HUB IQ

Inteligencia Artificial & Machine Learning

Applicaciones en movilidad

Dr. Iván S. Razo Zapata

EngineeringX

Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Reducción de Dimensionalidad

HUB IQ

EngineeringX
Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Problemas con la alta dimensionalidad de los datos

- En algunos dominios donde aplicamos algoritmos de aprendizaje de máquina nos enfrentamos a conjuntos de datos de alta dimensionalidad
  - Ejemplo: observaciones con decenas variables
- Costo de procesamiento y almacenamiento
- Atributos relevantes e irrelevantes
- Dificultad para evaluar distancias

HUB IQ

EngineeringX
Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Técnicas comunes para reducción de dimensionalidad

- Selección de atributos
  - Seleccionar un subconjunto de atributos de los atributos originales
- Generación de atributos
  - Mapear los atributos originales a un nuevo espacio de menor dimensión

Selección de atributos

- Seleccionar un **subconjunto** de atributos
  - Que **no afecte significativamente** a la tarea de regresión/clasificación
- Un atributo es **irrelevante** si no afecta a la tarea
- Un atributo es **redundante** si no añade nada nuevo
- Un atributo se considera **relevante** si no es irrelevante o redundante

Selección de atributos

- Regularización
  - Ridge
  - Lasso

Regularización – Idea general

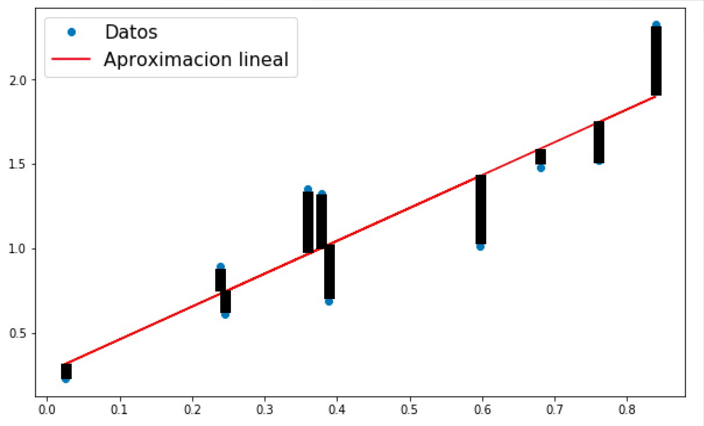
$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2$$

$$\theta_0 = \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x + \theta_2 x^2 - y)$$

$$\theta_1 = \theta_1 - \alpha\frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x + \theta_2 x^2 - y)x$$

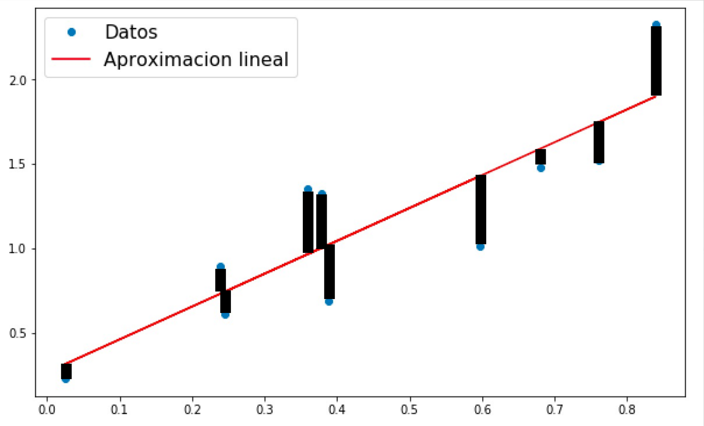
$$\theta_2 = \theta_2 - \alpha\frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x + \theta_2 x^2 - y)x^2$$



Regularización – Idea general

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2$$



$$\theta_j \leftarrow \theta_j - \alpha \times \sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x$$

Regularización – Ridge

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma \sum_{i=1}^{m} \theta^2\right]$$

Penalización L2

$$\theta_j \leftarrow \theta_j - \alpha \times \frac{E(\theta)}{\partial \theta_j}$$

Ridge – para regresión lineal

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma \sum_{i=1}^{m} \theta^2\right]$$

$$\theta_1 \leftarrow \theta_1 - \alpha \times \frac{E(\theta)}{\partial \theta_1}$$

$$\frac{E(\theta)}{\partial \theta_1} = \frac{\frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma \sum_{i=1}^{m} \theta^2\right]}{\partial \theta_1} = \frac{\partial}{\partial \theta_1}\frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \frac{\partial}{\partial \theta_1}\frac{1}{2m}\gamma \sum_{i=1}^{m} \theta^2$$

$$\frac{E(\theta)}{\partial \theta_1} = \frac{2}{2m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)\frac{\partial}{\partial \theta_1}(\hat{y}_\theta(x) - y) + \frac{2}{2m}\gamma\theta_1$$

HUB IQ

EngineeringX

Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Ridge – para regresión lineal

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma\sum_{i=1}^{m}\theta^2\right]$$

$$\theta_1 \leftarrow \theta_1 - \alpha \times \frac{E(\theta)}{\partial\theta_1}$$

$$\frac{E(\theta)}{\partial\theta_1} = \frac{2}{2m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)\frac{\partial}{\partial\theta_1}(\hat{y}_\theta(x) - y) + \frac{\gamma}{m}\theta_1$$

$$\frac{E(\theta)}{\partial\theta_1} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)\frac{\partial}{\partial\theta_1}(\theta_0 + \theta_1 x + \theta_2 x^2 - y) + \frac{\gamma}{m}\theta_1$$

$$\frac{E(\theta)}{\partial\theta_1} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x + \frac{\gamma}{m}\theta_1$$

Ridge – para regresión lineal

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma \sum_{i=1}^{m}\theta^2\right]$$

$$\theta_1 \leftarrow \theta_1 - \alpha \times \frac{E(\theta)}{\partial\theta_1}$$

$$\frac{E(\theta)}{\partial\theta_1} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x + \frac{\gamma}{m}\theta_1$$

$$\theta_1 \leftarrow \theta_1 - \alpha \times \left(\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x + \frac{\gamma}{m}\theta_1\right)$$

$$\theta_1 \leftarrow \theta_1 - \alpha \times \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x + \frac{\alpha\gamma}{m}\theta_1$$

$$\theta_1 \leftarrow \theta_1\left(1 - \frac{\alpha\gamma}{m}\right) - \alpha \times \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x$$

Ridge – para regresión lineal

$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma\sum_{i=1}^{m}\theta^2\right]$$

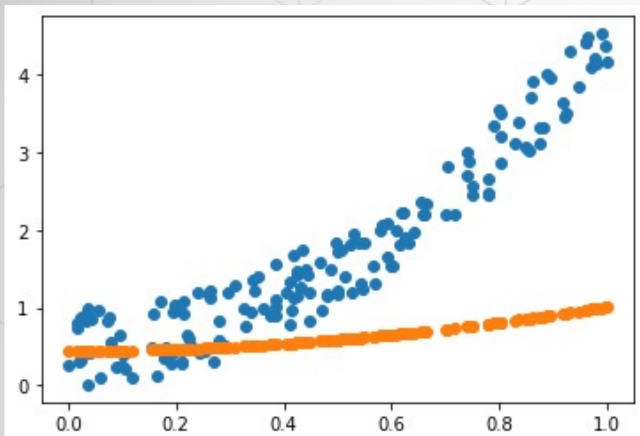
$$\theta_0 \leftarrow \theta_0 - \alpha\times\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x$$

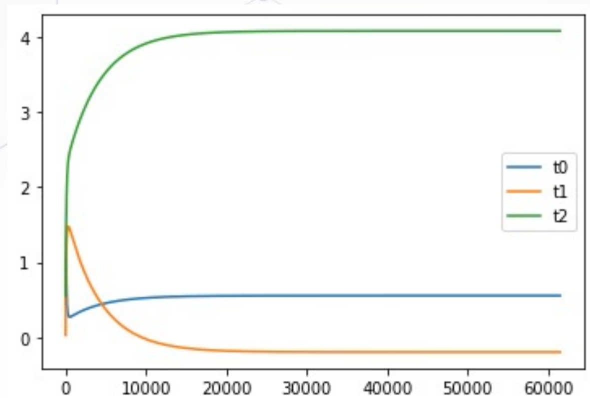
$$\theta_j \leftarrow \theta_j\left(1 - \frac{\alpha\gamma}{m}\right) - \alpha\times\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)x$$

