

Idea del Análisis de las Componentes Principales

Reducción de la Dimensión

Selección de Características	Extracción de Características
Eliminación hacia atrás	ACP
Selección directa	LDA
Eliminación bidimensional	ACP con Kernel
Comparación de Scores	

¿Y que es ACP?

De las m variables independientes del dataset, ACP extrae las $p \leq m$ nuevas variables independientes que explicar la mayor parte de la varianza del dataset, sin importar el valor de la variable dependiente.

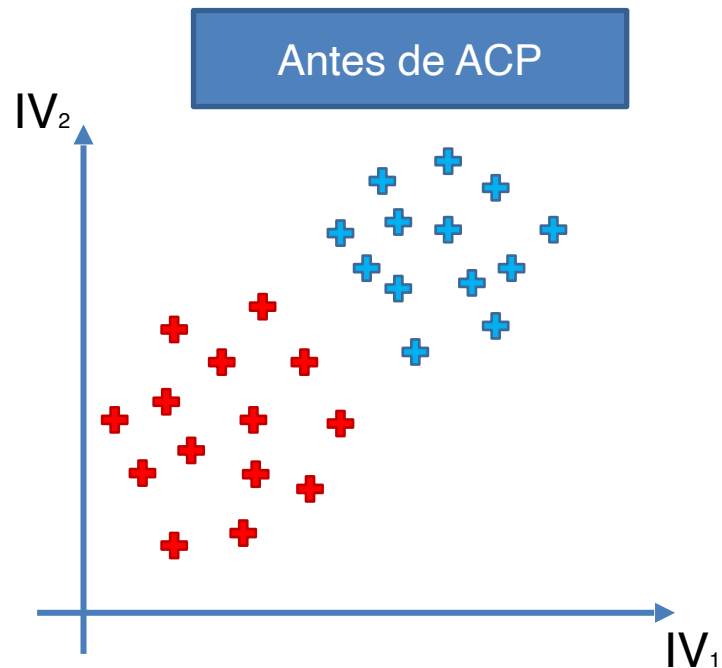
PCA in a few words

De las m variables independientes del dataset, ACP extrae las $p \leq m$ nuevas variables independientes que explicar la mayor parte de la varianza del dataset, sin importar el valor de la variable dependiente.

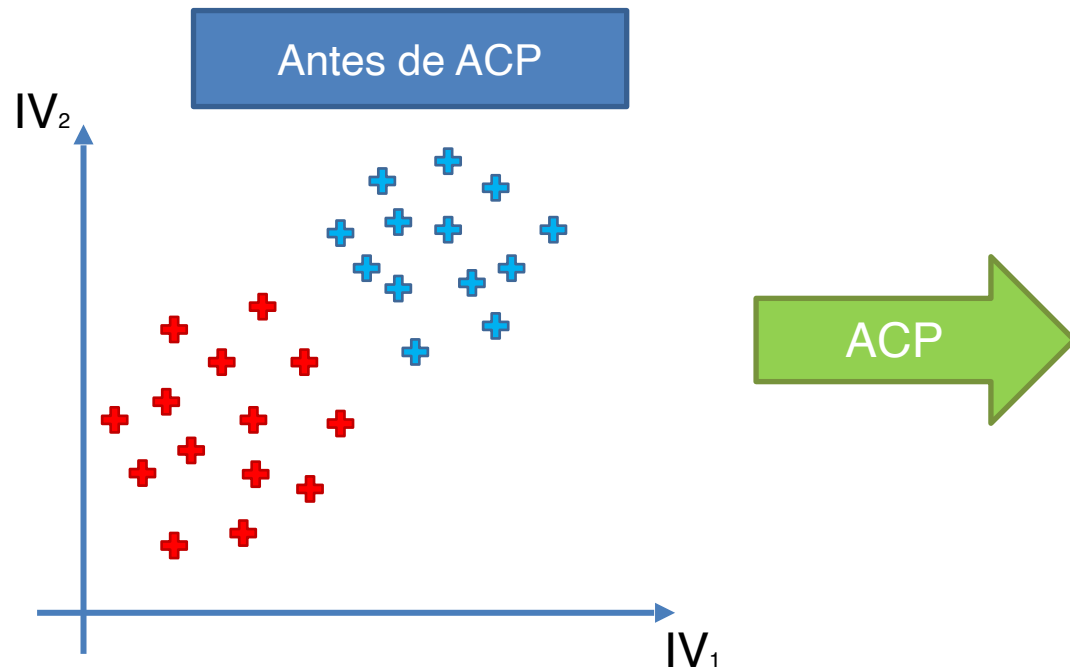


Como no se usa la VD, hace que consideremos el ACP un modelo no supervisado.

