

ACÀMICA

---

# ¡Bienvenidos/as a Data Science!



# Agenda

---

¿Cómo anduvieron?

Repaso: Machine Learning

Explicación: Regresión y Métricas de Evaluación para Regresión

Hands-On

Break

Lanzamiento Entrega 3 + Dudas comunitarias

Cierre



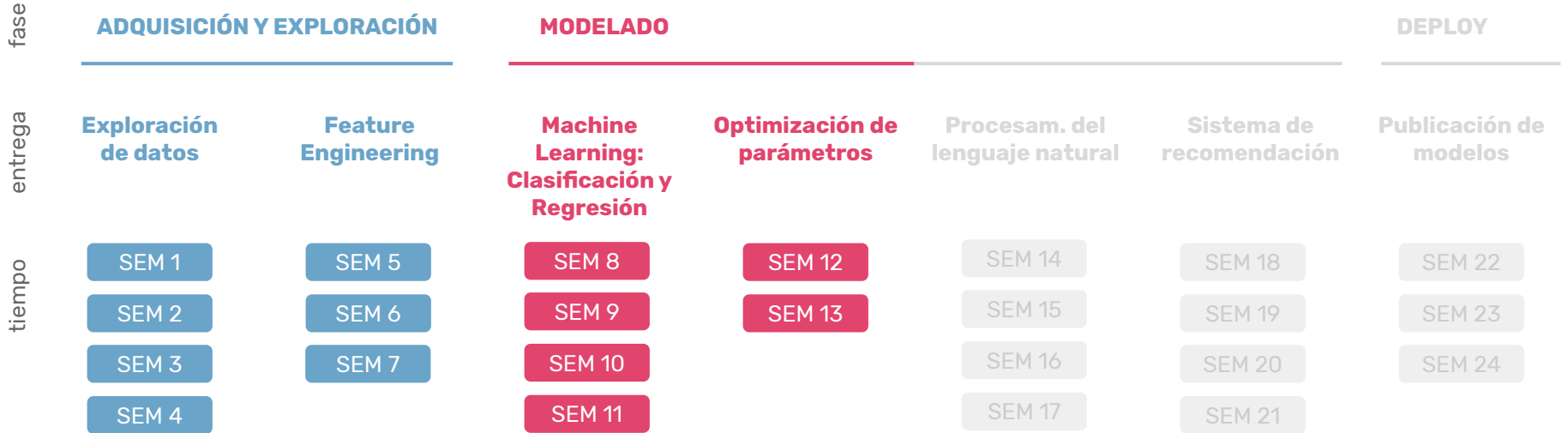
# ¿Cómo anduvieron?



# Proyecto 2: Modelado



# Análisis Exploratorio de Datos (EDA)



# Proyecto EDA: Hoja de ruta

---

## SEM 1 - 7

EDA:

- Python
- Numpy
- Pandas
- Visualización de datos: Matplotlib y Seaborn
- Estadística
- Transformación de Datos
- Outliers

## SEM 8

- Intro a Machine Learning
- Aprendizaje Supervisado: Clasificación
- Árboles de Decisión
- Overfitting y Underfitting, Train/Test Split