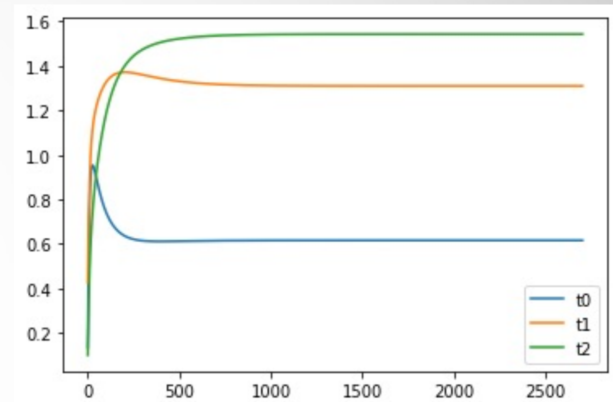
$$j \geq 1$$

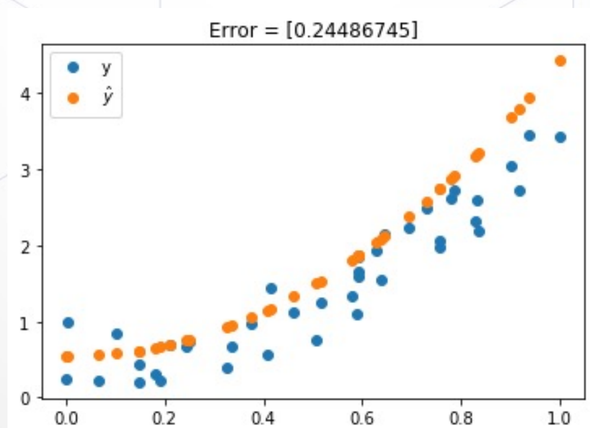
Ridge – para regresión lineal

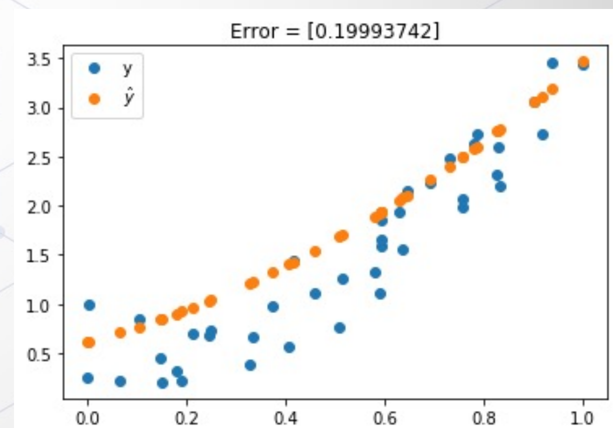
$$\gamma = 10$$





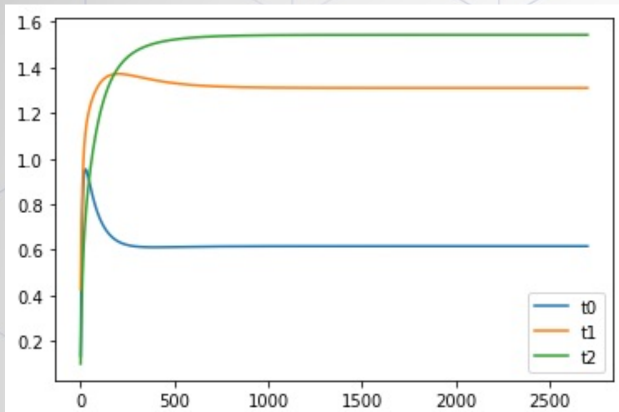




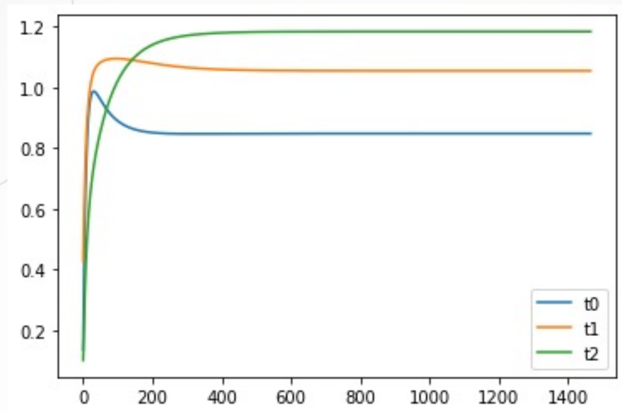


Ridge – para regresión lineal

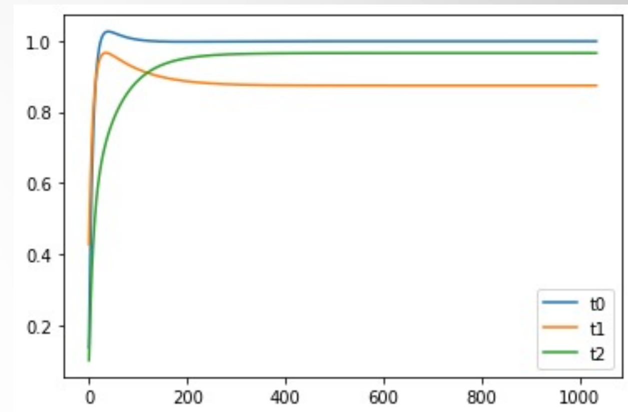
$\gamma = 10$

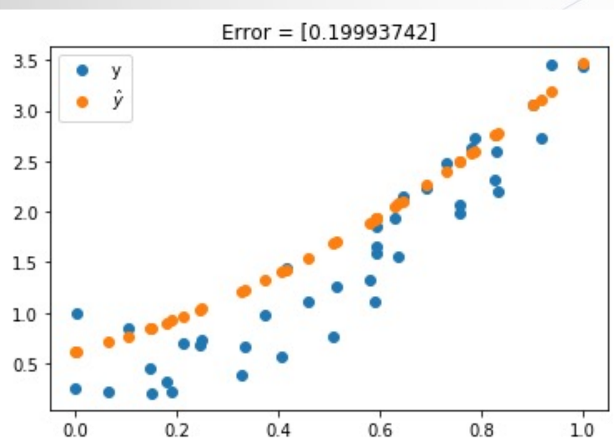


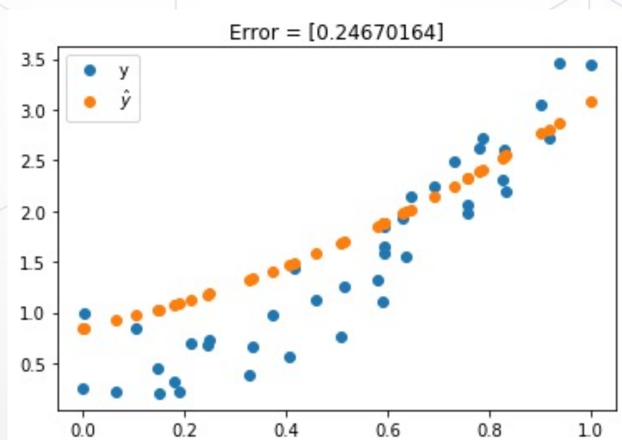
$\gamma = 20$

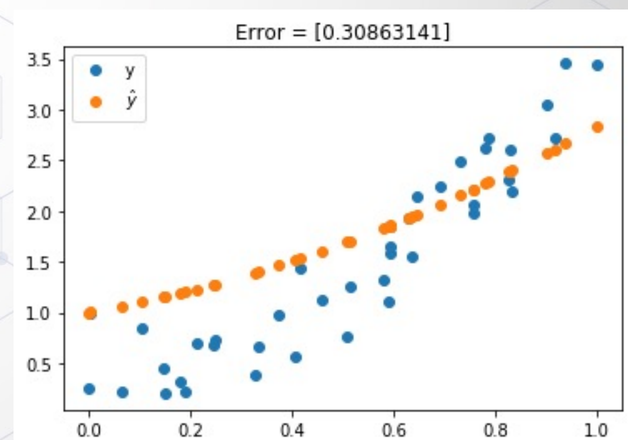


$\gamma = 30$

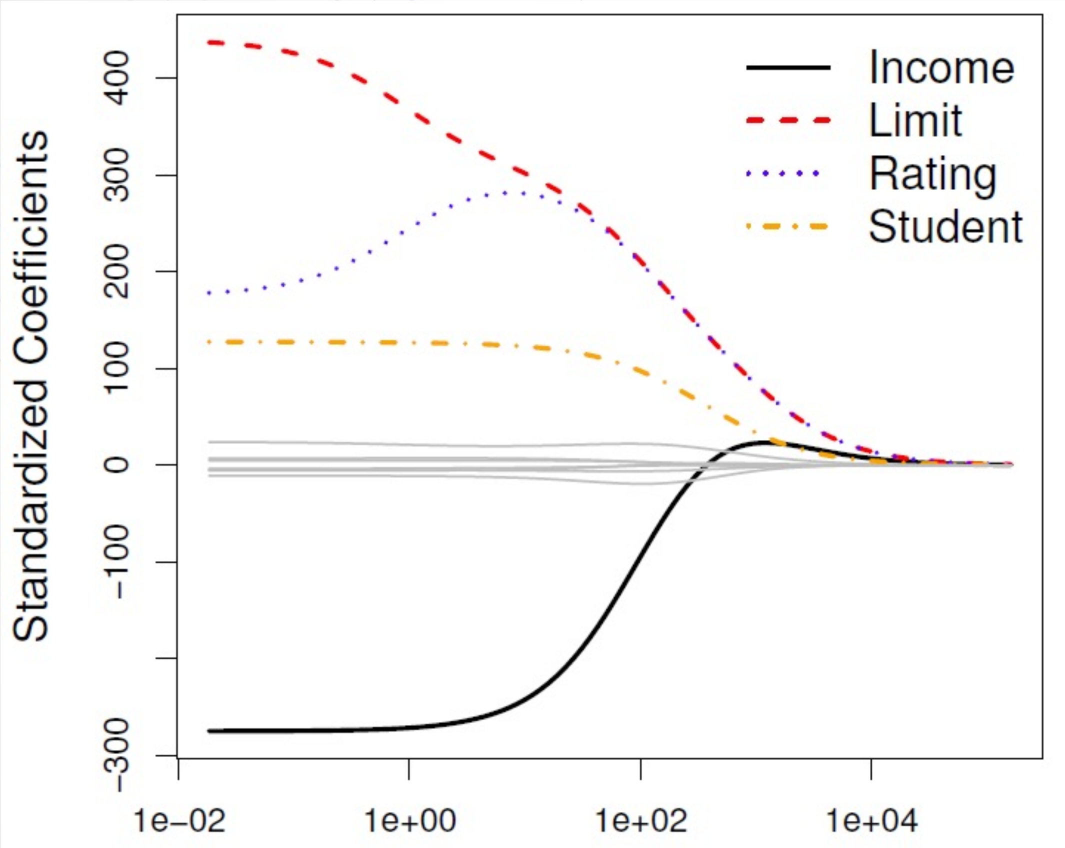




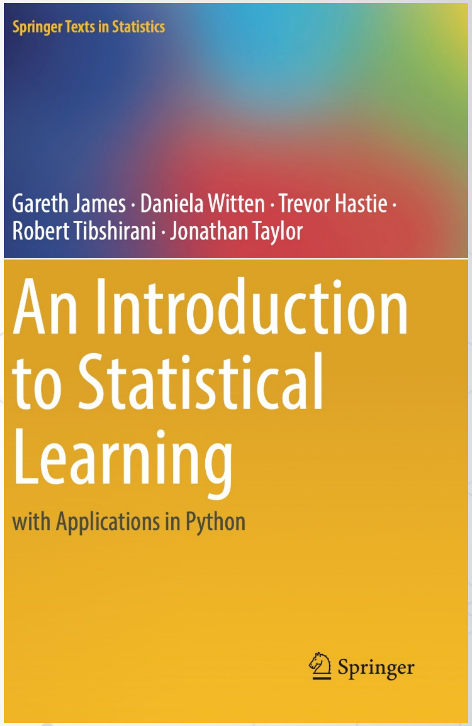




Ridge – para regresión lineal



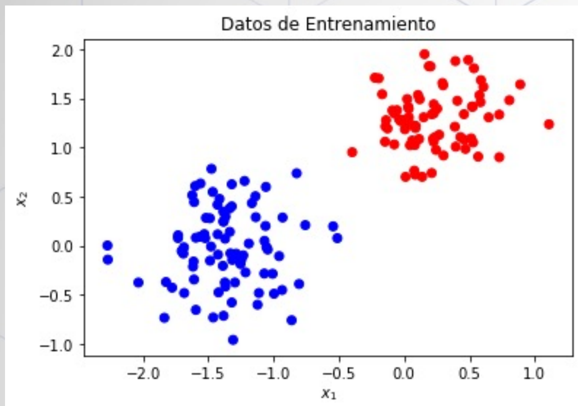
γ

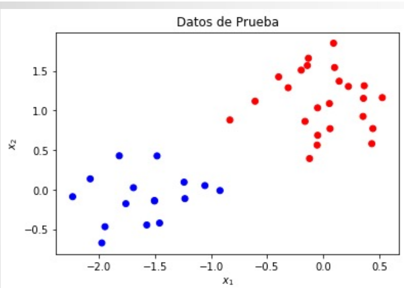


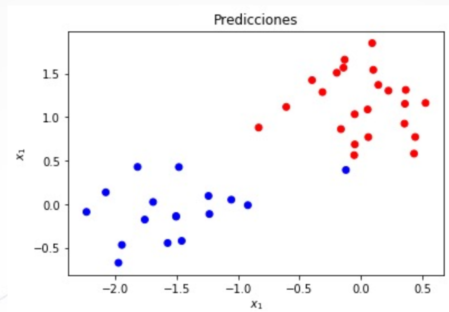
https://www.statlearning.com/

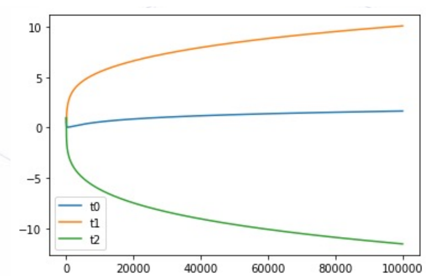
Ridge - para regresión logística

