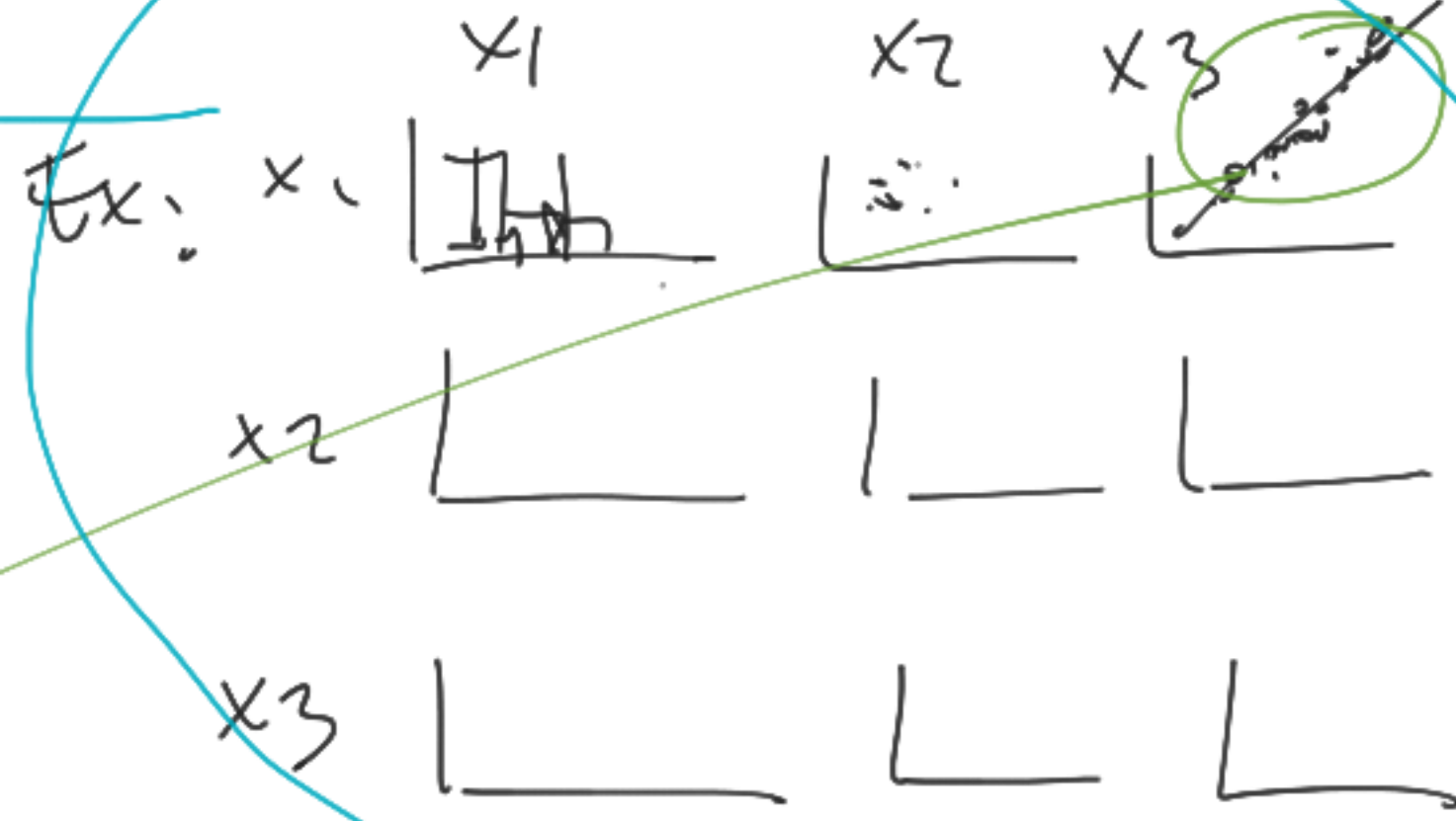
DATOS:  $A$   
 $N \times M$   
observaciones atributos/variables