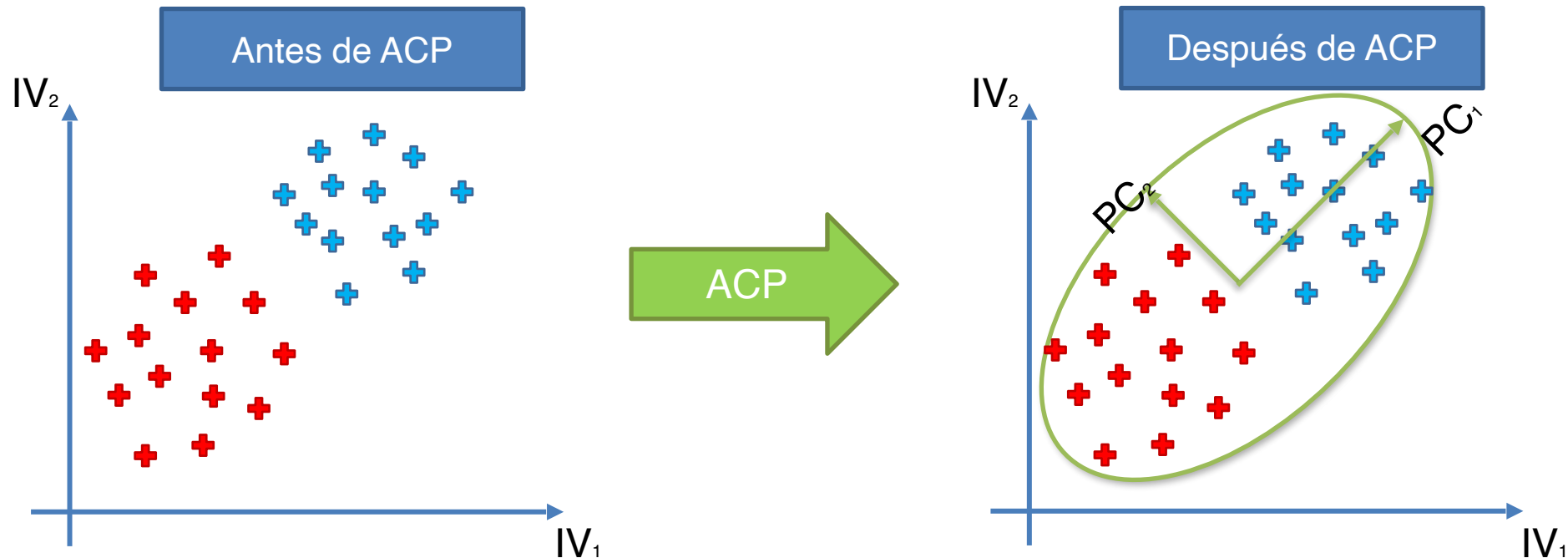
ACP encuentra dimensiones de máxima varianza



ACP encuentra dimensiones de máxima varianza



ACP encuentra dimensiones de máxima varianza



PC_1 y PC_2 son las direcciones de máxima varianza

Las matemáticas tras el ACP

PASO 1: Aplicar escalado de variables a la matriz de características X , formada por m variables independientes.



PASO 2: Calcular la matriz de covarianzas de las m variables independientes de X .



PASO 3: Calcular los valores y vectores propios de la matriz de covarianzas.



PASO 4: Elegir un porcentaje P de varianza explicada y elegir los $p \leq m$ valores propios más grandes tales que:



$$\frac{\sum_{j=1}^p \lambda_j}{\sum_{i=1}^m \lambda_i} \geq P$$

PASO 5: Los p vectores propios asociados a estos p valores más grandes son las componentes principales. El espacio m -dimensional del dataset original se proyecta al nuevo subespacio p -dimensional de características, aplicando la matriz de proyecciones (que tiene los p vectores propios por columnas).