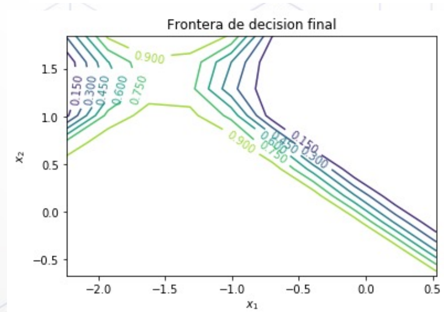
$\gamma = 10$

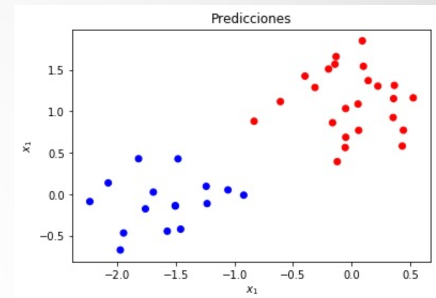


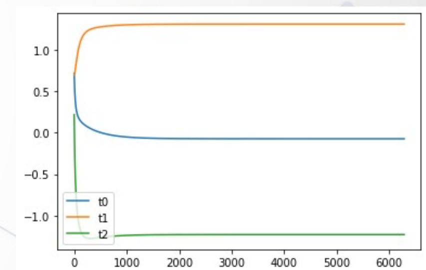


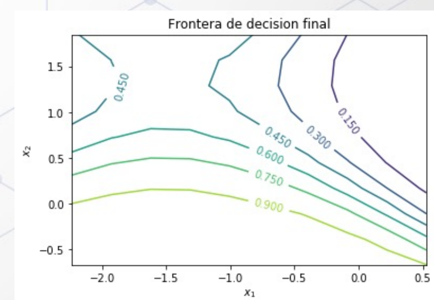






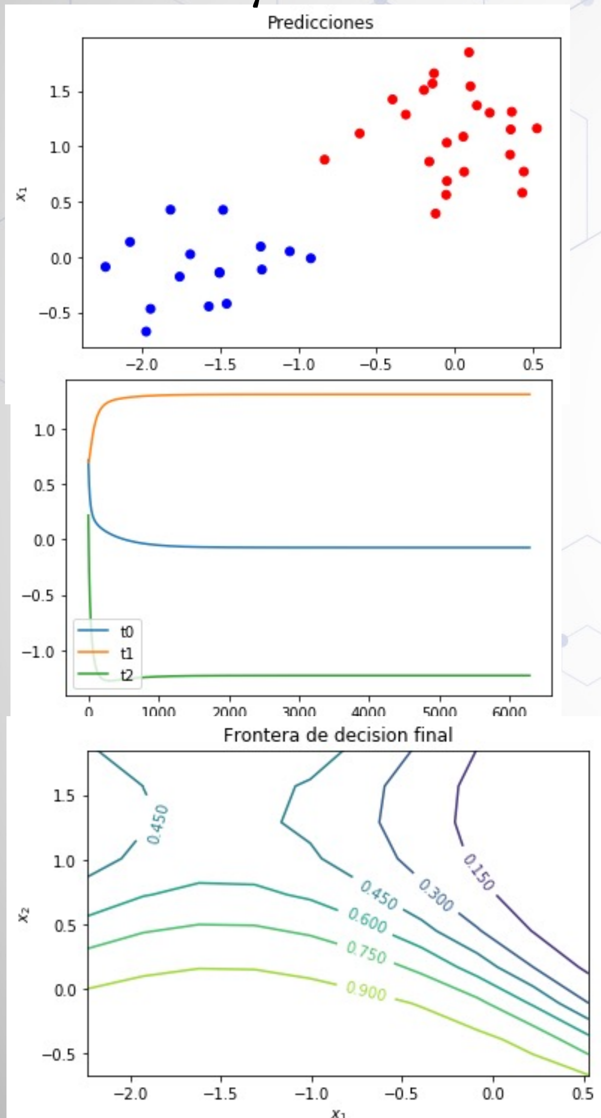




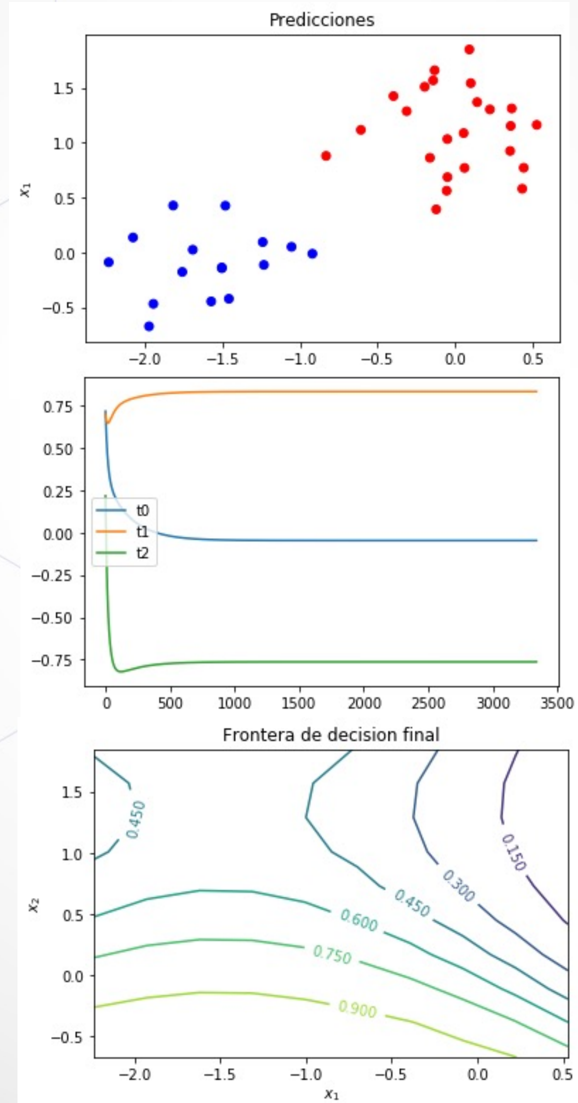


Ridge - para regresión logística

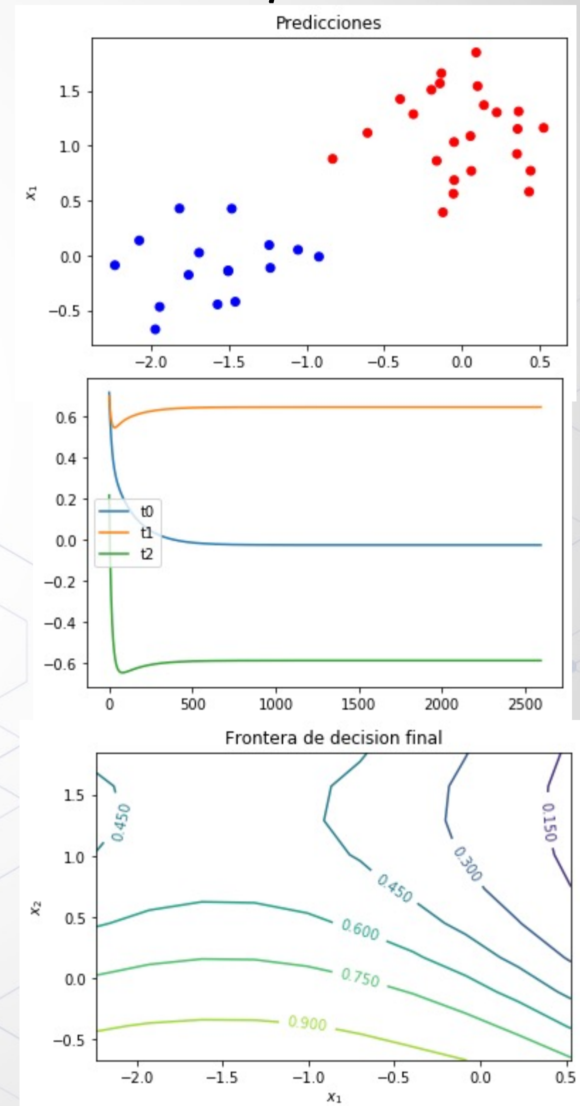
$\gamma = 10$



$\gamma = 30$



$\gamma = 50$



HUB IQ

EngineeringX

Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Regularización – Lasso

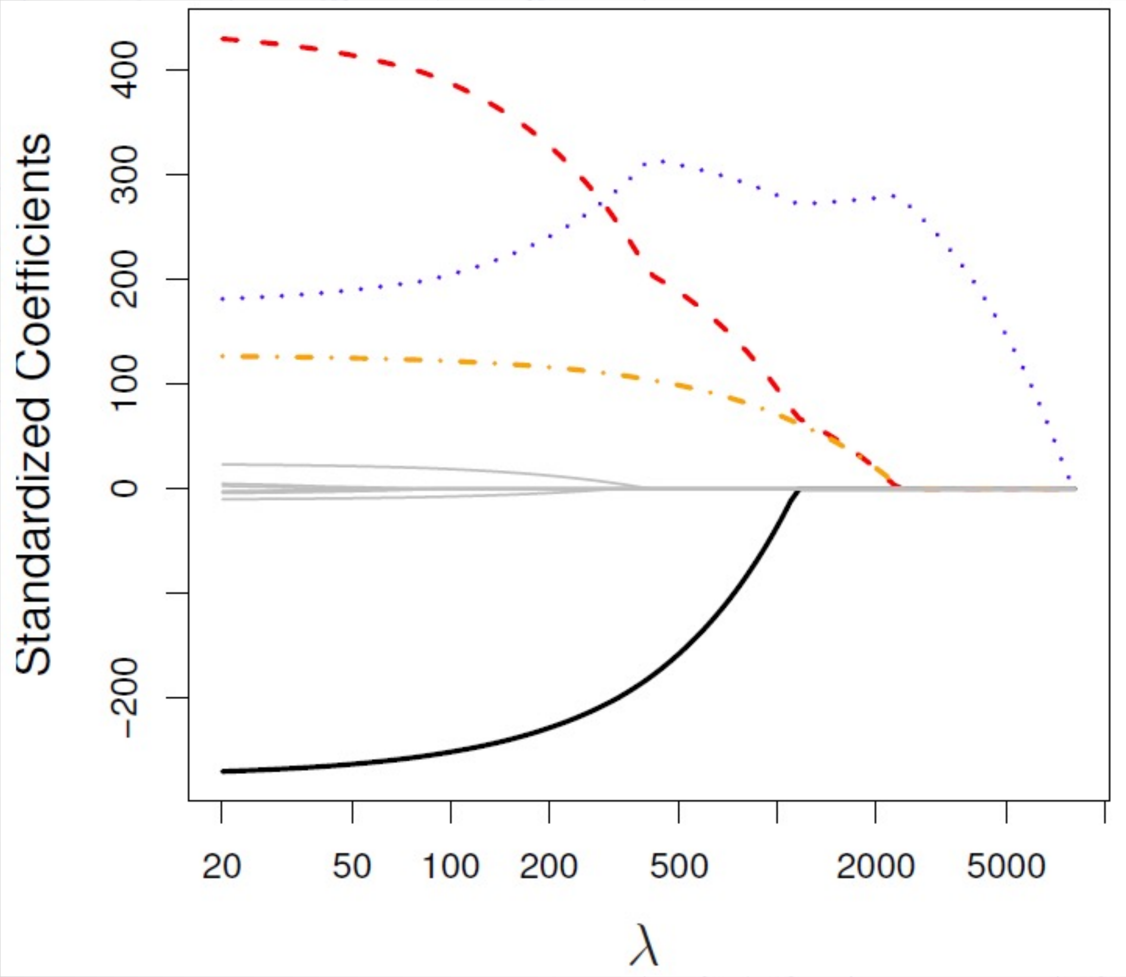
$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

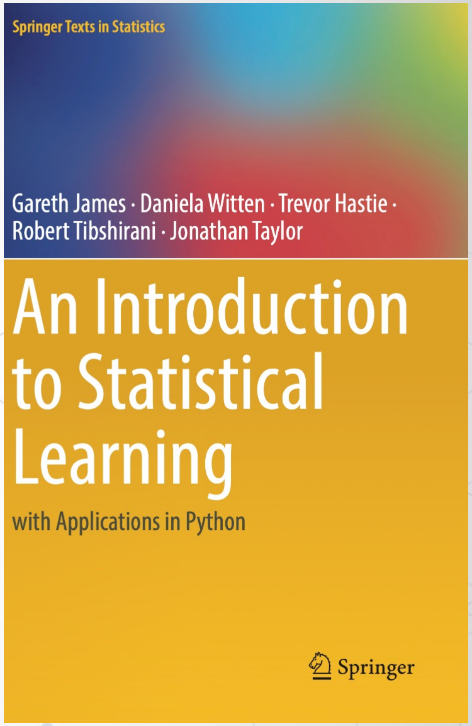
$$E(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}_\theta(x) - y)^2 + \gamma\sum_{i=1}^{m}|\theta|\right]$$

Penalización L1

$$\theta_j \leftarrow \theta_j - \alpha \times \frac{E(\theta)}{\partial\theta_j}$$