SELECCIÓN ATRIBUTOS

$(x_1, x_2, x_3)$

↓ eliminar  $x_3$

$(x_1, x_2)$  tener  
VARIABLES  
IMPORTANCIA



varias variables altamente correlacionadas

REDUCCIÓN DIMENSIONALIDAD

3D

→

2D

$(x_1, x_2, x_3)$

$(z_1, z_2)$

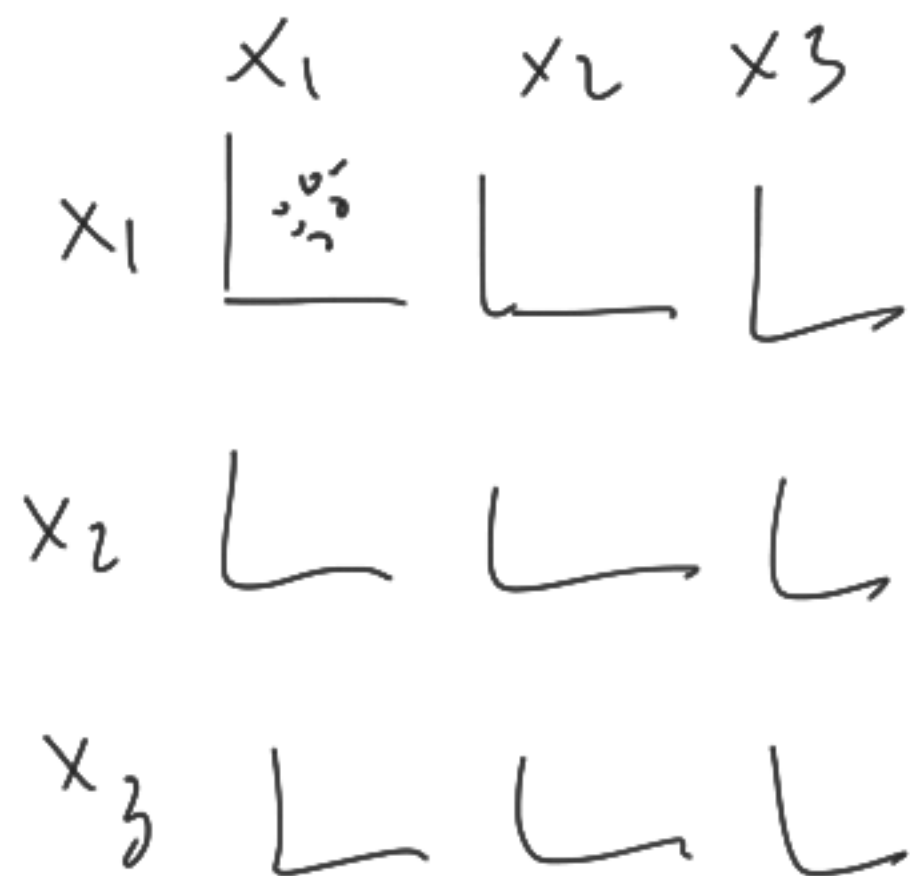
(PCA)

SIMPLIFICAR DATOS 3D

$$z_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3$$

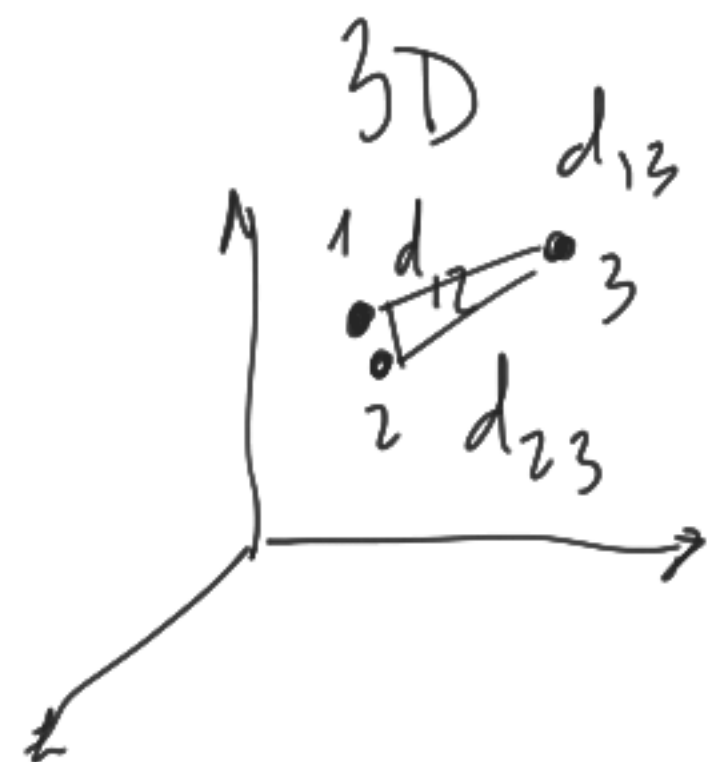
$$z_2 = \alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3$$