## SEM 9

- k-Vecinos más cercanos
- Métricas de Evaluación para Clasificación
- Repaso

Usted  
Está Aquí

## SEM 10

- Aprendizaje Supervisado: Regresión
- Métrica de Evaluación para Regresión

## SEM 11

Entrega 3

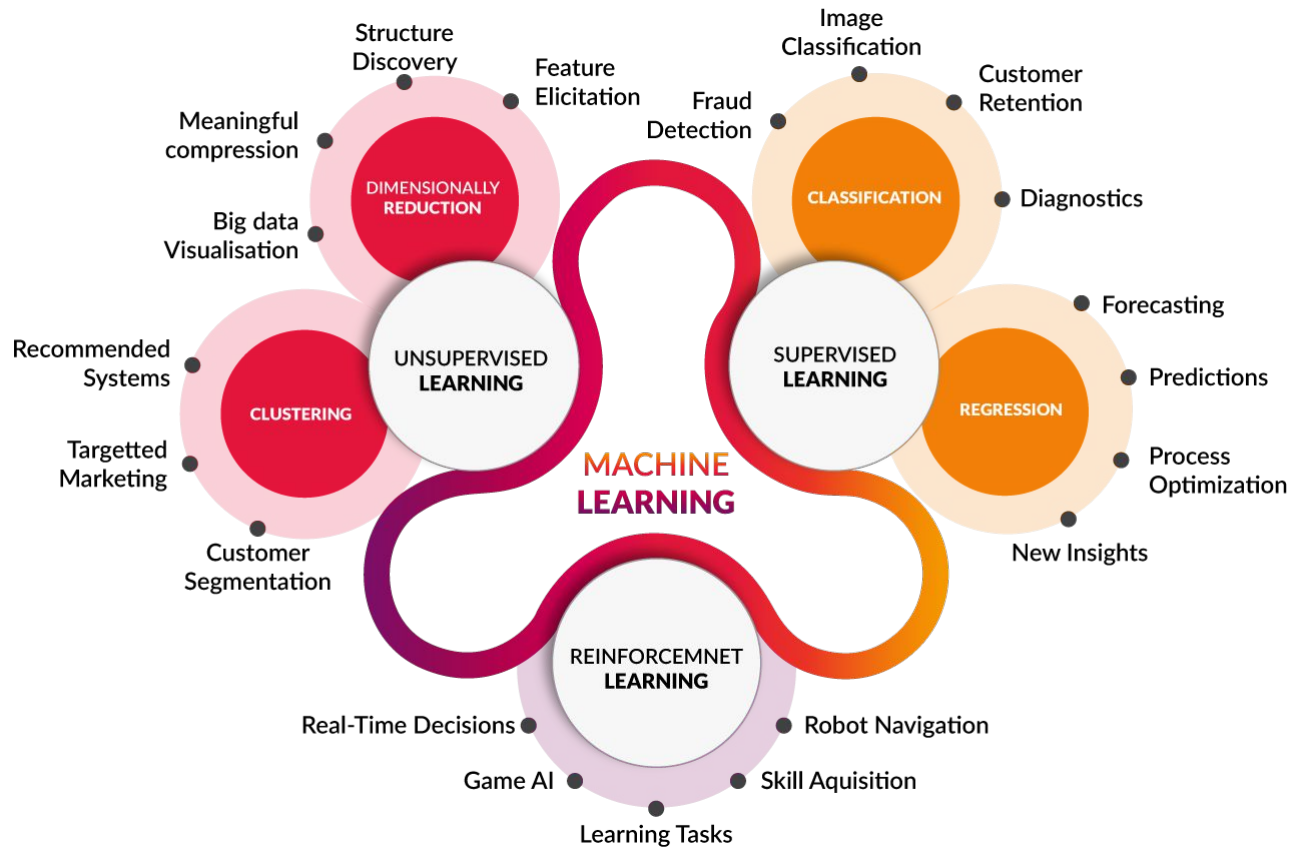


# Repaso: Machine Learning

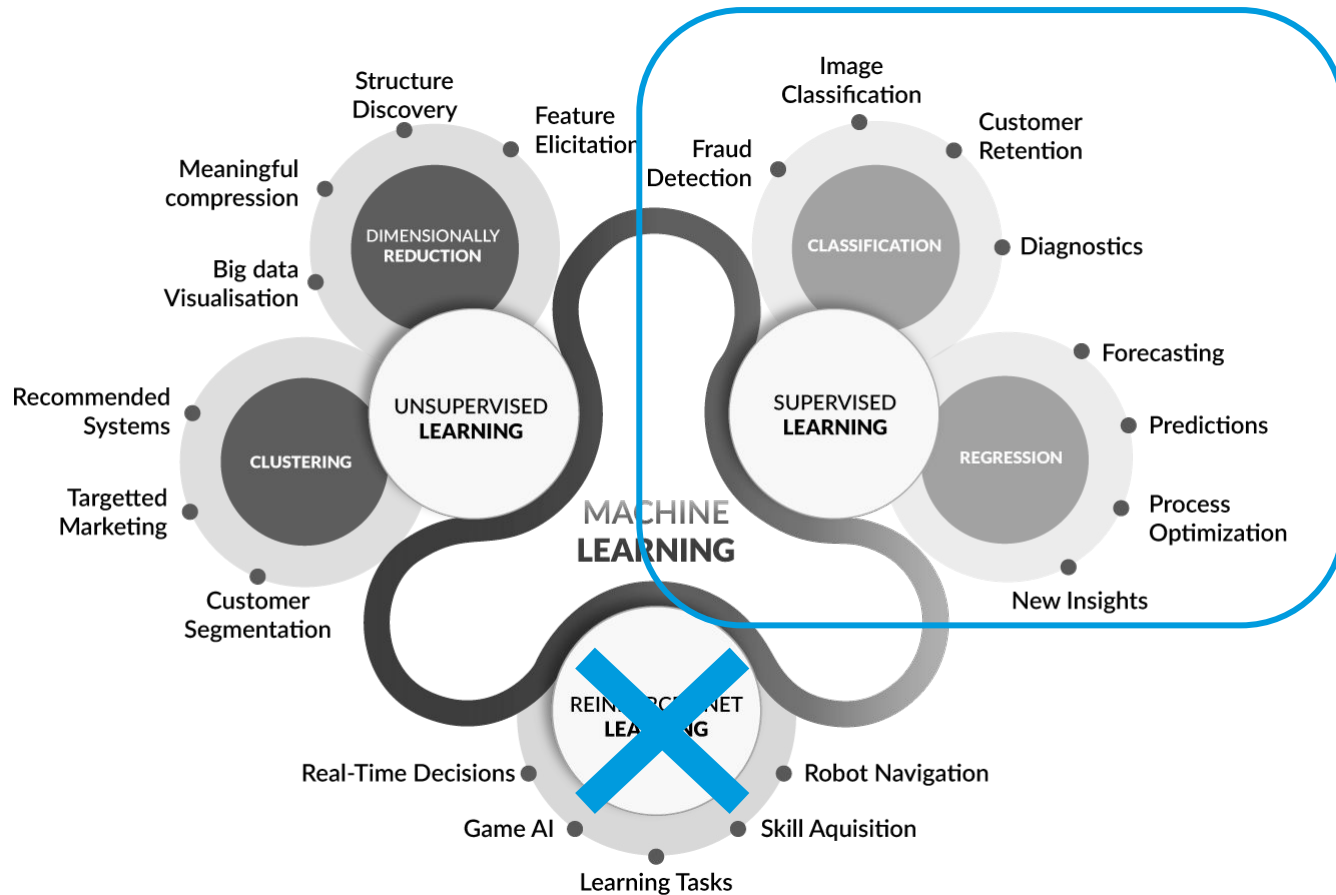




# Mapa



# Mapa



¡Vamos a ver que es  
todo muy parecido!

Machine Learning



Aprendizaje Supervisado



## Clasificación



## Modelos

- Árbol de Decisión
- Support Vector Machines
- k-nearest neighbors
- Random Forest
- Perceptrón
- etc...

Machine Learning



Aprendizaje Supervisado



**Clasificación**



## Modelos

- Árbol de Decisión
- Support Vector Machines
- k-nearest neighbors
- Random Forest
- Perceptrón
- etc...