Regularización – Lasso





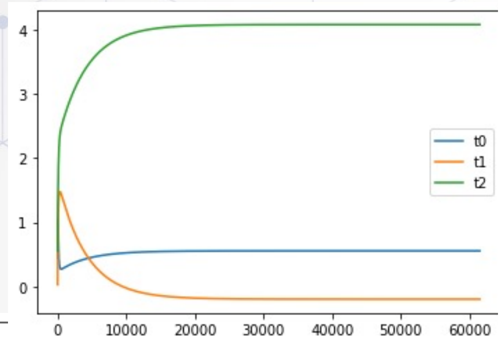
https://www.statlearning.com/

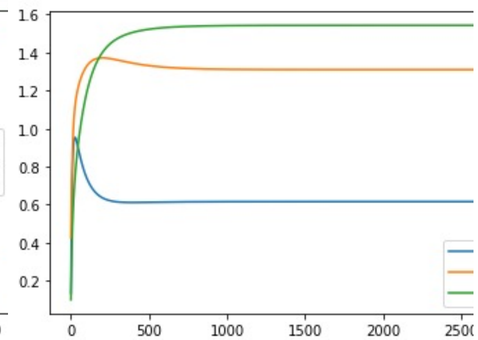
Lasso – para regresión lineal

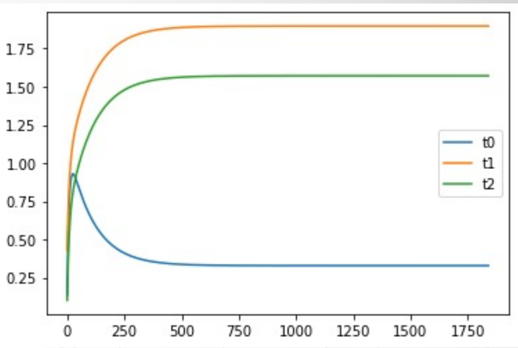
$\gamma = 10$

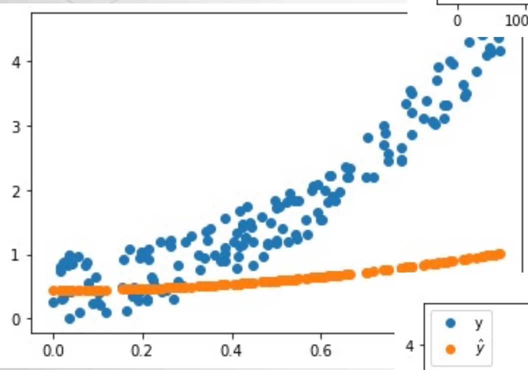
Ridge

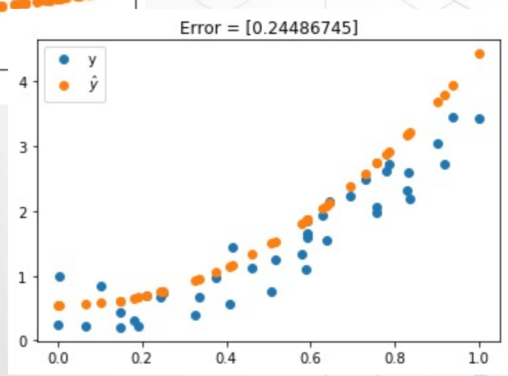
Lasso

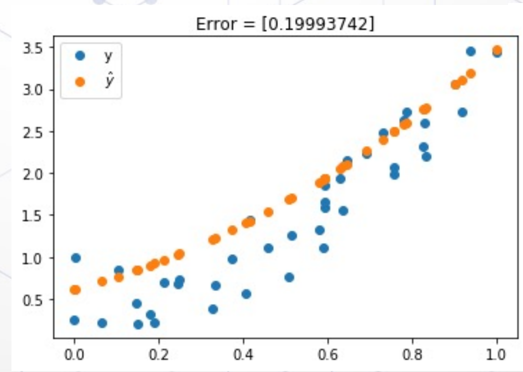


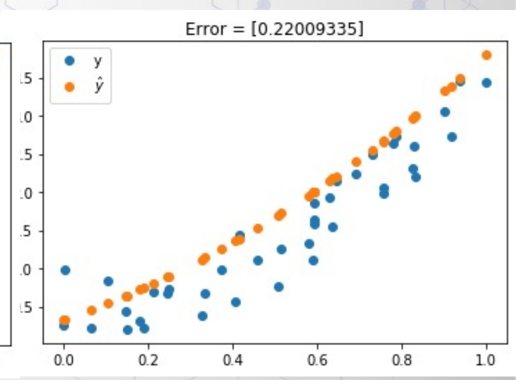






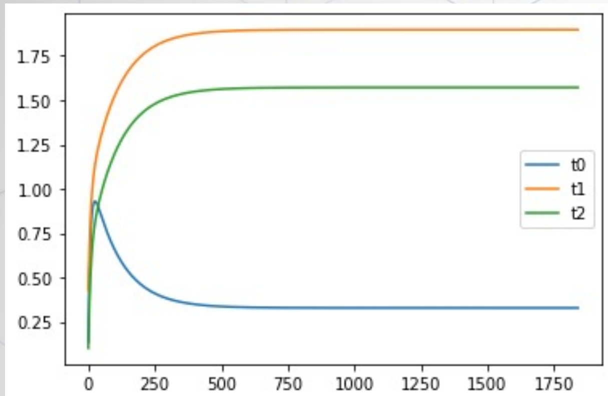


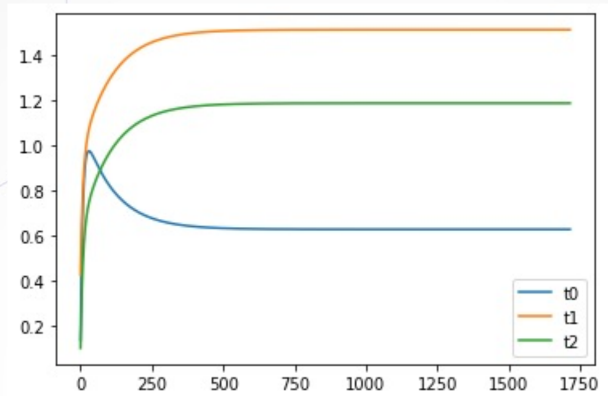


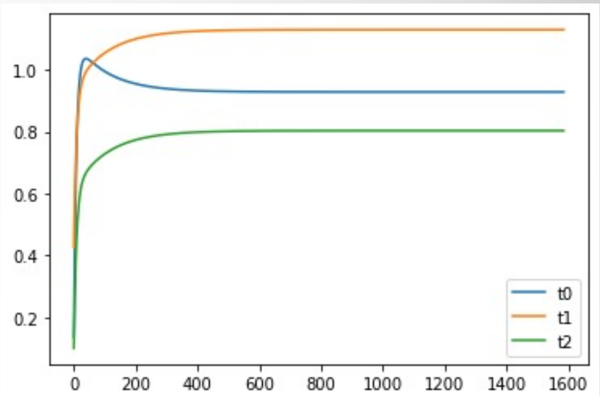


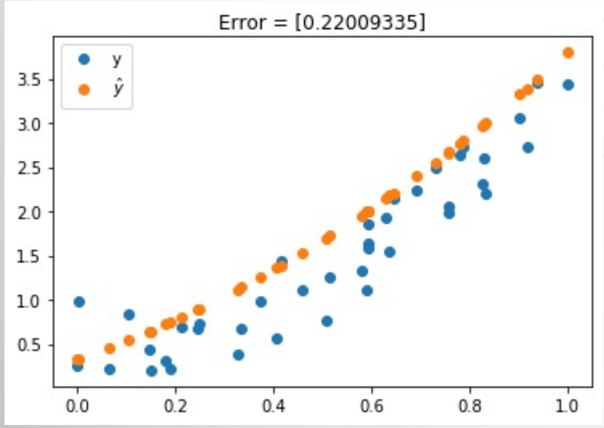
Lasso – para regresión lineal

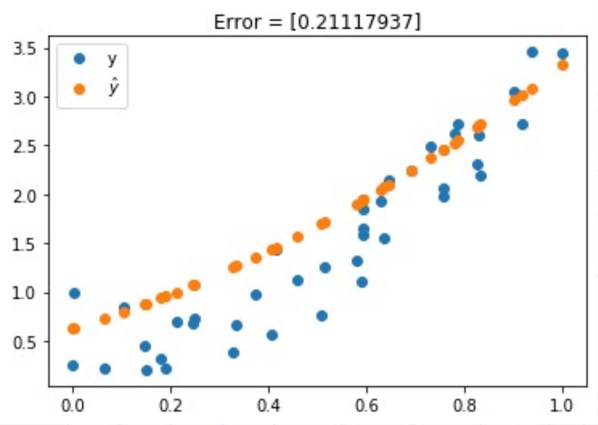
$\gamma = 10$ | $\gamma = 30$ | $\gamma = 50$