# Proyección DATOS DIMEUS. REDUCIDA

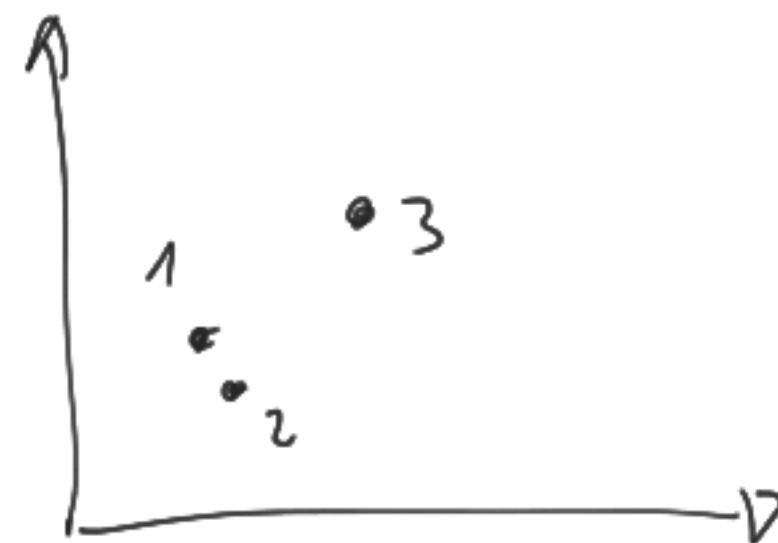


$f(x_1, x_2, x_3)$

→



MDS  
t-SNE → VISUALIZAR  
DATOS



for crop mantener las dist.  
relativas entre observaciones  
(optimización)

# Reducción de dimensionalidad: Análisis de Componentes Principales (PCA)

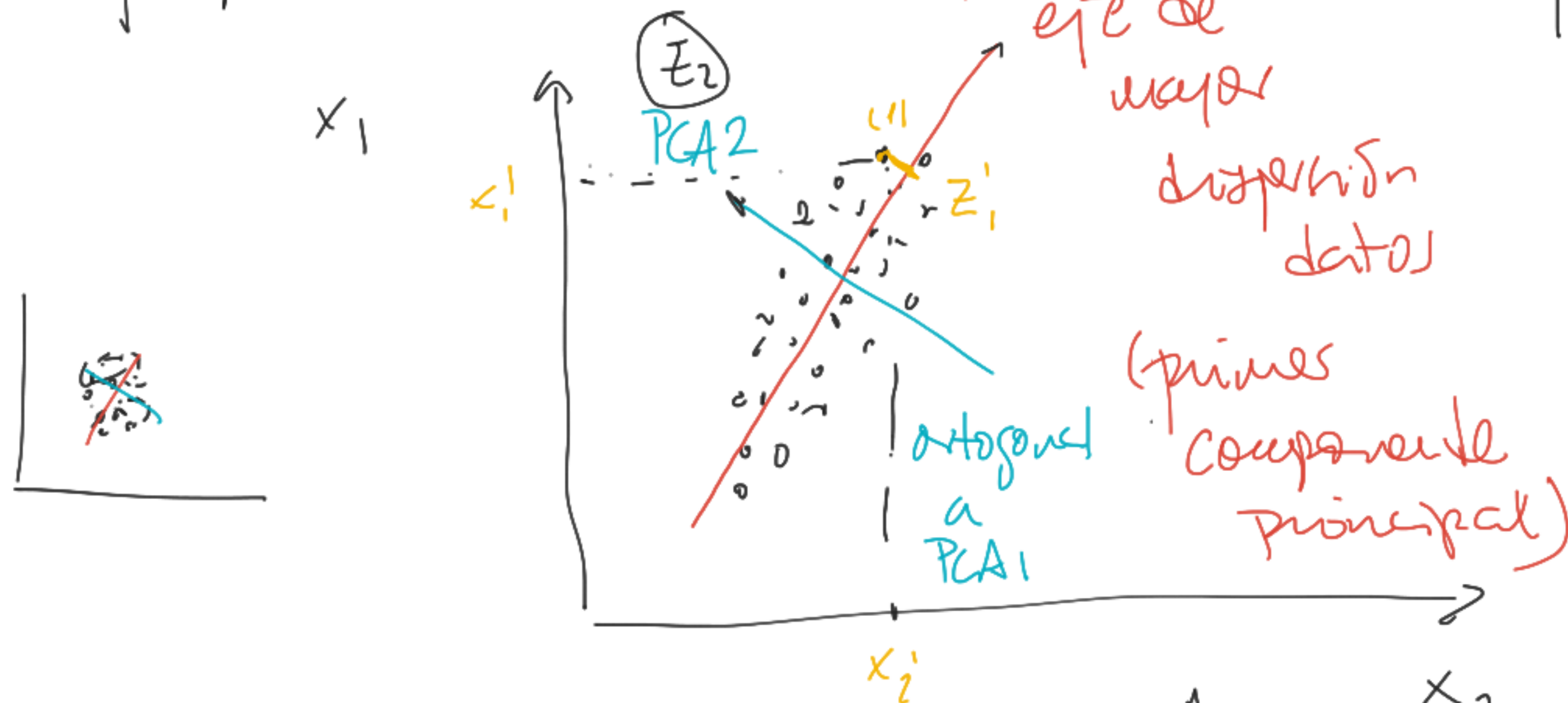
Principales (PCA)