Machine Learning



Aprendizaje Supervisado



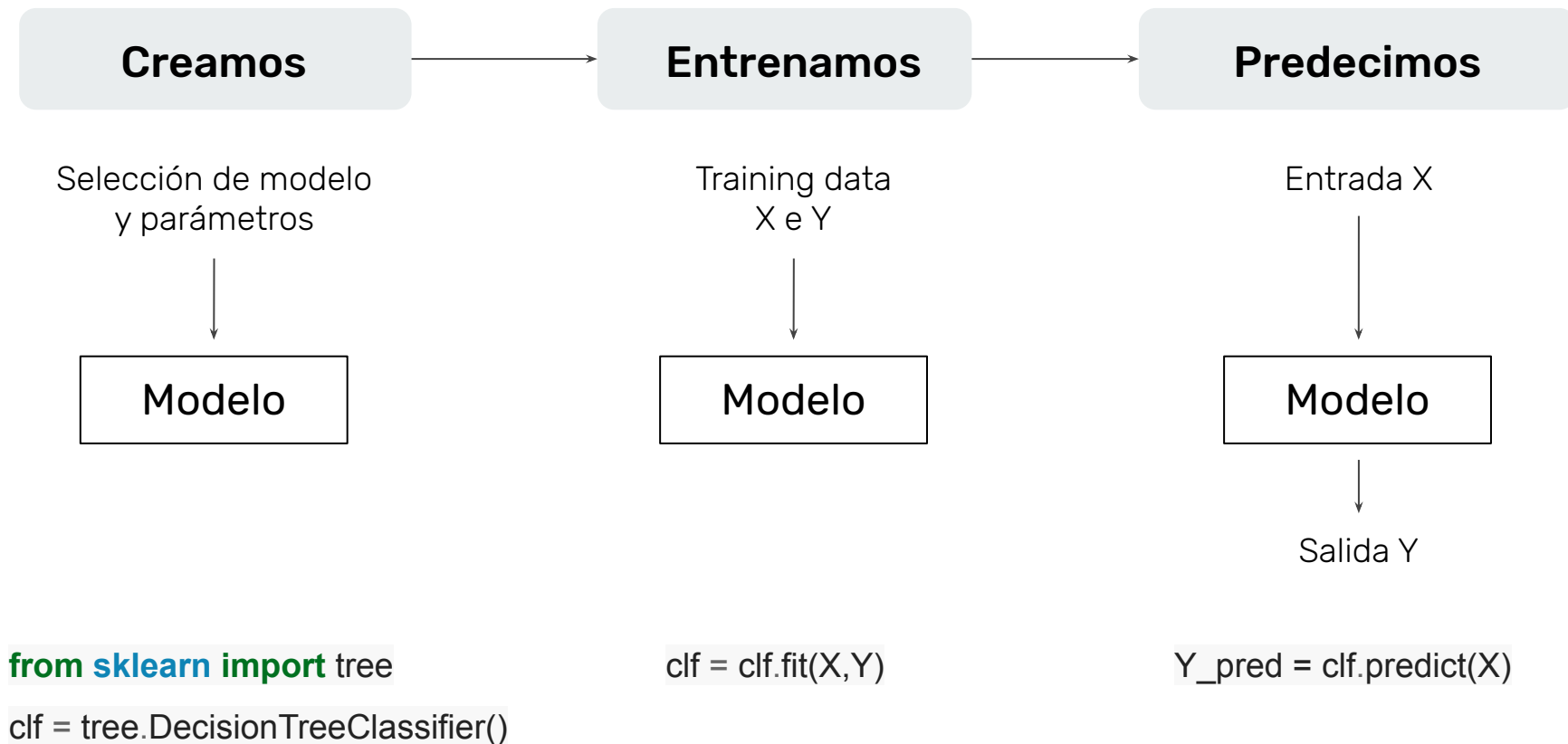
**Regresión**



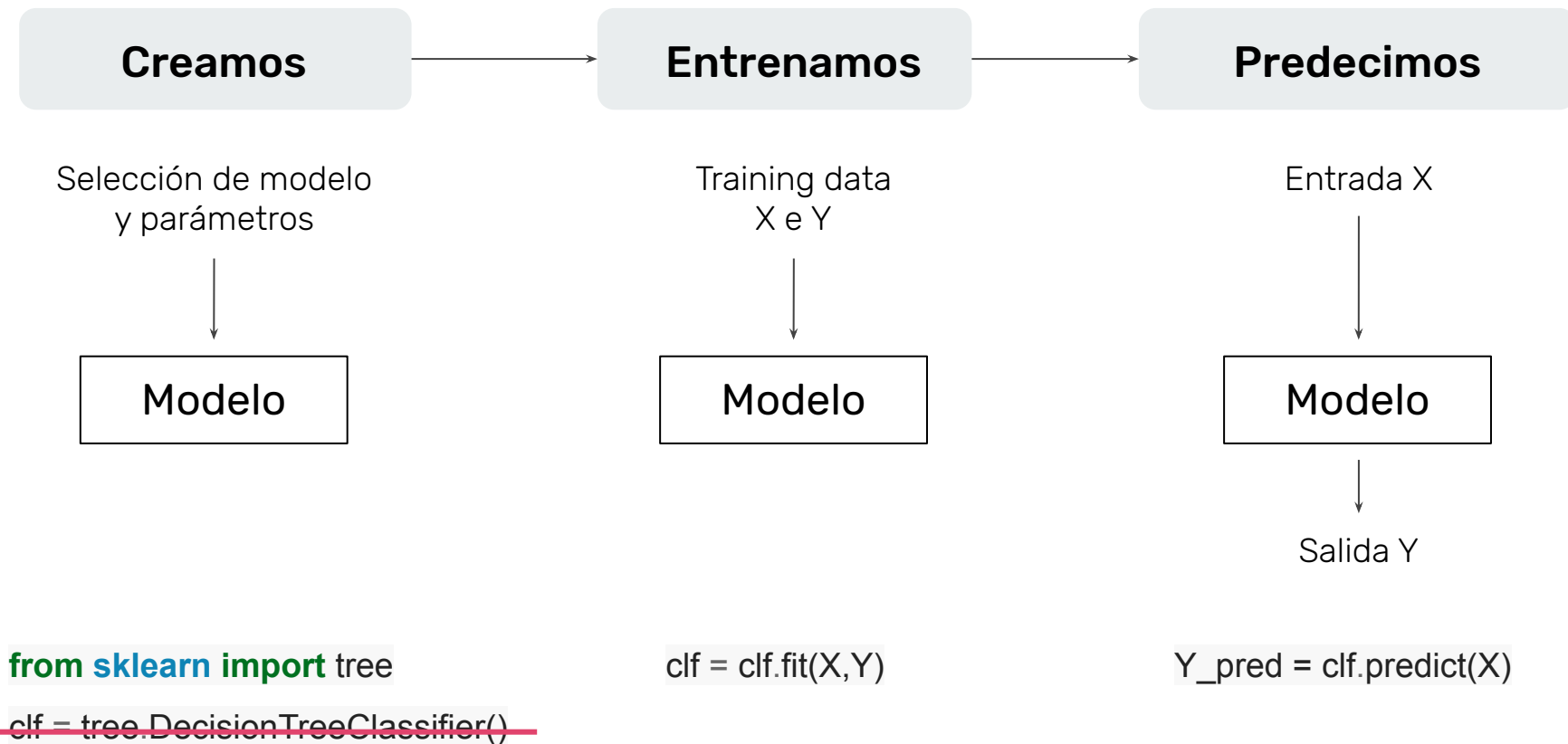
**Modelos**

- Árbol de Decisión
- Support Vector Machines
- k-nearest neighbors
- Random Forest
- Perceptrón
- etc...