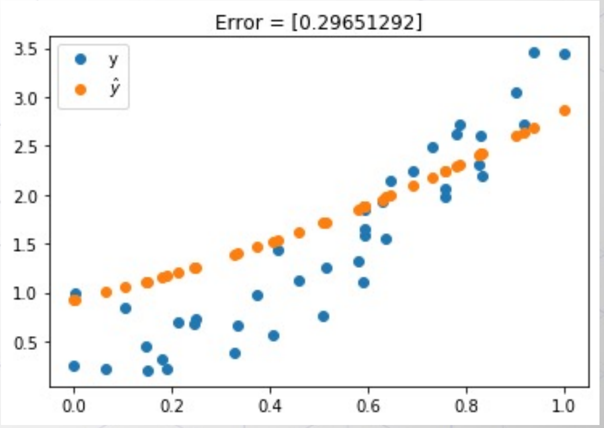






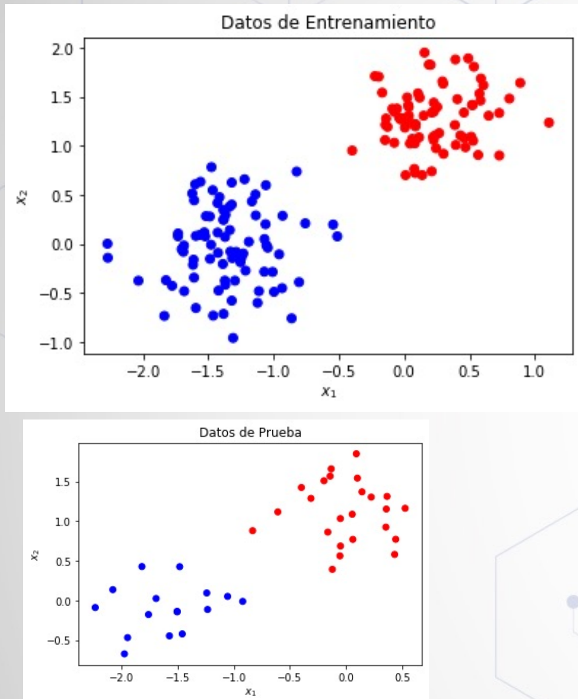


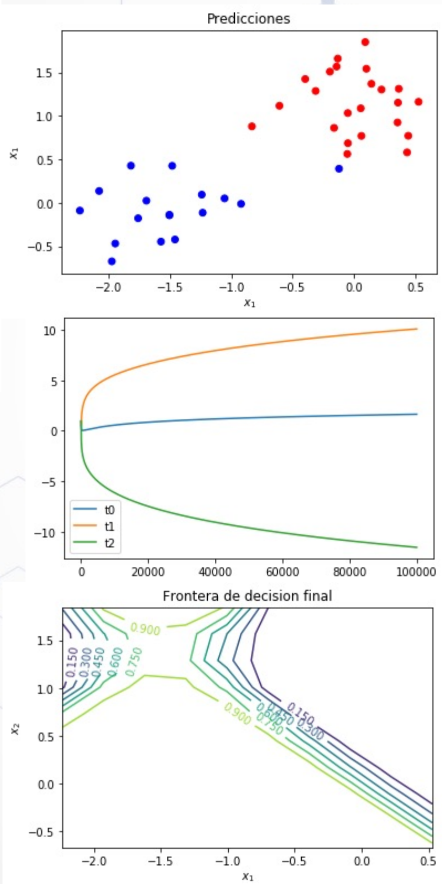




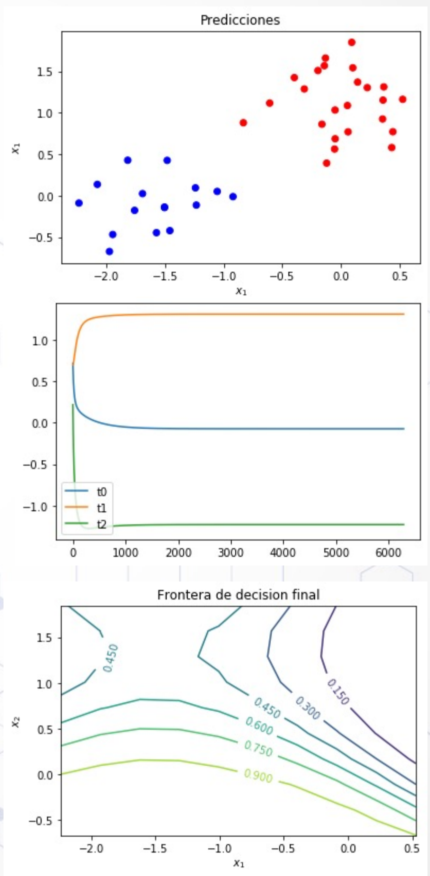
Lasso - para regresión logística

$\gamma = 10$

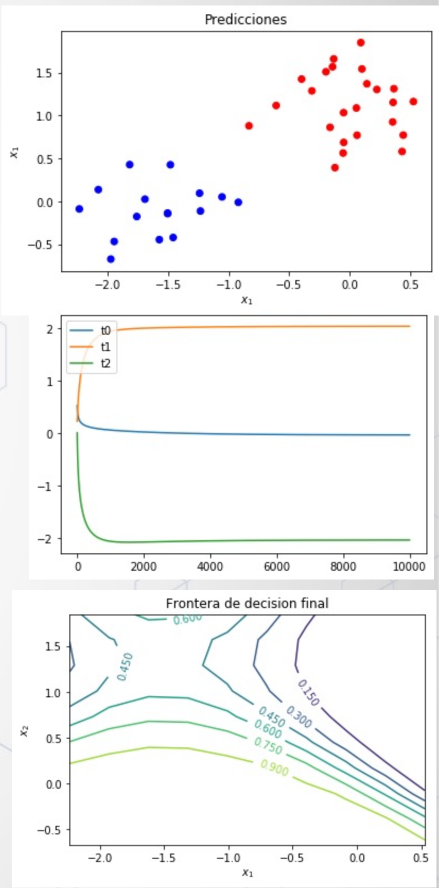




Ridge

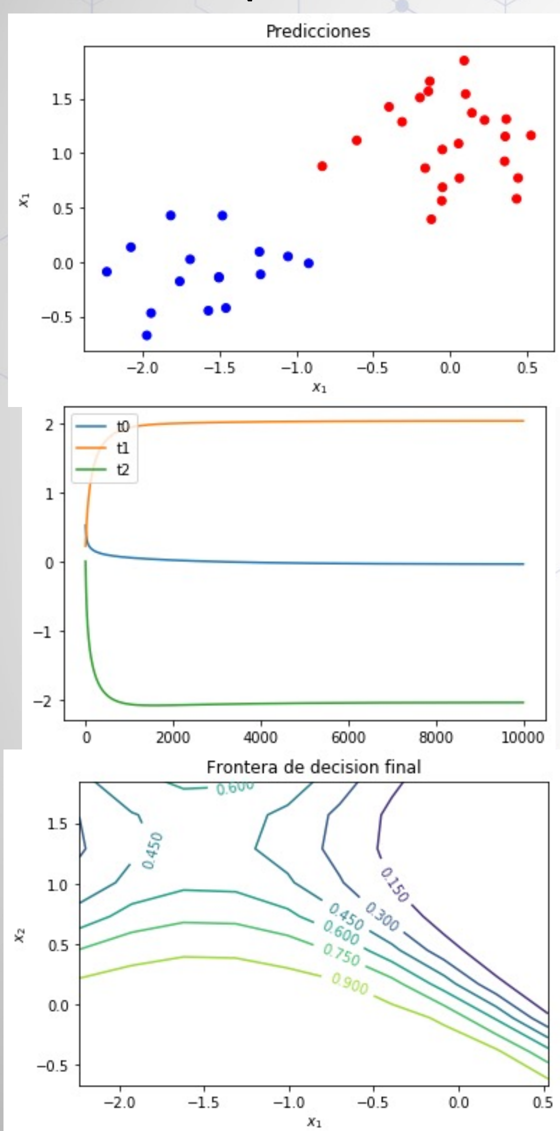


Lasso

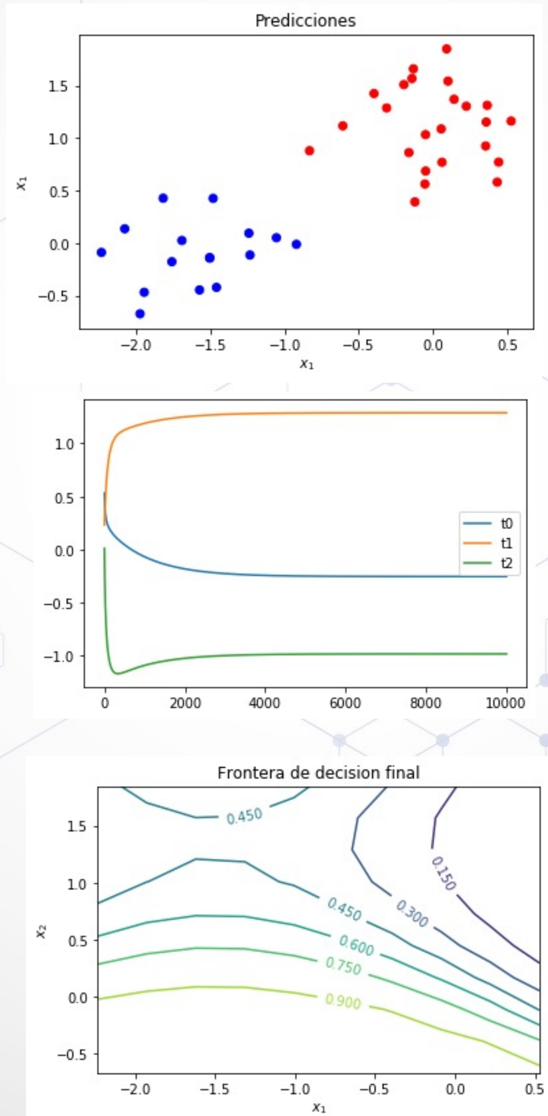


Lasso - para regresión logística

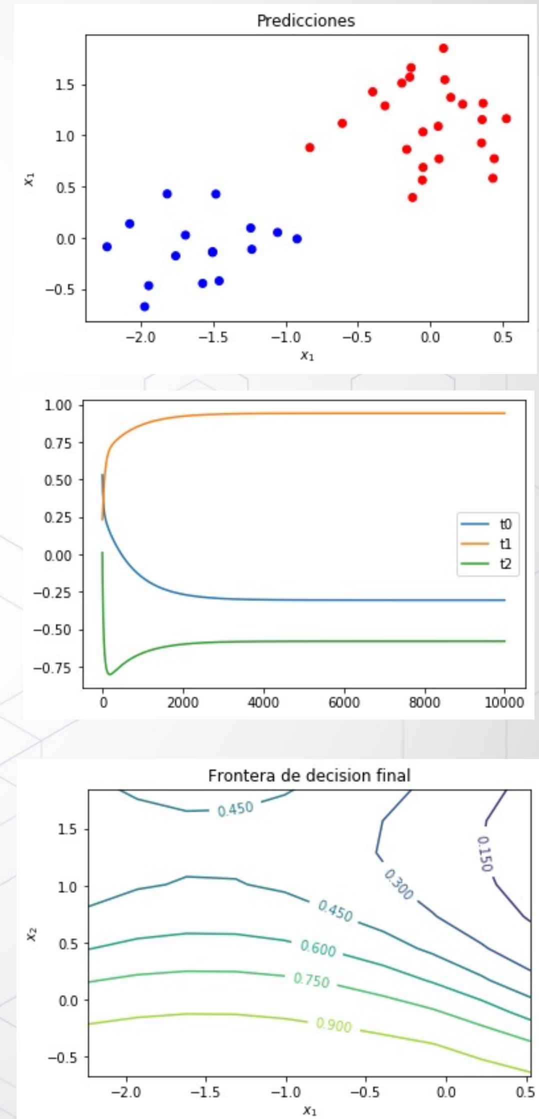
$\gamma = 10$



$\gamma = 30$



$\gamma = 50$



Código

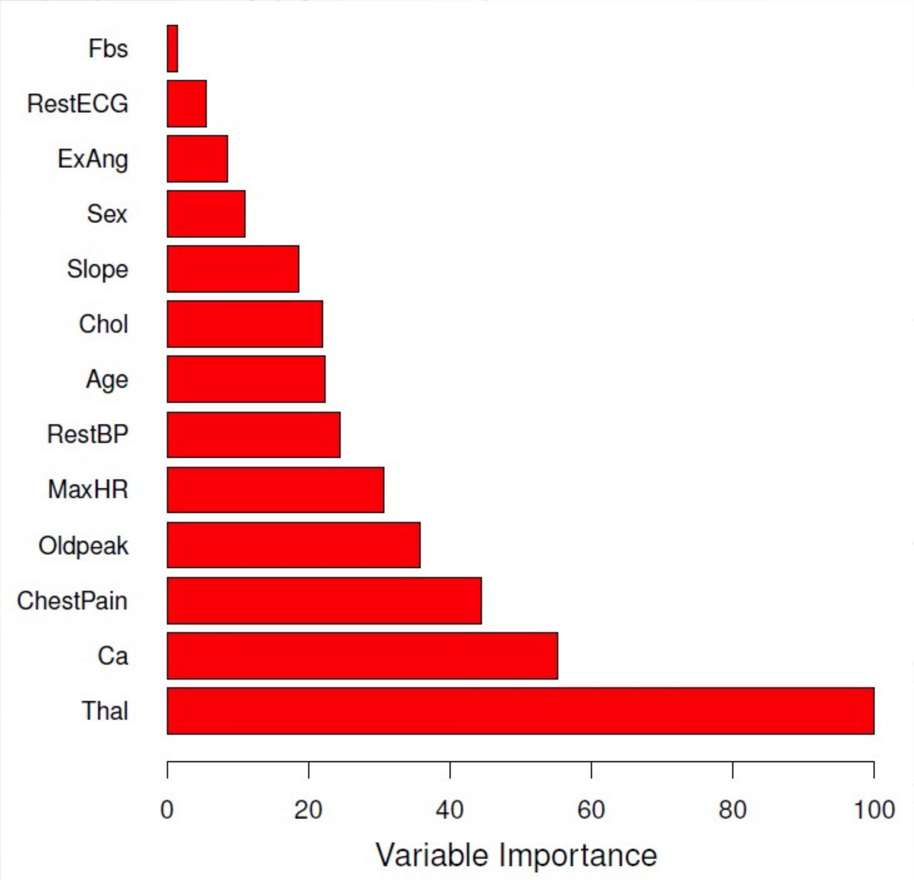
Ejercicio

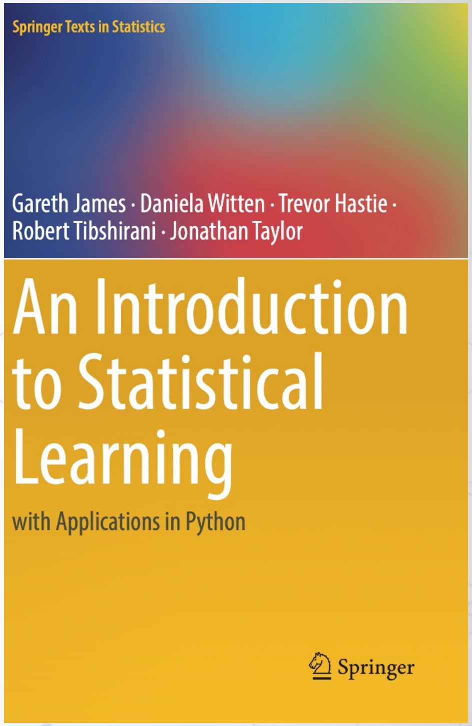
- Implementar regularización Ridge
- Regresión lineal
- Regresión logística

$$\theta_j \leftarrow \theta_j \left(1 - \frac{\alpha\gamma}{m}\right) - \alpha \times \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_\theta(x) - y)x$$

$$j \geq 1$$

Selección de atributos – Random Forest





HUB IQ

EngineeringX
Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

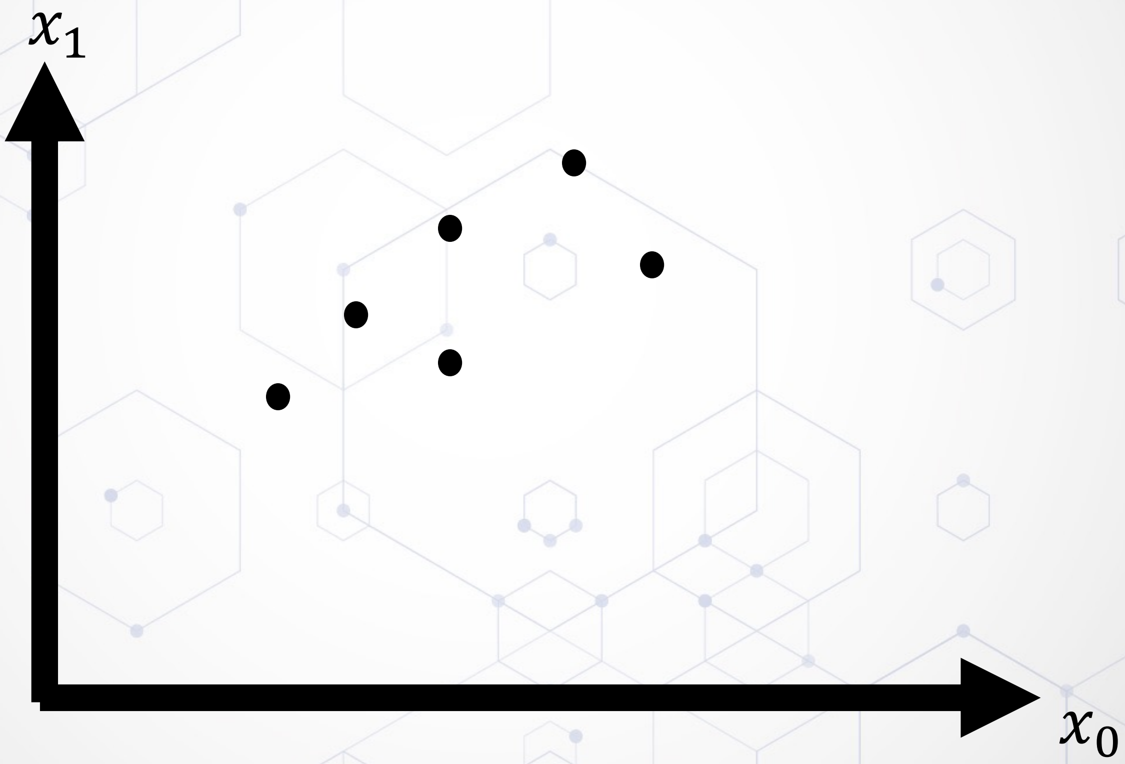
Generación de atributos

- Mapear los atributos originales a un nuevo espacio de menor dimensión

Principal Component Analysis (PCA)

