

# 1

## ① NORMALIZACIÓN:

$$X \longrightarrow X_{\text{norm}}$$

↖ cada columna tiene  
 $\mu = 0$  ,  $\sigma = 1$

o bien

$\max = 1$  ,  $\min = 0$

Ejemplo: p3.m con y sin  
normalización

## ② CLEAN:

- Elimina columnas que tengan  $\sigma \approx 0$
- ✓ ✓ ✓ están correlacionados

Ejemplo: PAT03\_dataclean

## ③ STRATIFY:

NO SELECCIONA COLUMNAS

→ Selecciona filas

↗ Training

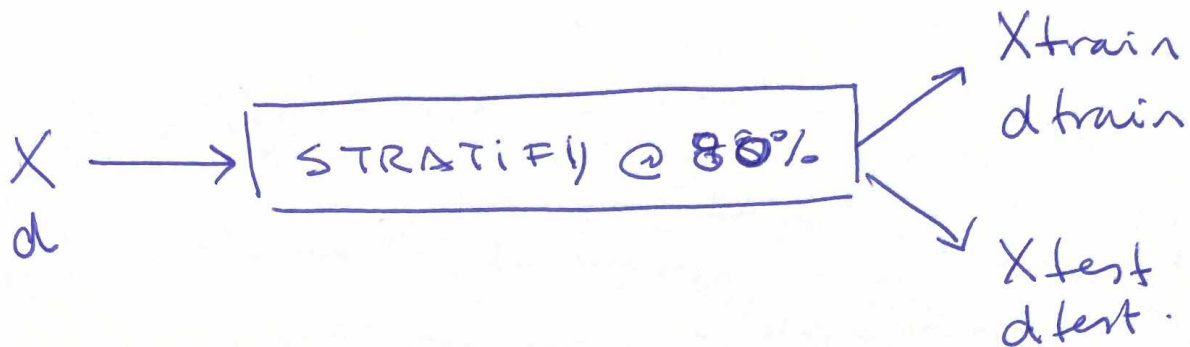
↘ Testing

Mantiene la distribución de datos.

Ejemplo:

$X \leftarrow$  datos originales (1000 filas, 50 columnas)

750 clase 1  
250 clase 0



$X_{train}$  tiene 800 filas:

75% = 600 clase 1  
25% = 200 clase 0

$X_{test}$  tiene 200 filas:

75% = 150 clase 1  
25% = 50 clase 0

OJO:  $d$ : clase ideal: vector de 1000 elementos  
donde hay 750 que son 1  
250 que son 0