# Flujo de trabajo **Scikit Learn**

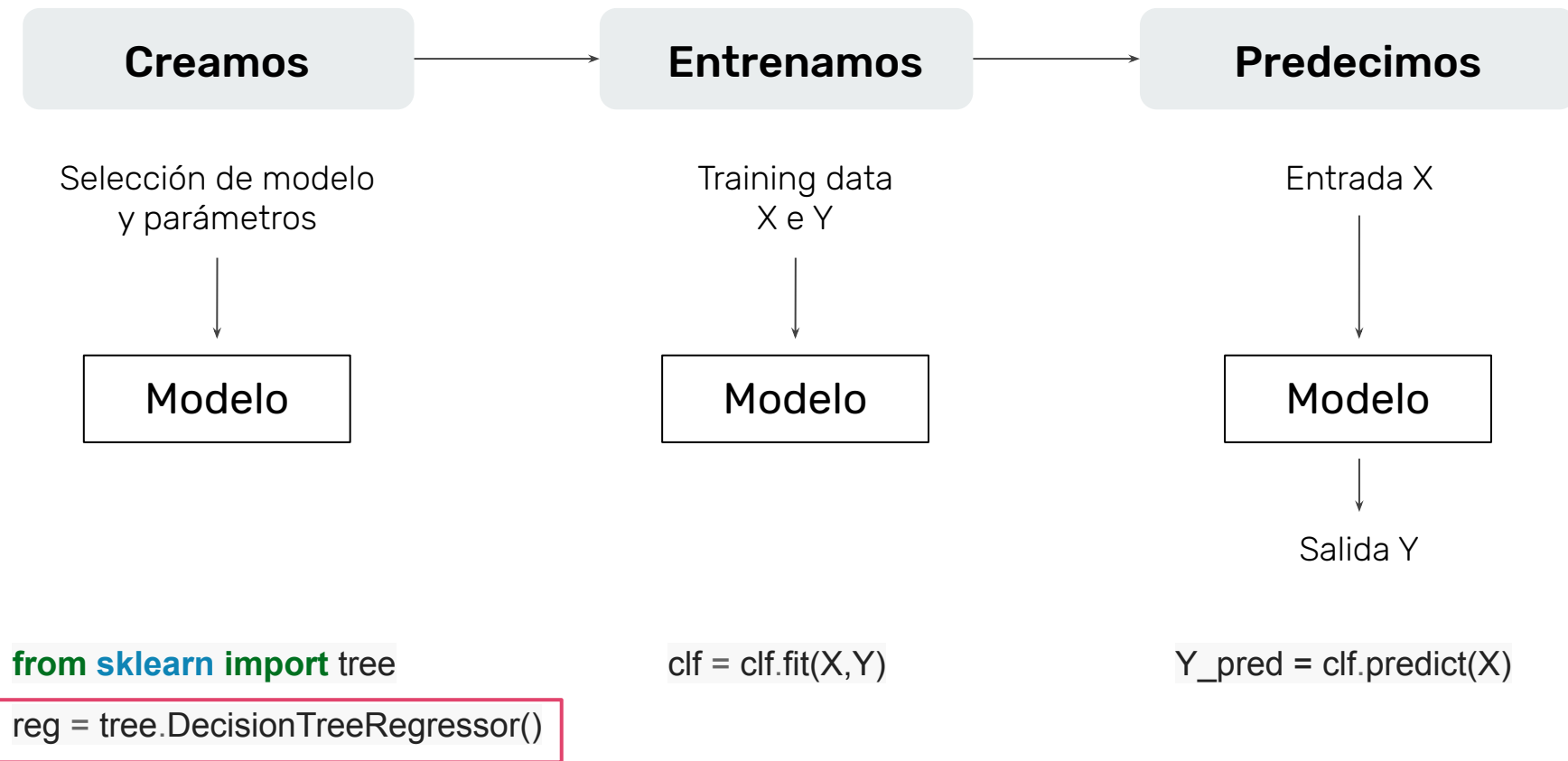


# Flujo de trabajo Scikit Learn





# Flujo de trabajo **Scikit Learn**



# Regresión



Machine Learning



Aprendizaje Supervisado

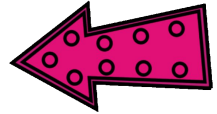


## Regresión



## Modelos

- **Regresión Lineal**
- **Árbol de Decisión**
- **k-nearest neighbors**
- Support Vector Machines
- Random Forest
- Perceptrón
- etc...



# Problema de regresión

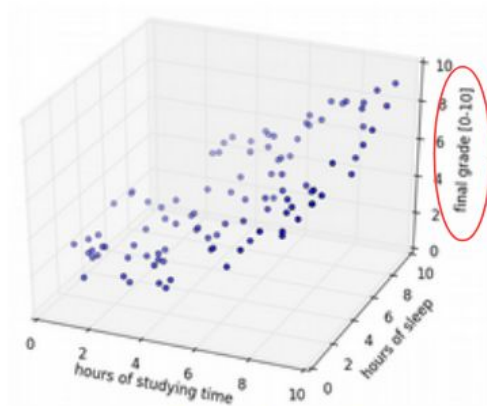
La variable dependiente  $Y$  es numérica.

Ejemplos:

- Precio de una propiedad.
- Precios en general
- Edad de una persona dado los consumos en su tarjeta de crédito
- Nota en un examen

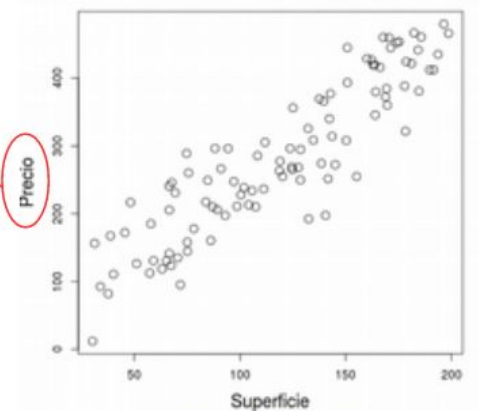
# Problema de regresión

Nota en un examen



(2 atributos)

Precio de una propiedad



(1 atributo)

Variable  
dependiente

# Problema de regresión

Consiste en predecir una respuesta numérica  $Y$  en base a atributos  $X_1, X_2, \dots, X_p$ .

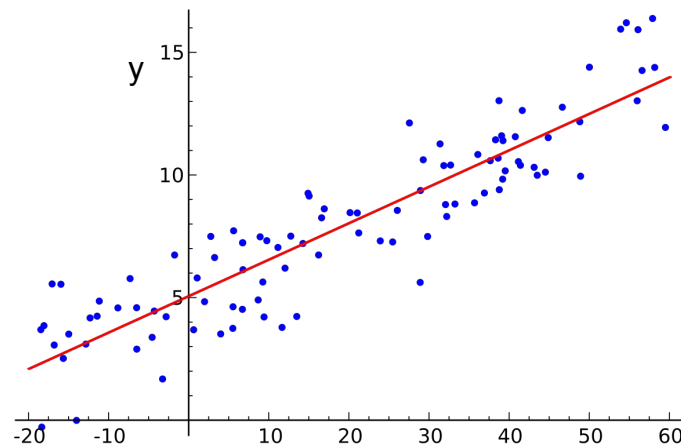
$$Y \approx f(X_1, X_2, \dots, X_p)$$

# Problema de regresión

Consiste en predecir una respuesta numérica  $Y$  en base a atributos  $X_1, X_2, \dots, X_p$ .

$$Y \approx f(X_1, X_2, \dots, X_p)$$

El caso más sencillo  
es una regresión lineal.