↓  
transformación  
lineal  
coordenadas

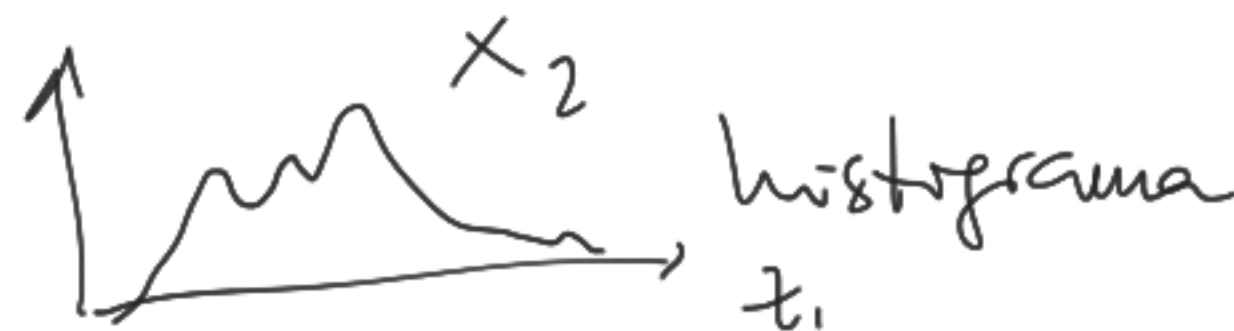
$$(x_1, x_2) \rightarrow (z_1, z_2)$$

- traslación
- rotación

Ejemplo: datos 2D



$$(x_1, x_2) \rightarrow z_1$$



PCA

datos  $M$ -dimensionales



$M$  componentes principales

PCA1    PCA2    ...    PCAM  
|        |        |  
 $z_1, z_2, \dots, z_M$



ordenar los PC por importancia (varianza explicada)



seleccionar los  $d$  PC más importantes



proyectar los datos al espacio PCA de dimensión reducida ( $d$ )

PCA

$(x_1, x_2, \dots, x_m)$

↓

$(z_1, z_2, \dots, z_d)$

} • more dimensional  
• more correlations entre variables

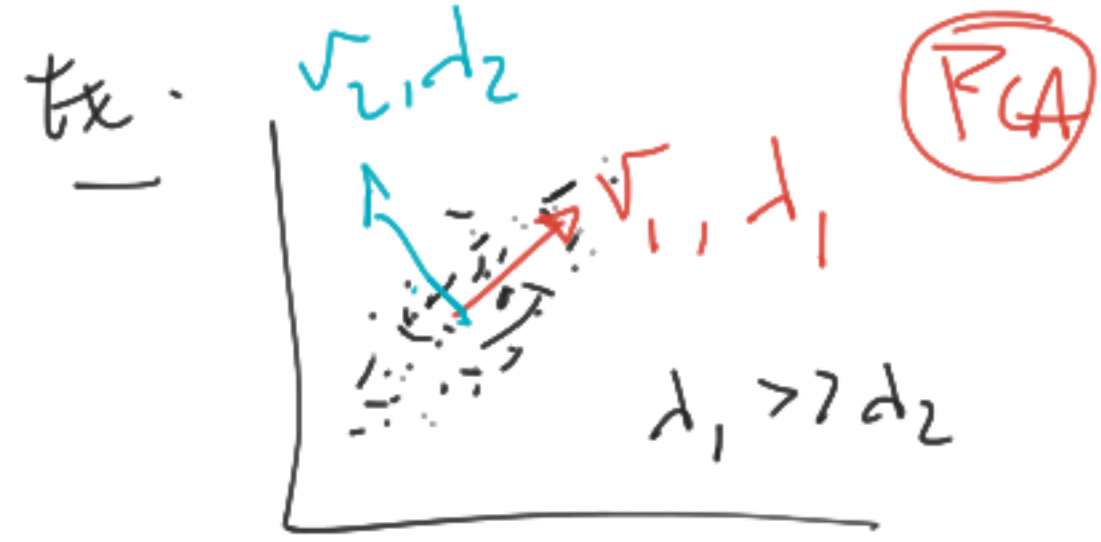


# MATEMÁTICAS PCA

$$A_{N \times M}$$

=

$$\begin{pmatrix} x_1 & \dots & x_M \\ \vdots & & \vdots \\ \text{obj1} & & \text{objM} \end{pmatrix}$$