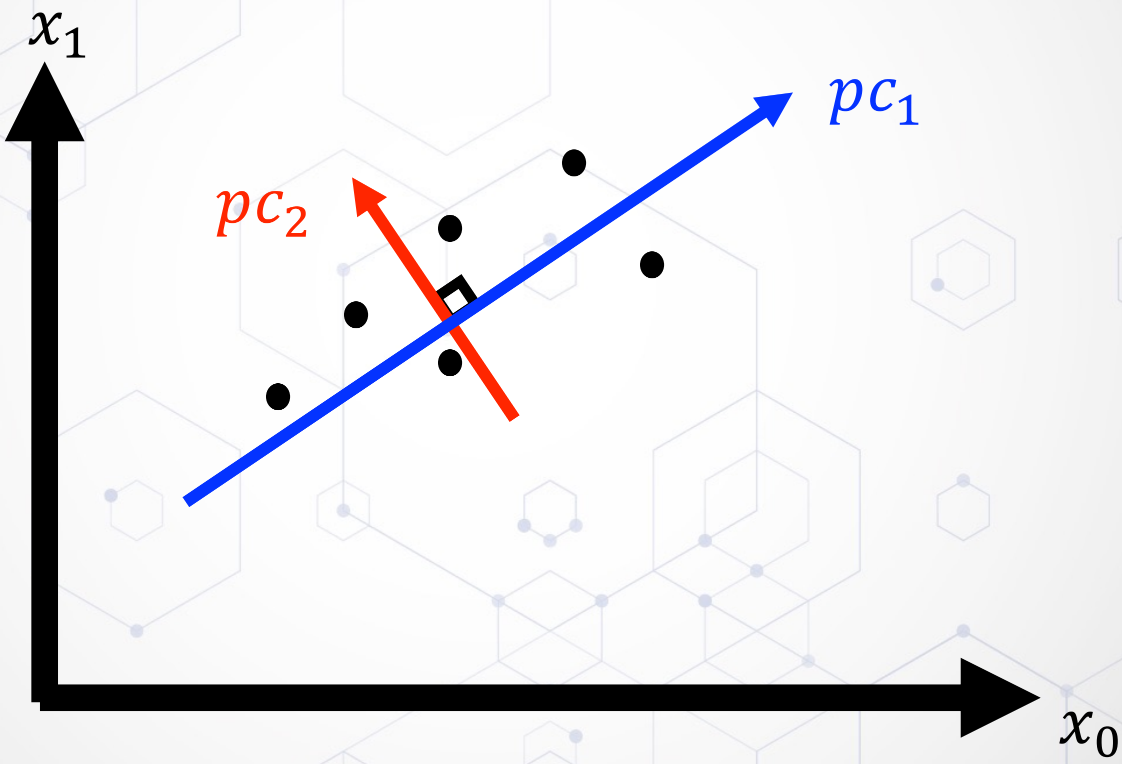
Idea General

- Datos en dos dimensiones



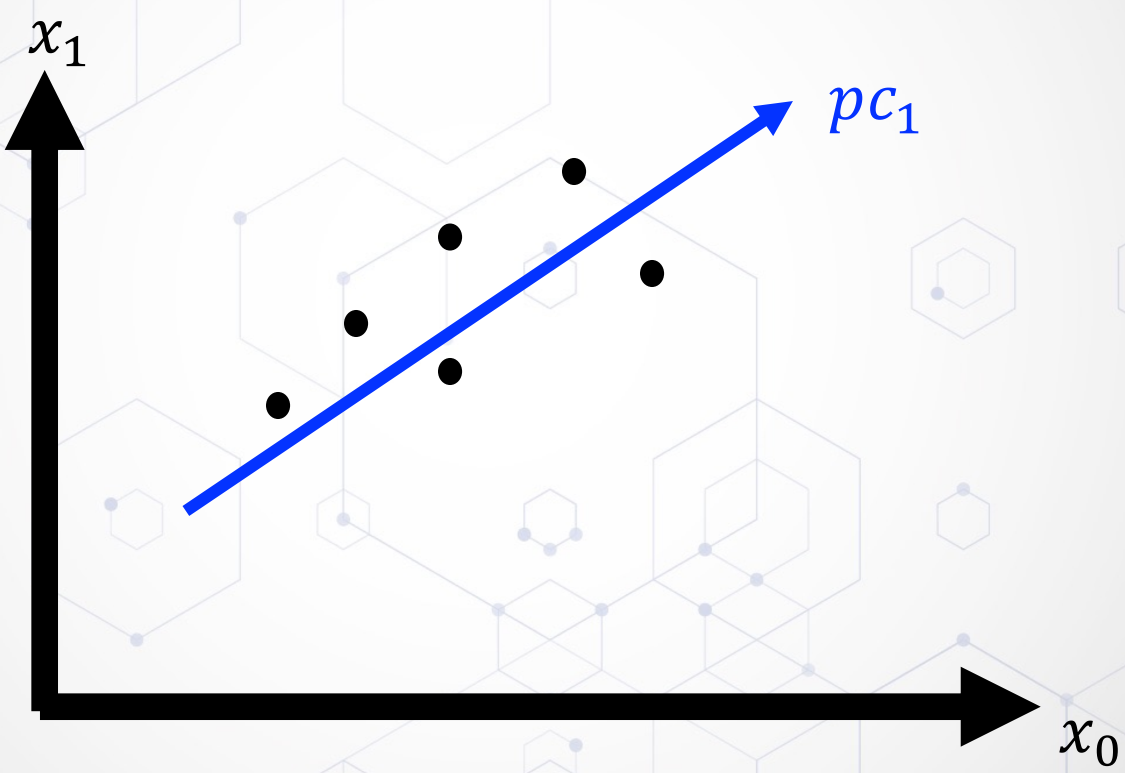
Idea General

- Bases en otro espacio de dimensión -> principal components



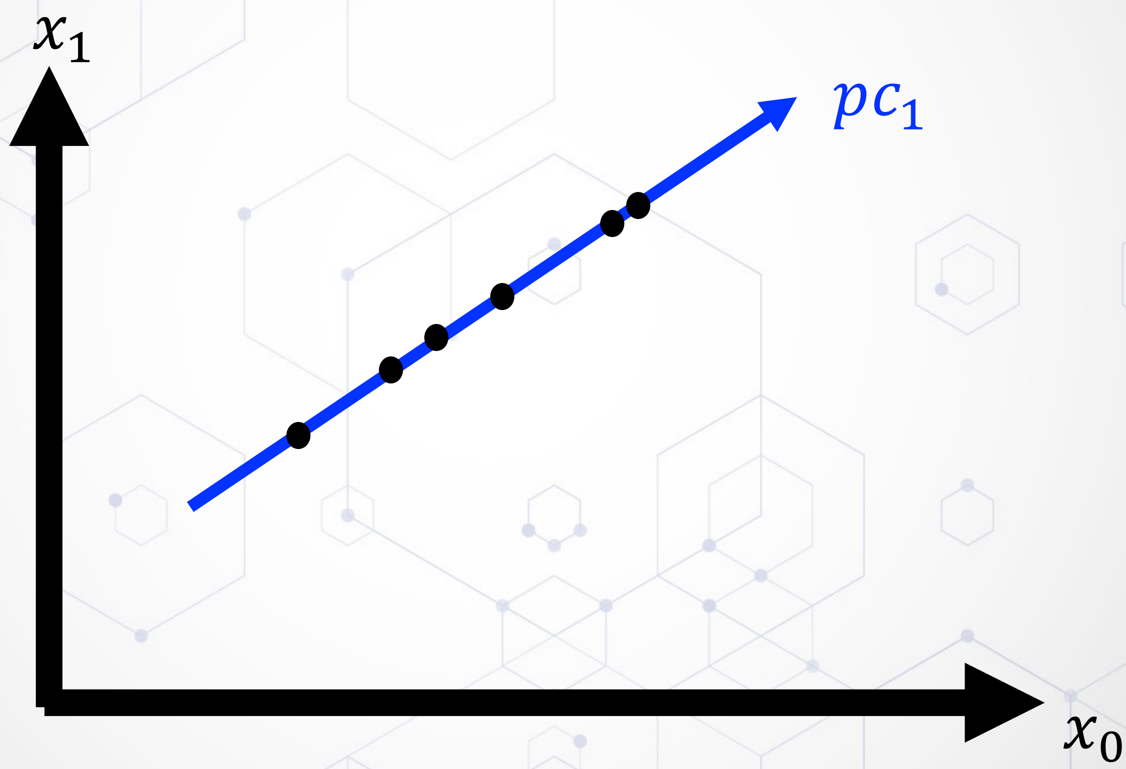
Idea General

- Una dimensión / una base



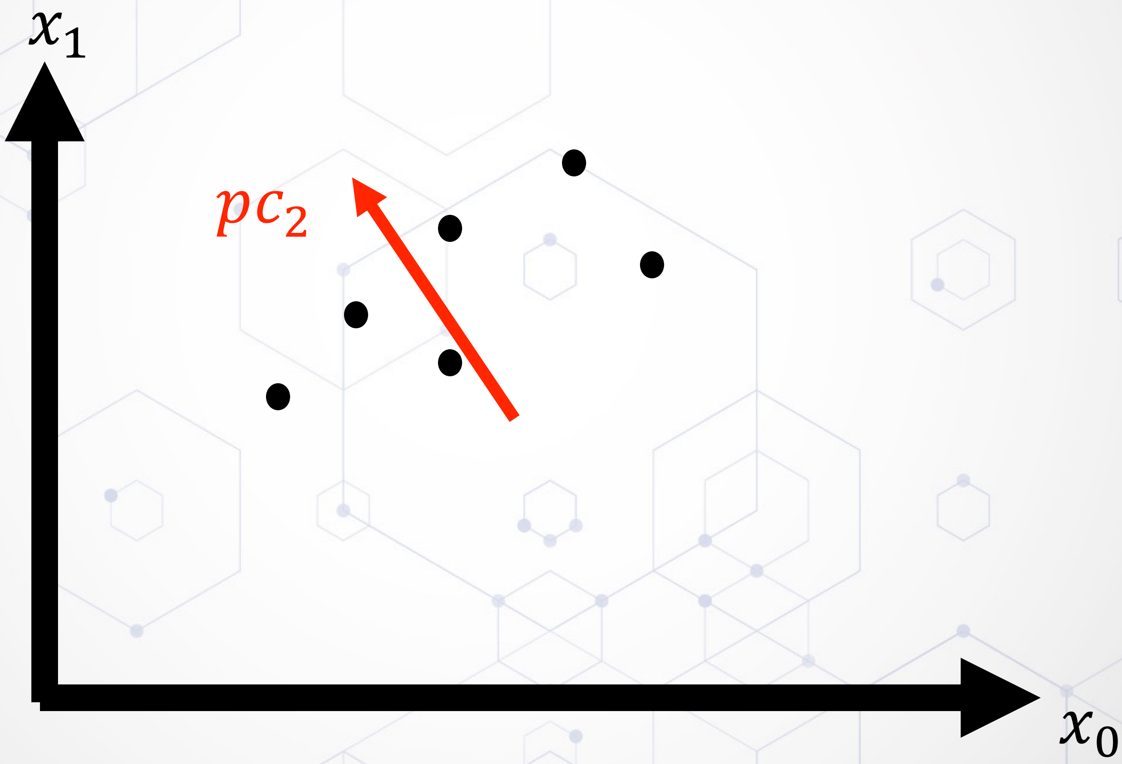
Idea General

- Proyección en una base



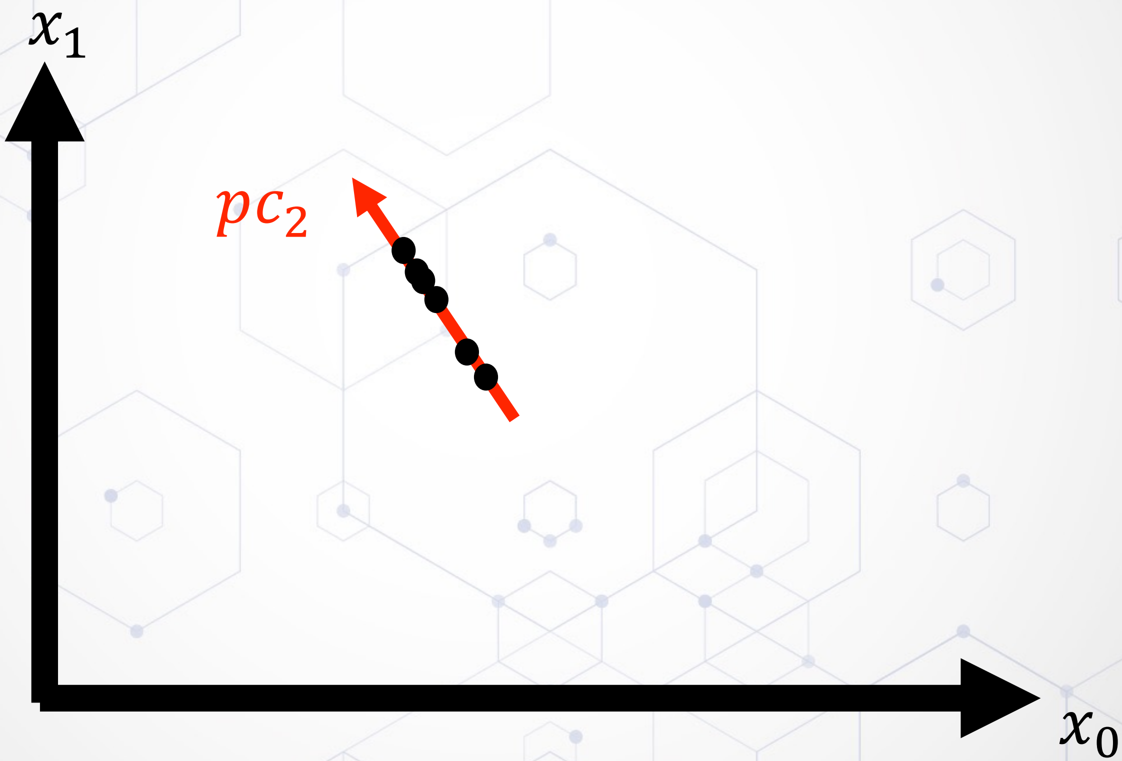
Idea General

- Segunda base



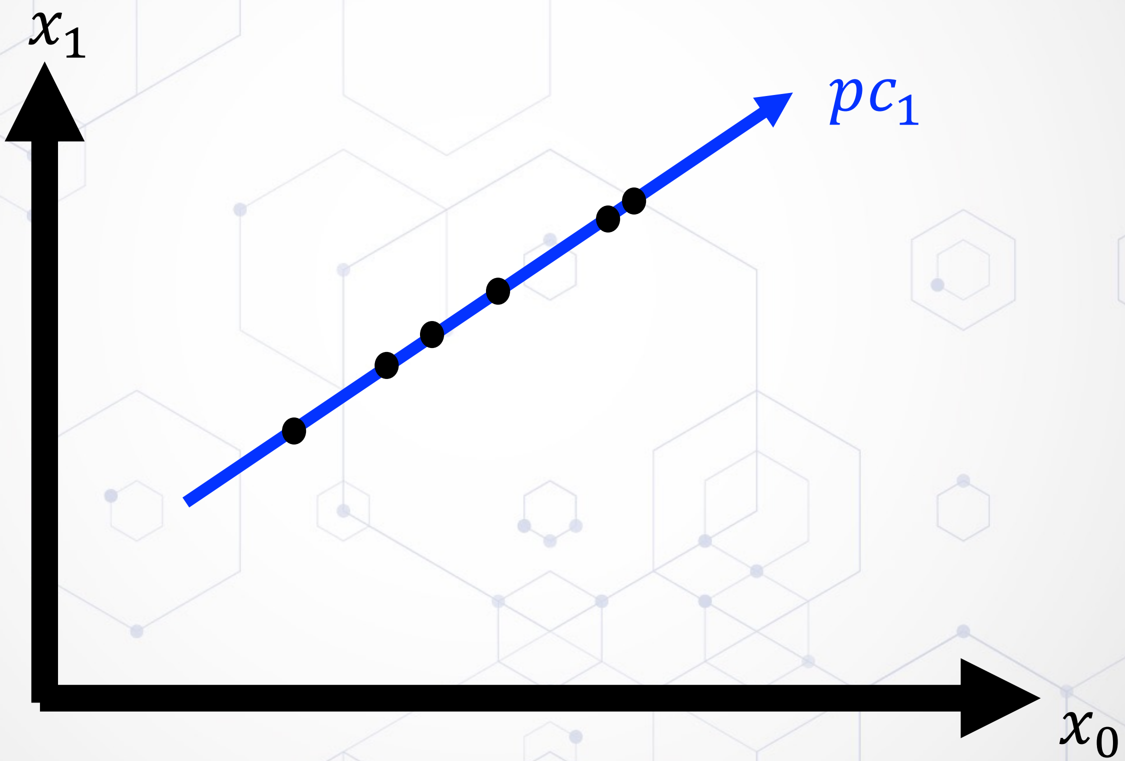
Idea General

- Proyección en la segunda base



Idea General

- "Mejor" base / componente principal

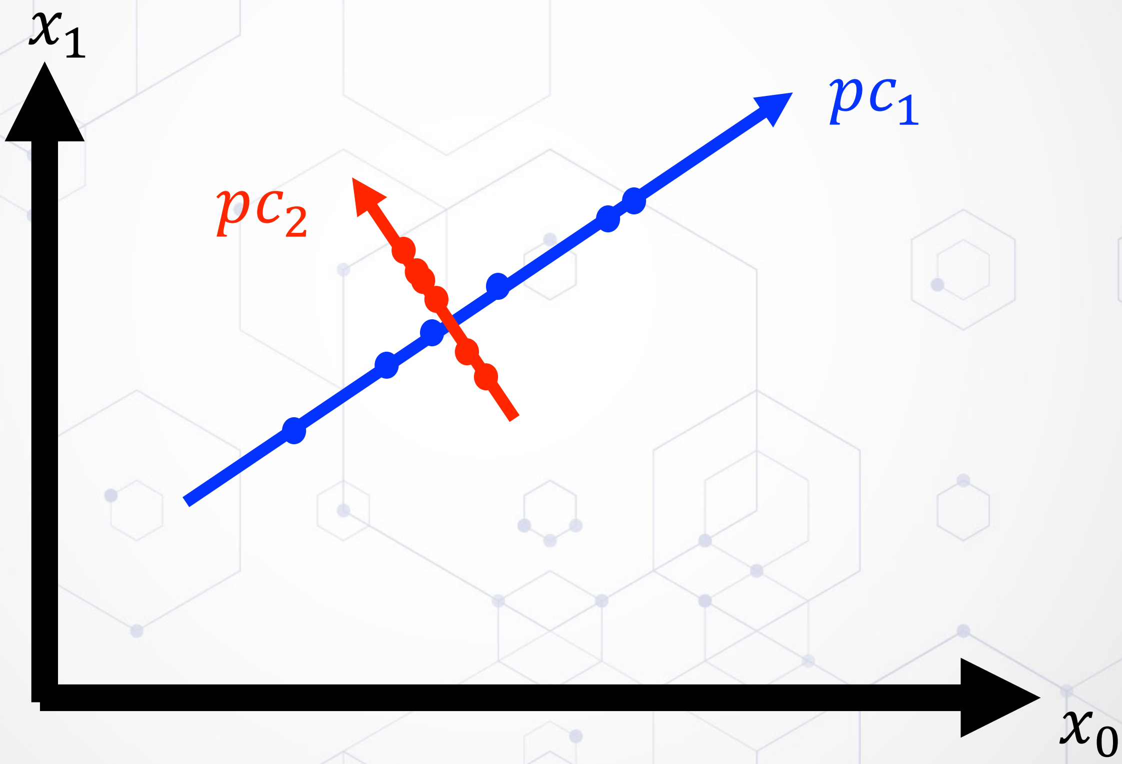


Idea General

- 3D -> 2D
- nD -> mD
  - Tal que n >> m