# Problema de regresión

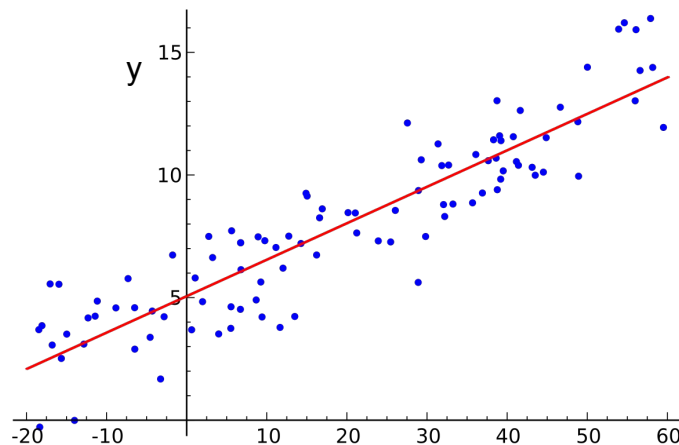
Consiste en predecir una respuesta numérica  $Y$  en base a atributos  $X_1, X_2, \dots, X_p$ .

$$Y \approx f(X_1, X_2, \dots, X_p)$$

El caso más sencillo es una regresión lineal.

Buscamos  $Y = mX + b$  que mejor ajuste a los datos:

- **m: pendiente**
- **b: ordenada al origen**



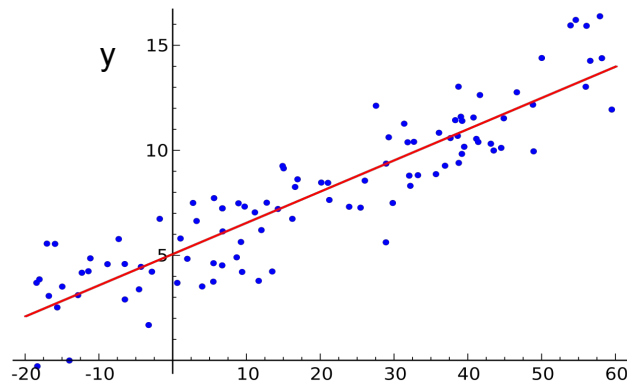


# Regresión **lineal**

**Un atributo:  $x_1$**

$$Y = m_1x_1 + b$$

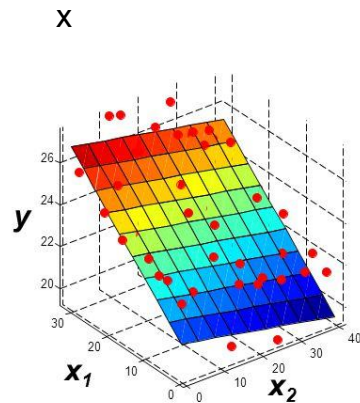
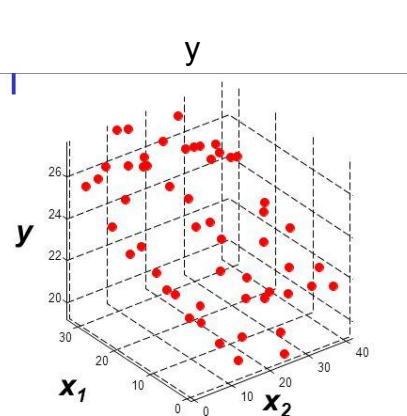
---



**Dos atributos:  $x_1, x_2$**

$$Y = m_1x_1 + m_2x_2 + b$$

---



**Tres atributos:  $x_1, x_2, x_3$**

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + b$$

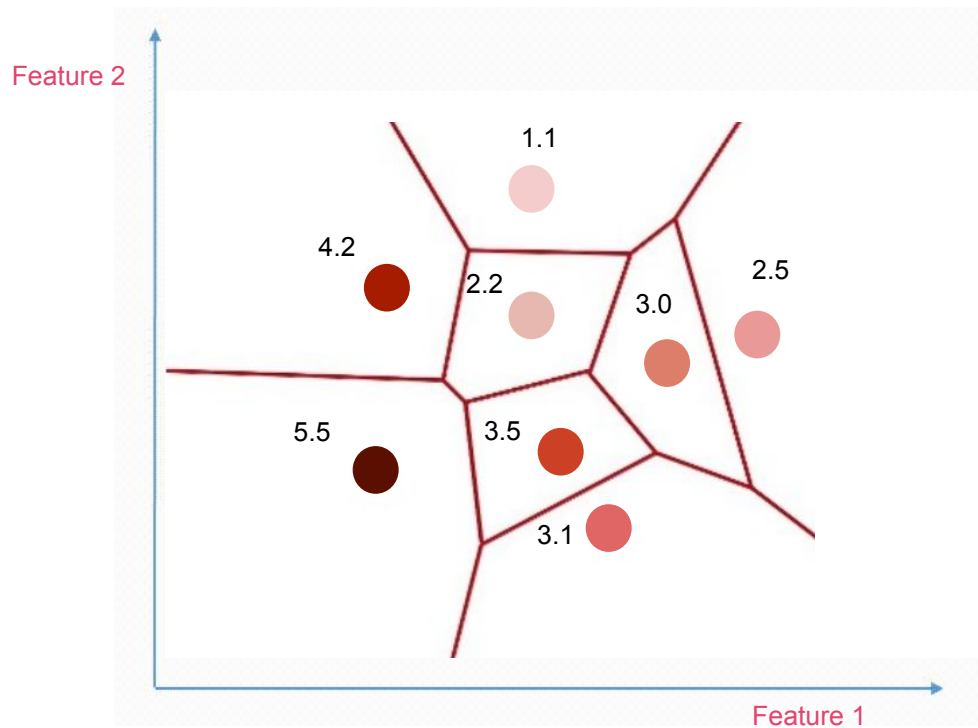
---

**$p$  atributos:  $x_1, x_2, x_3, \dots, x_p$**

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_px_p + b$$

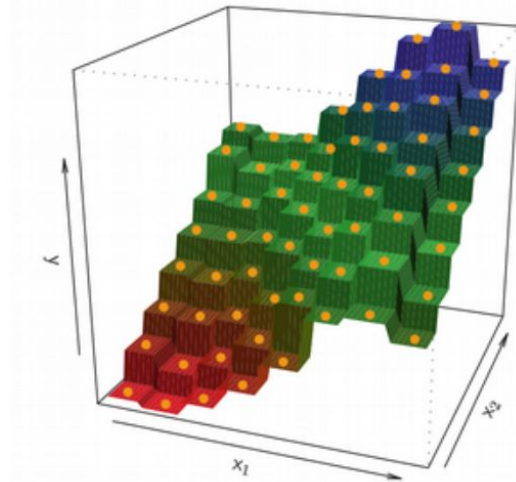
# Regresión con KNN

Dada una nueva instancia, devolver el promedio (ponderado) de los valores de sus vecinos.