- ① center los datos
  - ② escale los datos
- } estandarizar

$$\frac{A - \bar{A}}{\text{std}(A)}$$

- ③ contruyes matrix covarianza  $C = \frac{1}{N-1} \cdot (A - \bar{A})^T \cdot (A - \bar{A})$

- ④ diagonalizar C

↳ vectores propios  
↳ valores propios

ejes PC

$$\begin{matrix} \vec{v}_i, i=1 \dots M \\ \lambda_i, i=1 \dots M \end{matrix}$$

importancia de cada PC

$$\underbrace{M \times N}_{\text{input data}} \cdot \underbrace{N \times M}_{\text{output data}} = M \times M$$

# Regla reducción dimensionalidad:

PCA

$$M \rightarrow \underline{d}$$

$\lambda_1, \lambda_2, \dots, \lambda_M \rightarrow$  valores propios

$\frac{\lambda_1}{\sum_i \lambda_i}, \frac{\lambda_2}{\sum_i \lambda_i}, \dots, \frac{\lambda_M}{\sum_i \lambda_i} \rightarrow$  fracción explicada

$\frac{1}{\sum_i \lambda_i} (\lambda_1 + \lambda_2 + \dots + \lambda_d) \geq 0.95 \rightarrow$  fracción acumulada (cumsum)

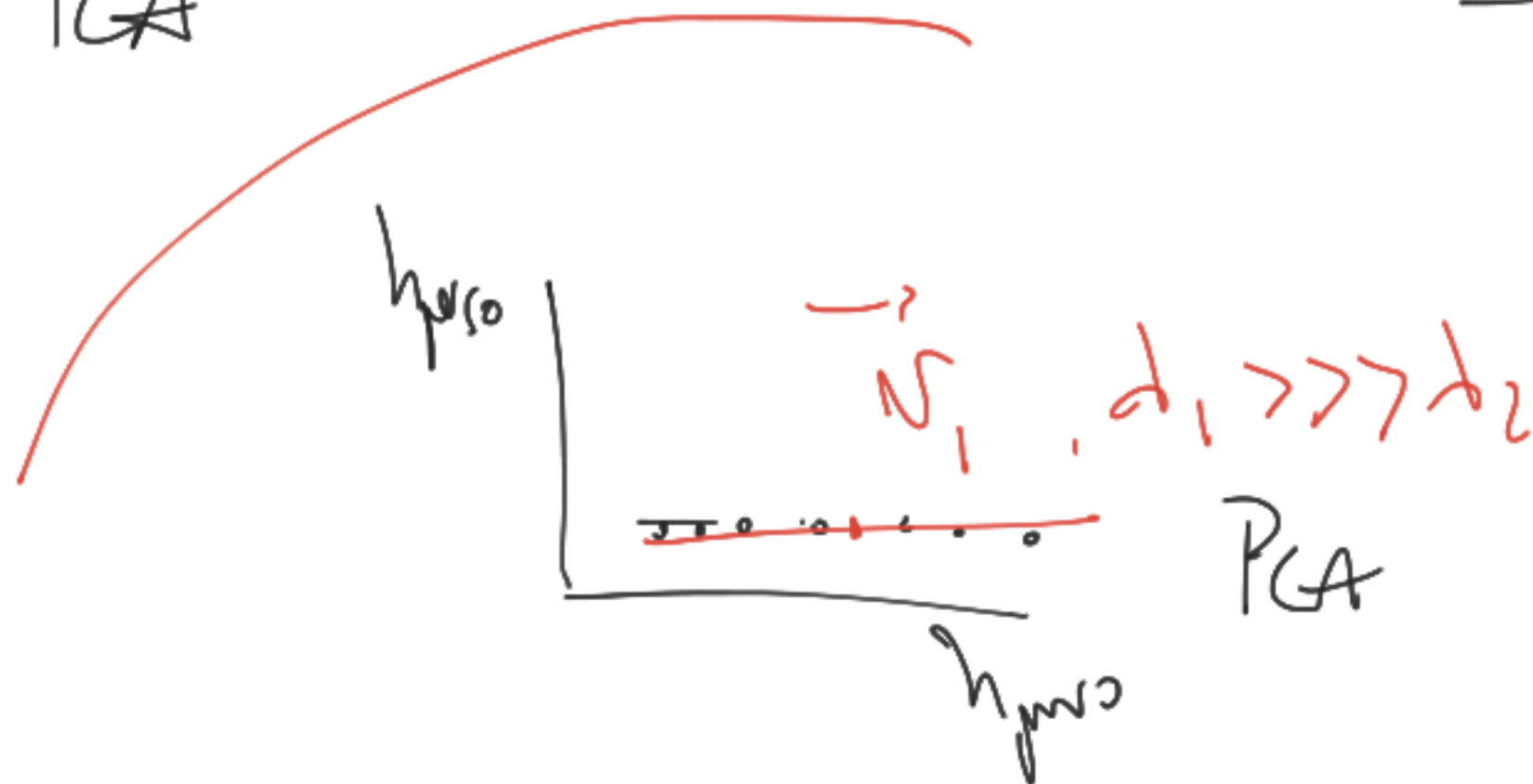
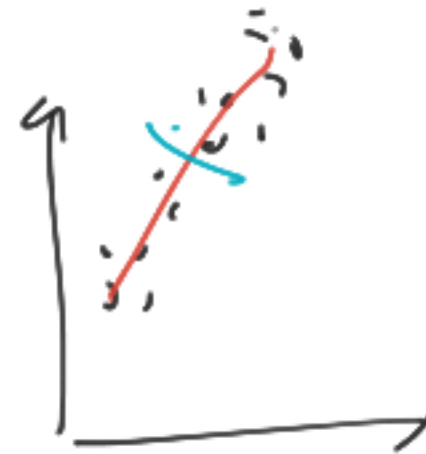
$\hookrightarrow$  Quedarse con tantos componentes como sea necesario para explicar el 95% de la dispersión