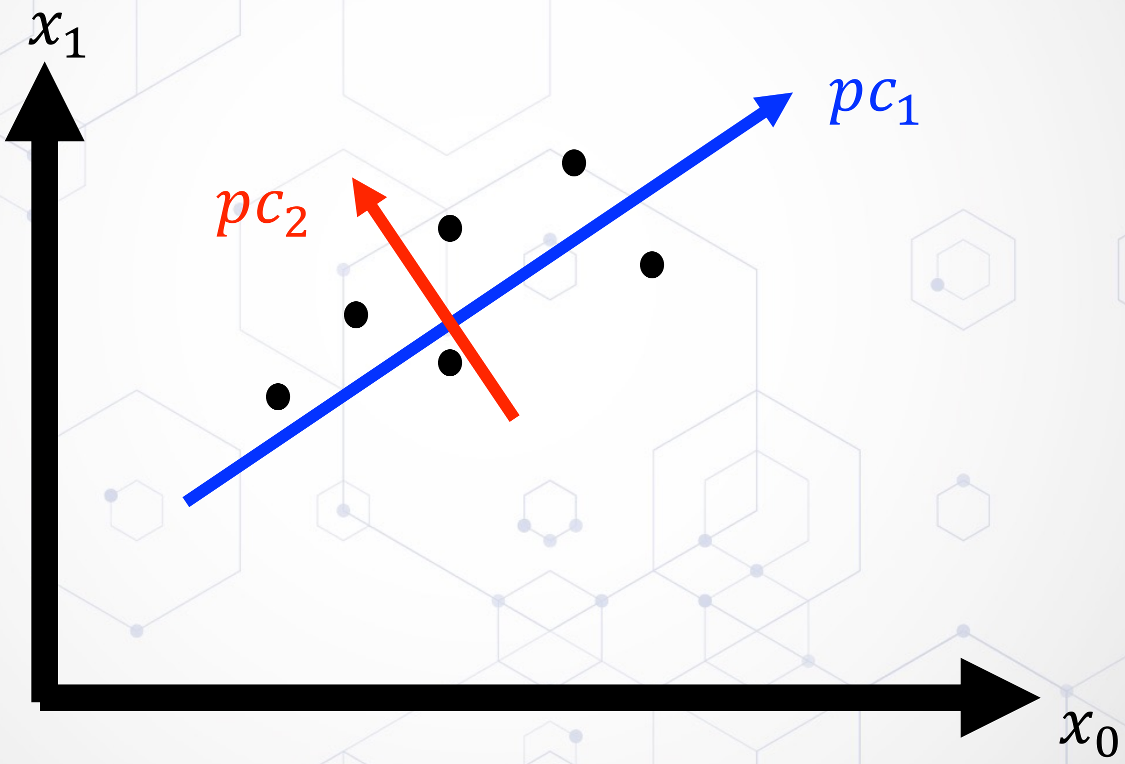
Idea General

- Recuperación de la información



Idea General

- Recuperación de la información



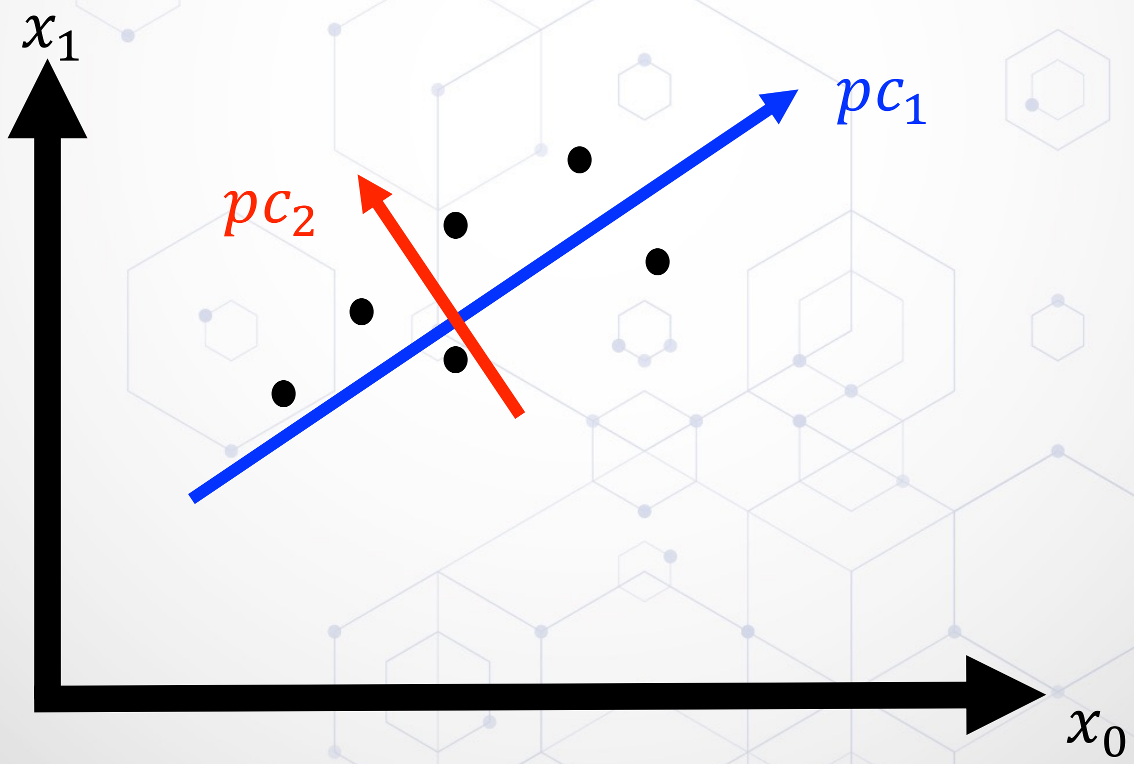
¿Cómo?

Idea General

- PCA es un método que encuentra un conjunto de **bases** que:
  - maximizan la varianza de los datos originales
    - encuentran las direcciones de mayor variación
  - son ortogonales entre sí
- Las bases o componentes principales (Principal components) se obtienen de los **vectores propios (eigenvectors)** de la matriz de covarianza … i.e. proporcionan la "dirección" con mayor varianza

Principal Component Analysis

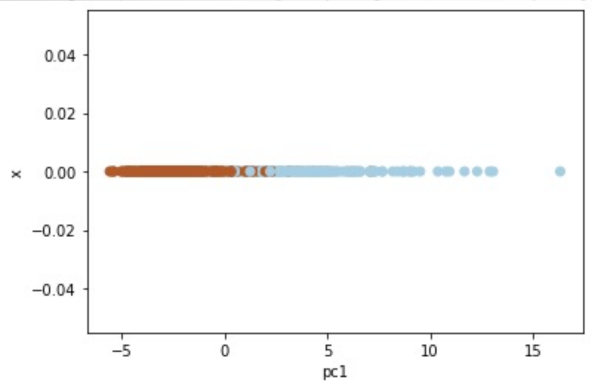
- Eigenvector (vector propio)
- Eigenvalue (valor propio)