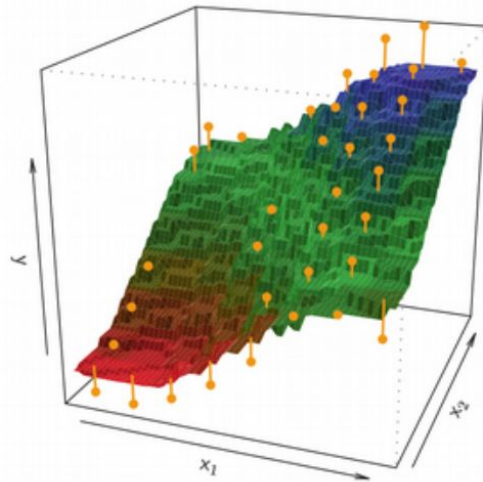


# Regresión con KNN

Dada una nueva instancia, devolver el promedio (ponderado) de los valores de sus vecinos.



K = 1



K = 9

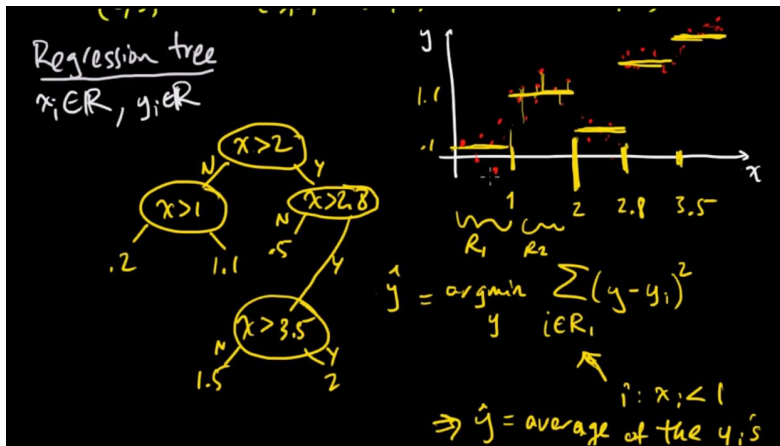
# Regresión con Árboles

## Construcción

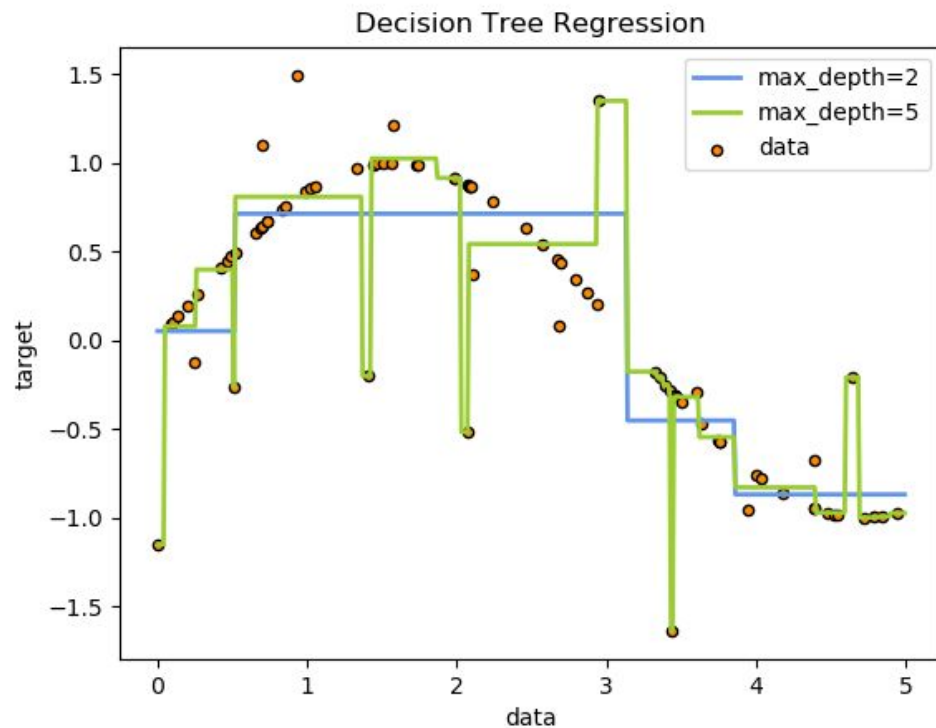
En cada nodo, usar reducción de desvío estándar de Y en lugar de gini/info gain.

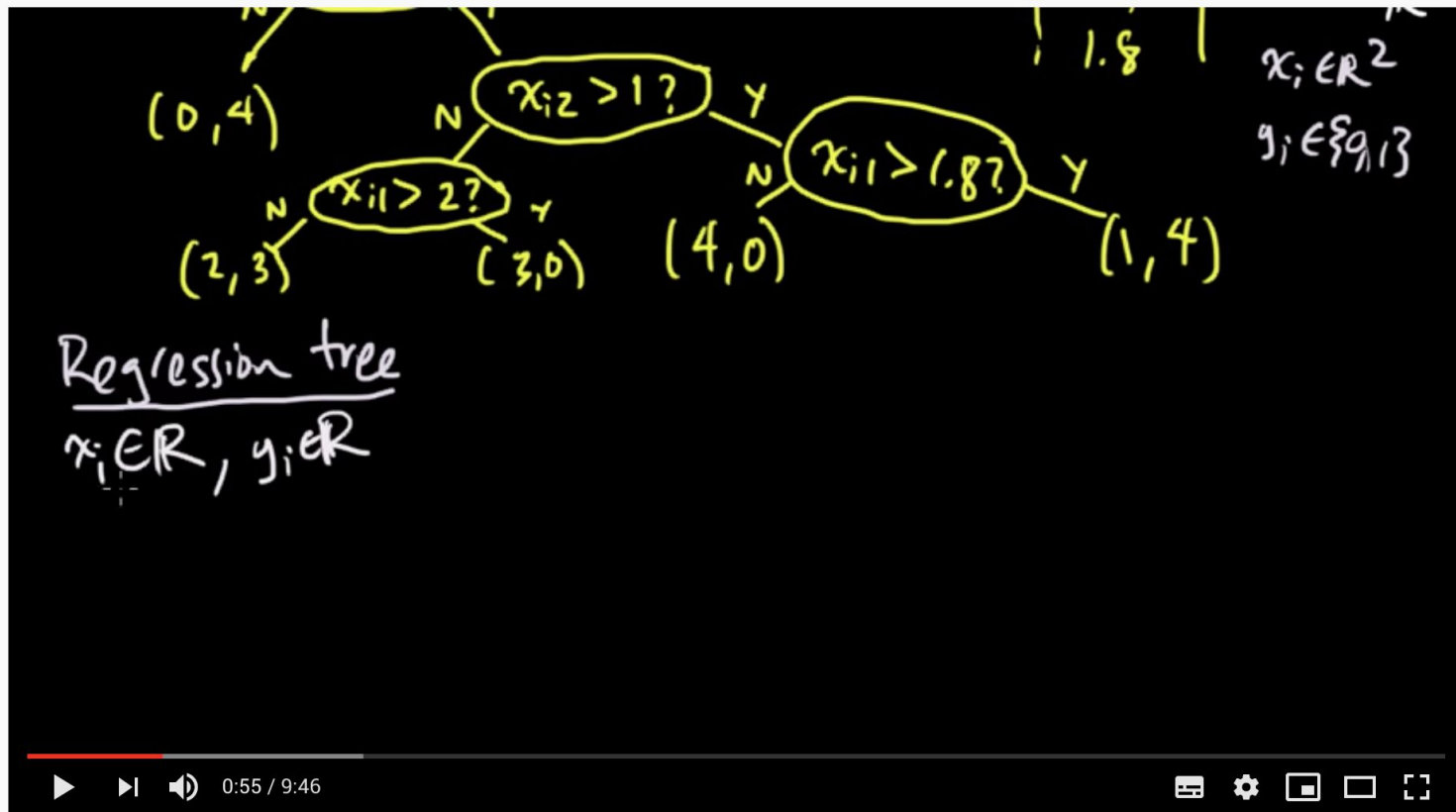
## Consulta

Al llegar a una hoja, devolver el promedio de Y sobre las instancias de la hoja.