# ESCALADO DATOS PCA:

Ex:

$h_{persona}(m)$	$h_{marco/persono}(m)$
1.76	0.43
1.61	0.63
$\vdots$	$\vdots$

PCA



PCA

Ex:  $h_{persona}(m)$   $h_{marco/persono}(m)$

1.76

43

1.61

63

$\vdots$

↓

↓

$$\frac{1.56 - 1.89}{0.33}$$

$\lambda_2$

escale  
unidades



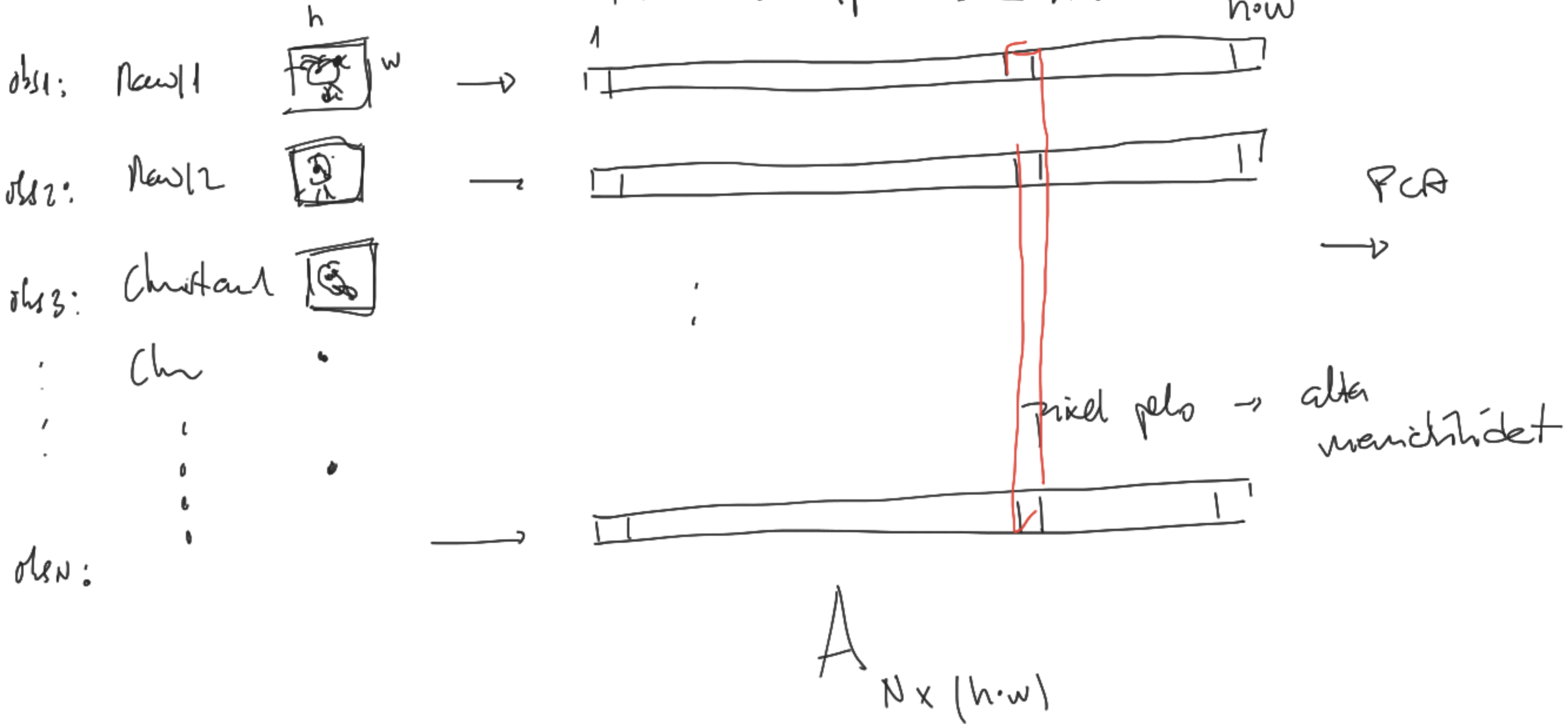
dispersion variable



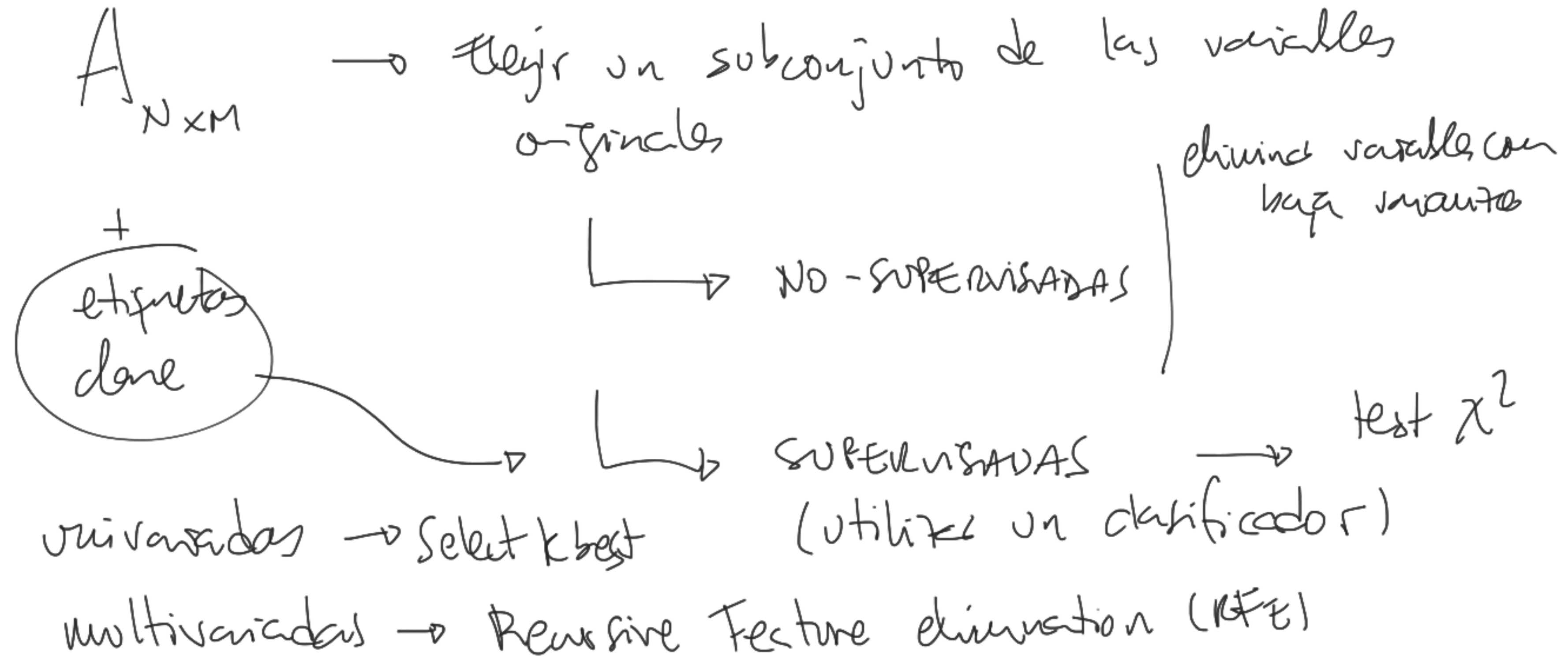
# PCA imágenes: Reconocimiento de caras (eigenfaces)