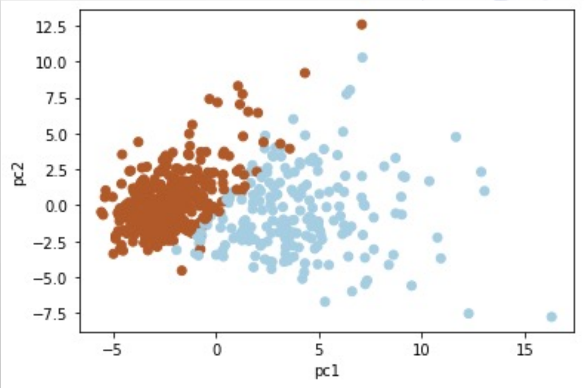


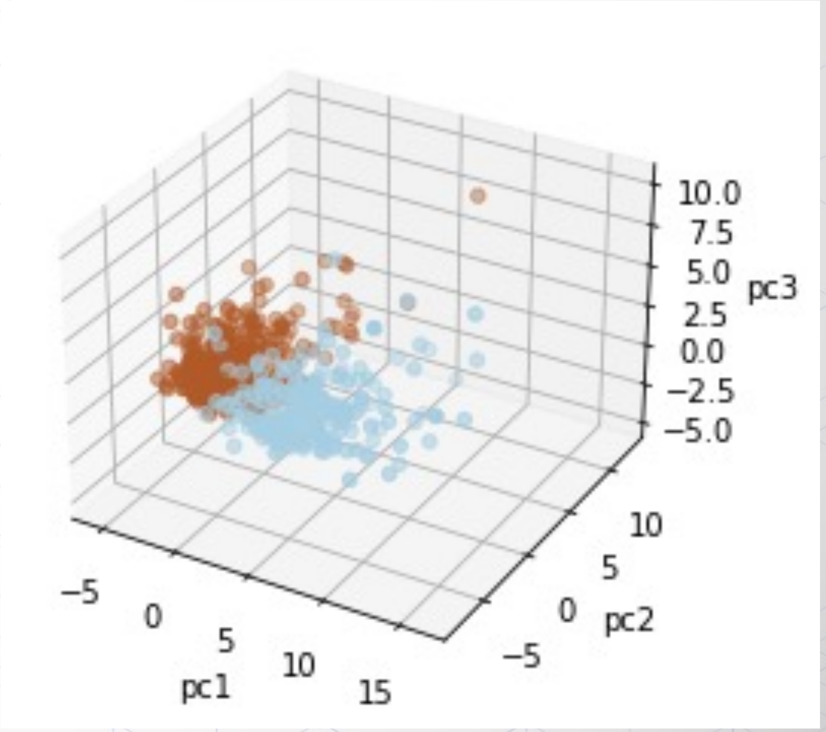
Código

Ejemplo

Breast cancer: Dos clases, 30 dim

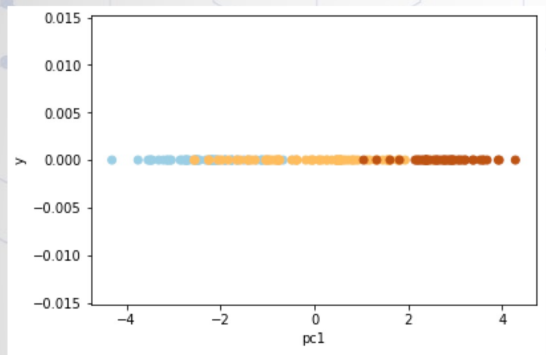


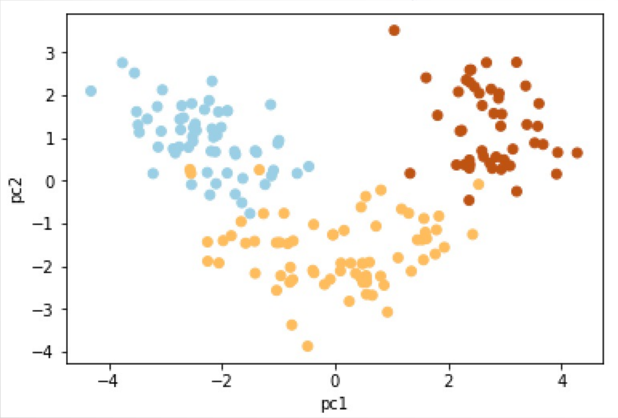


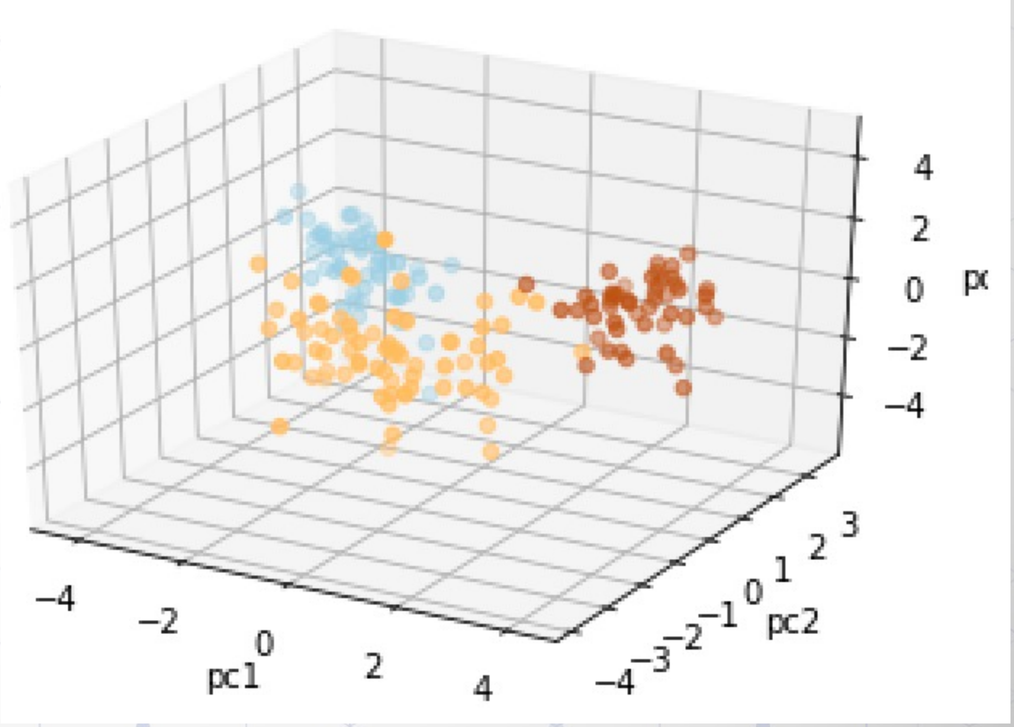


Ejemplo 2

Wine: Tres clases, 13 dim

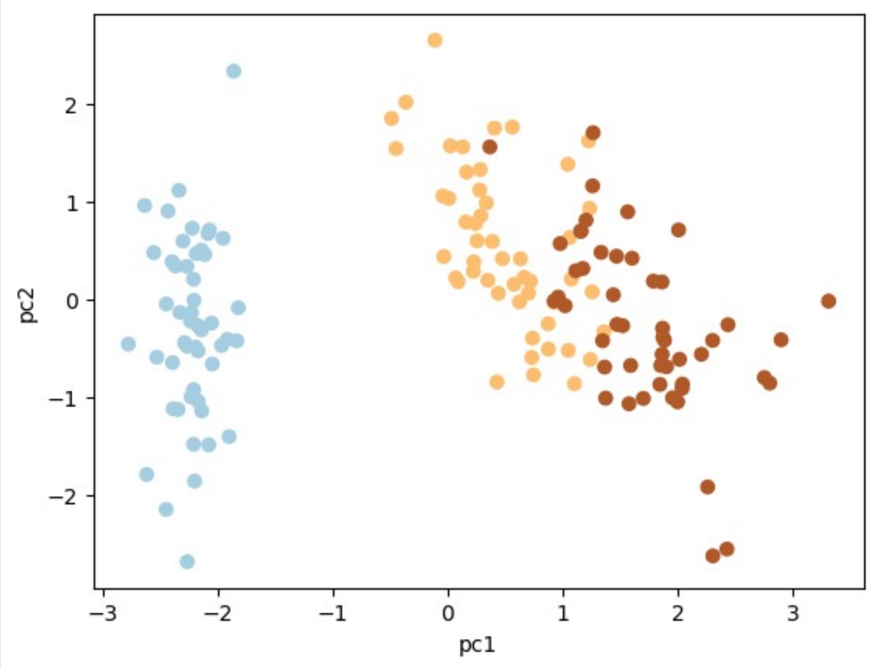


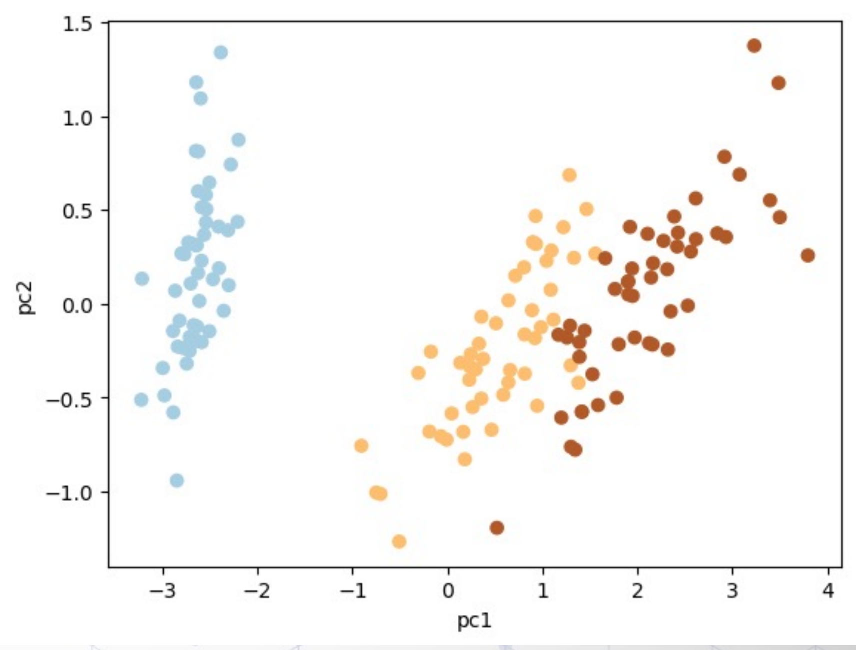




Ejemplo 3

Iris: Tres clases, dim 4






```python
# PCA

# Calcular matriz de covarianza
matriz_cov = np.cov(xNew.T)

# Calcular vectores y valores propios
eig_val, eig_vecs = np.linalg.eig(matriz_cov)
```

```python
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit(X).transform(X)
```

Ejercicio

- PCA para clasificación
- Conjunto de datos con más de 10 variables

PCA

- Simple
- No paramétrico
- Extrae información relevante a partir de un conjunto de datos
- Provee forma de reducir un conjunto de datos (redundante/complicado) a otro con dimensión menor
- Revela estructuras simplificadas (algunas veces ocultas)
- Permite remover ruido, información no relevante
- “también es considerado como un método de aprendizaje no supervisado”

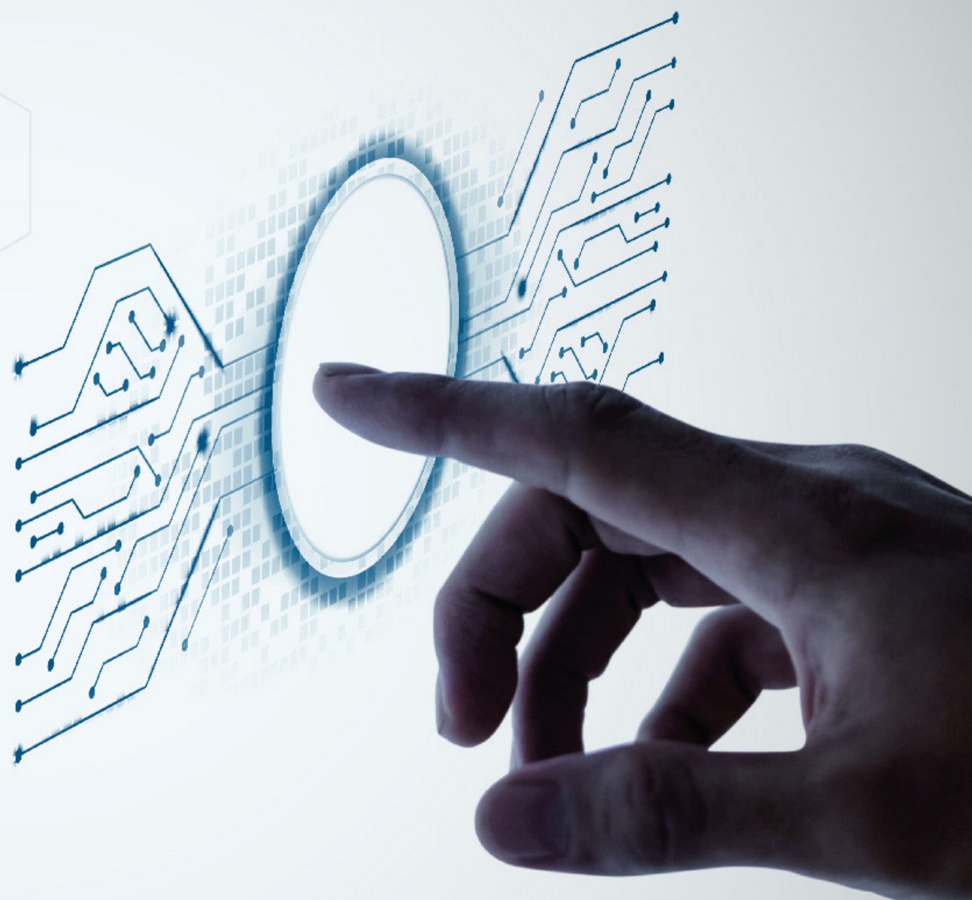
Selección de atributos

- A favor
  - Interpretabilidad de las dimensiones
  - Podría reducir la complejidad de procesamiento
- En contra
  - Perdida potencial de información

Engineering X

Founded by the Royal Academy of Engineering
and Lloyd's Register Foundation

GRACIAS



https://hubiq.mx/

HUBIQRO HUBIQ HUBIQRO