# Regresión con Árboles



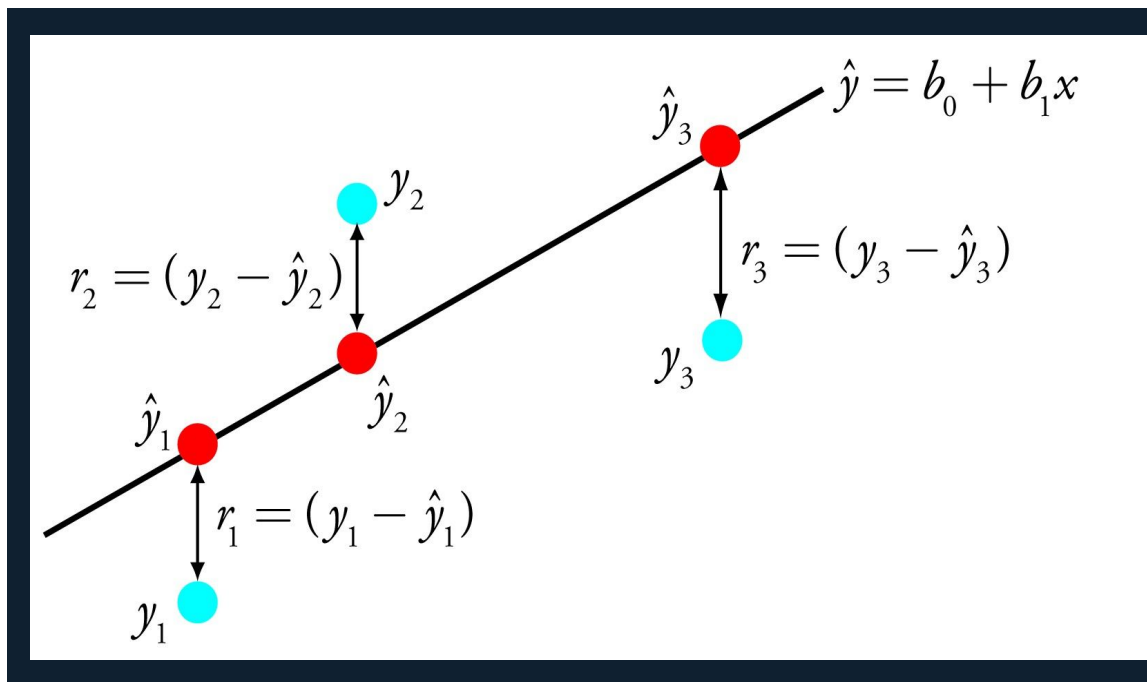


# Métricas de Evaluación para Regresión



# Errores

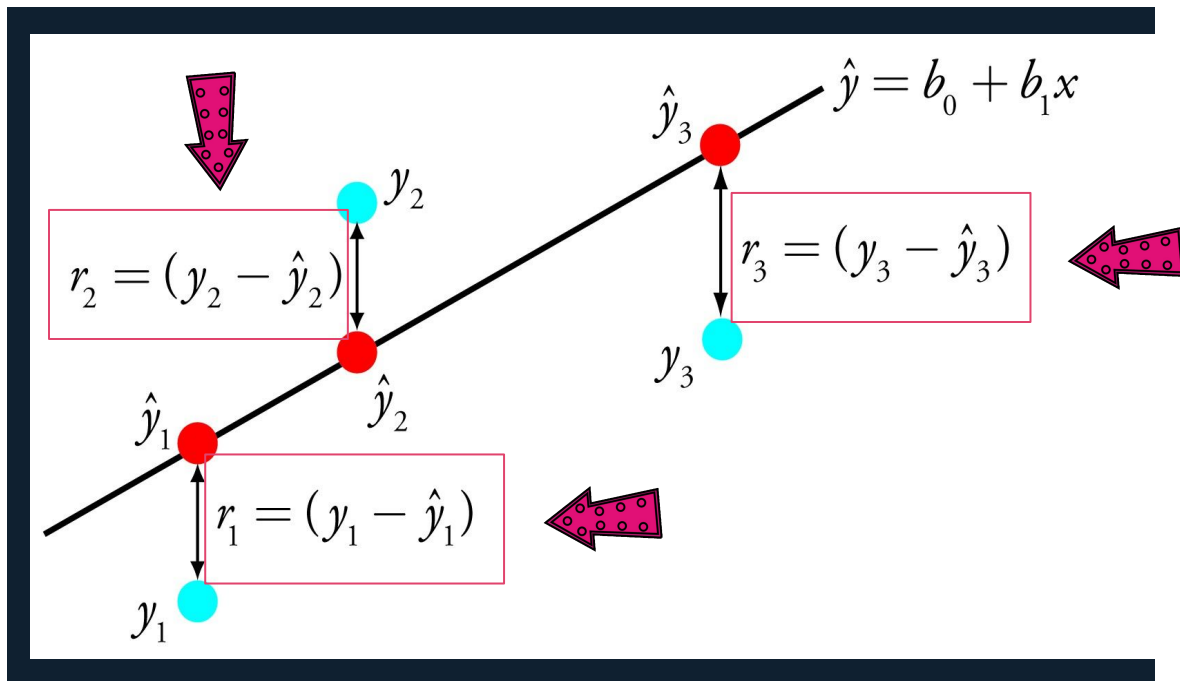
¿Pueden detectar los **errores** en la siguiente regresión?