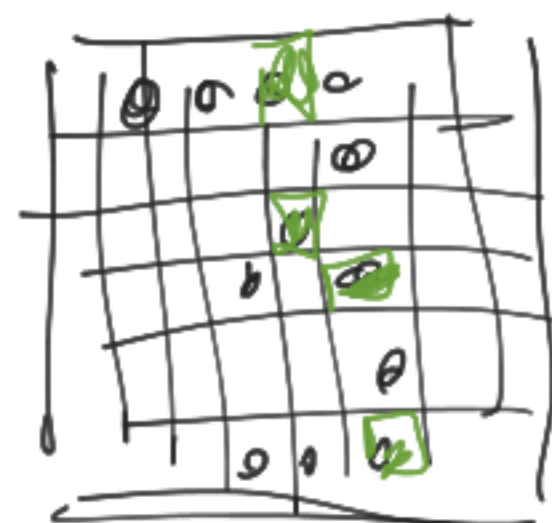
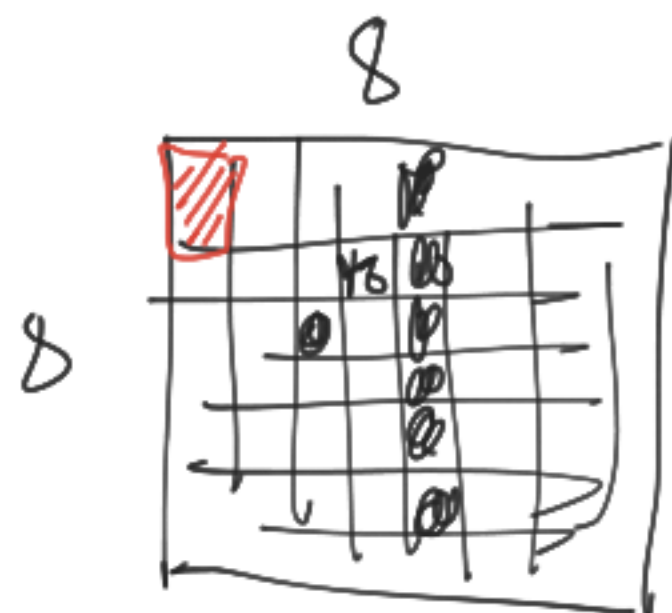
PCA

#variables = #píxeles =  $h \cdot w$



# SELECCIÓN DE ATRIBUTOS





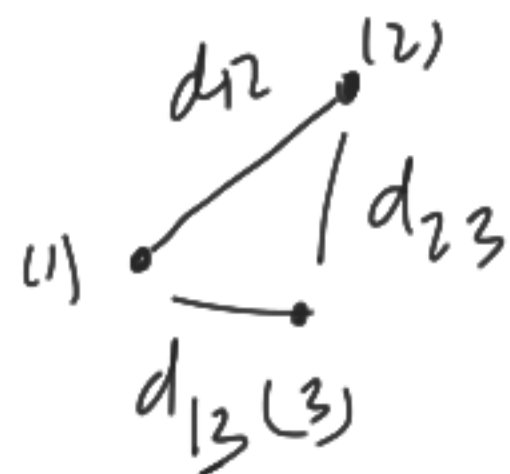
RFE 64 features/variables  $\rightarrow$  pixels



4 variables  $\rightarrow$  4 pixels

# PROYECCION DATOS MULTIDIMENSIONALES $\rightarrow$ VISUALIZACION

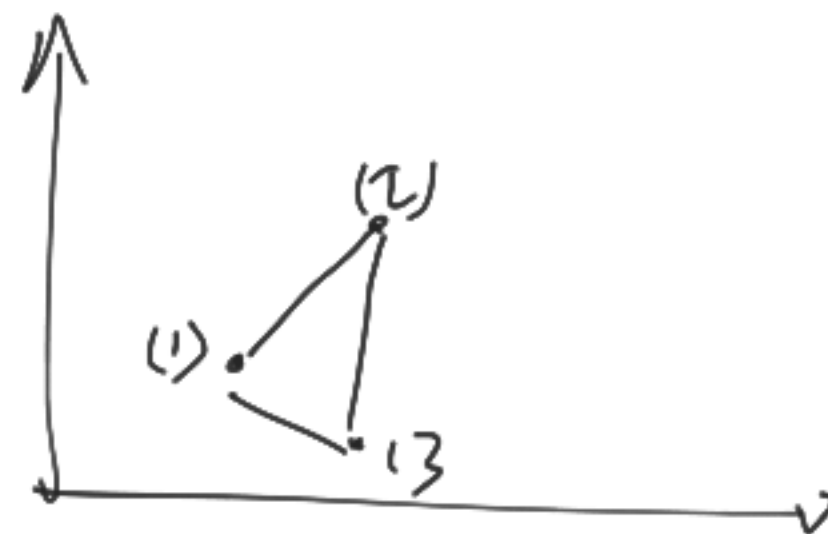
M-dimensiones



$f$

\_\_\_\_\_  $\rightarrow$

2 dimensiones



• MDS

• t-SNE

ERRO

o similar

ESTADO

BAJA

DIMENSIONALIDAD