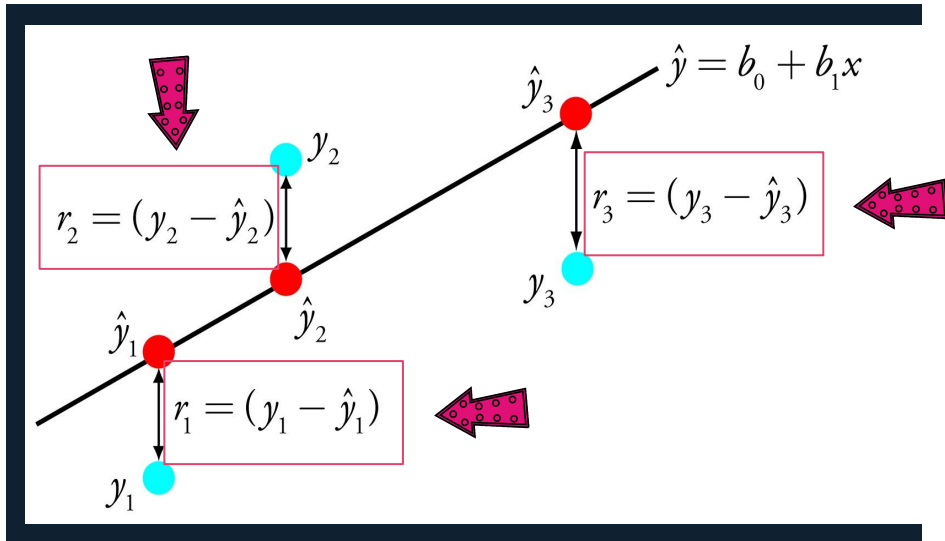
# Errores

¿Pueden detectar los **errores** en la siguiente regresión?



# Errores

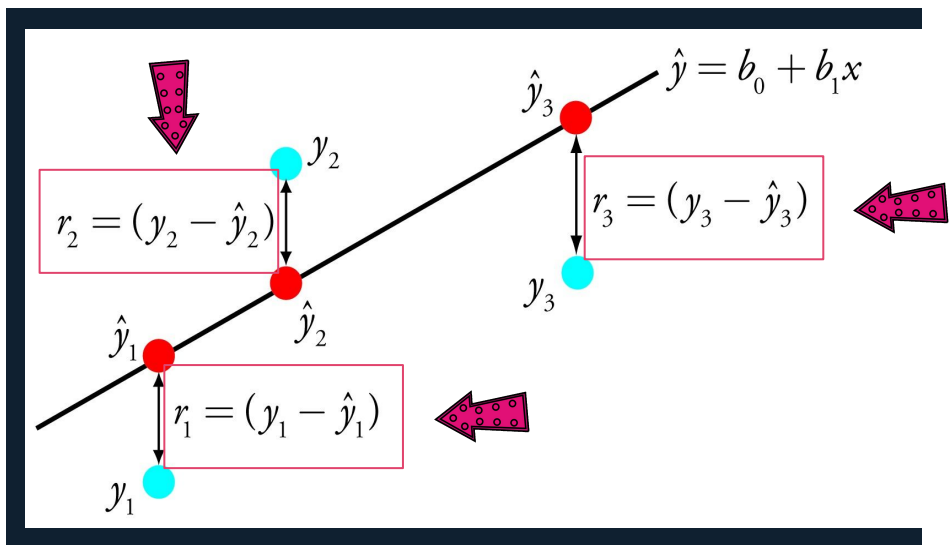
Los errores tienen distinto signo.  
Si sumamos sin considerar eso,  
podría suceder que se cancelen.



# Errores

Los errores tienen distinto signo.  $\longrightarrow$   
Si sumamos sin considerar eso,  
podría suceder que se cancelen.

**Solución:** podemos sumar  
sus valores **absolutos**:



$$\sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i|$$

# Errores

**Solución:** podemos sumar sus valores **absolutos**:

$$\left| \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i| \right|$$

*n es el número de muestras*

# Errores

**Solución:** podemos sumar sus valores **absolutos**:

$$\left| \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i| \right|$$

*n es el número de muestras*

¡Pero ahora el error va a ser mayor si hay más muestras!

$$\left| \text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i| \right|$$

*MAE: Mean Absolute Error*

## Errores

¿Y si usamos otros valores  
en lugar de los absolutos?

# Errores

**Solución:** podemos sumar sus valores **cuadrados:**



# Errores

**Solución:** podemos sumar sus valores **cuadrados**:

$$\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2$$

*n es el número de muestras*



# Errores

**Solución:** podemos sumar sus valores **cuadrados**:

$$\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2$$

*n es el número de muestras*

¡Pero ahora el error va a ser mayor si hay más muestras!

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2$$

*MSE: Mean Squared Error*

# En Scikit-Learn



<https://scikit-learn.org/stable/modules/classes.html#sklearn-metrics-metrics>



Para pensar o investigar: **FUNCIONES DE COSTO**

# Operativamente, ¿qué cambia y qué no?

## Cambia

Métrica de evaluación

## NO cambia

- Hipótesis de aprendizaje inductivo
- “Entrenar un modelo” consiste en ajustar sus parámetros para un dado dado un conjunto de datos.
- Overfitting/underfitting
  - Recursos para trabajar con el overfitting:
    - Datos de train/test
    - Curva de complejidad
- Queremos elegir el modelo más simple, que sea razonablemente bueno.

# Hands-on training





DS\_Clase\_19\_Regresión.ipynb

A close-up photograph of a white ceramic cup filled with a latte. The coffee has a thick layer of white foam with intricate brown latte art designs, including a central heart shape and radiating lines. The cup sits on a matching white saucer. In the background, a white napkin with a silver fork is visible, all set against a dark, textured surface.

**¡BREAK!**

---

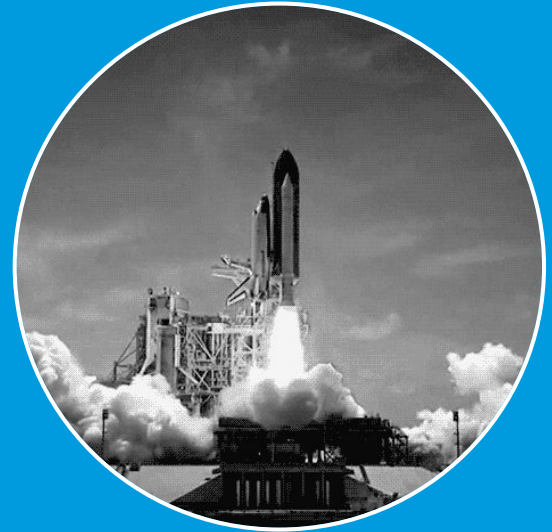
Ph. Credit: Drew Coffmann



# Proyecto 2:

## Lanzamiento

### Entrega 3



# Proyecto 2: Modelado

---



## Entrega 3: Regresión

Realizá tus primeras predicciones  
utilizando técnicas de regresión

● Beginner

by



Francisco Dorr



# Actividad:

## Dudas comunitarias



# Para la próxima

---

1. Ver los videos de la plataforma “Validación y testeo de Modelos”.
2. Trabajar en la Entrega 03.
3. Completar Notebooks atrasados.

ACÀMICA