

ACÀMICA

ESTADO DEL PROYECTO 1

Mi Primer Modelo de Machine Learning

Identifica el estado de avance de tu proyecto. Te será muy útil saberlo para poder esquematizar tu trabajo en las próximas semanas y entregar a tiempo.

Puedes estimar tu grado de avance de la siguiente manera:

Grupo 25%: Recién estoy comenzando

Grupo 50%: He avanzado pero aún me falta

Grupo 75%: Estoy muy avanzado

Grupo 100%: He finalizado el proyecto



¿En qué grupo te ubicas?

Grupo 25% Recién estoy comenzando	?
Grupo 50% He avanzado pero aún me falta	?
Grupo 75% Estoy muy avanzado	?
Grupo 100% He finalizado el proyecto	?



TEMA DEL DÍA

Regresión

Éste es el último tema que veremos antes del cierre del Sprint: aplicaremos las herramientas de Aprendizaje Supervisado a los problemas de Regresión.



Agenda

Daily

Explicación: Regresión Lineal, KNN y Árboles. Métricas.

Break.

Hands-on training

Cierre.



Daily



Daily



Sincronizando...

Toolbox



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Challenge

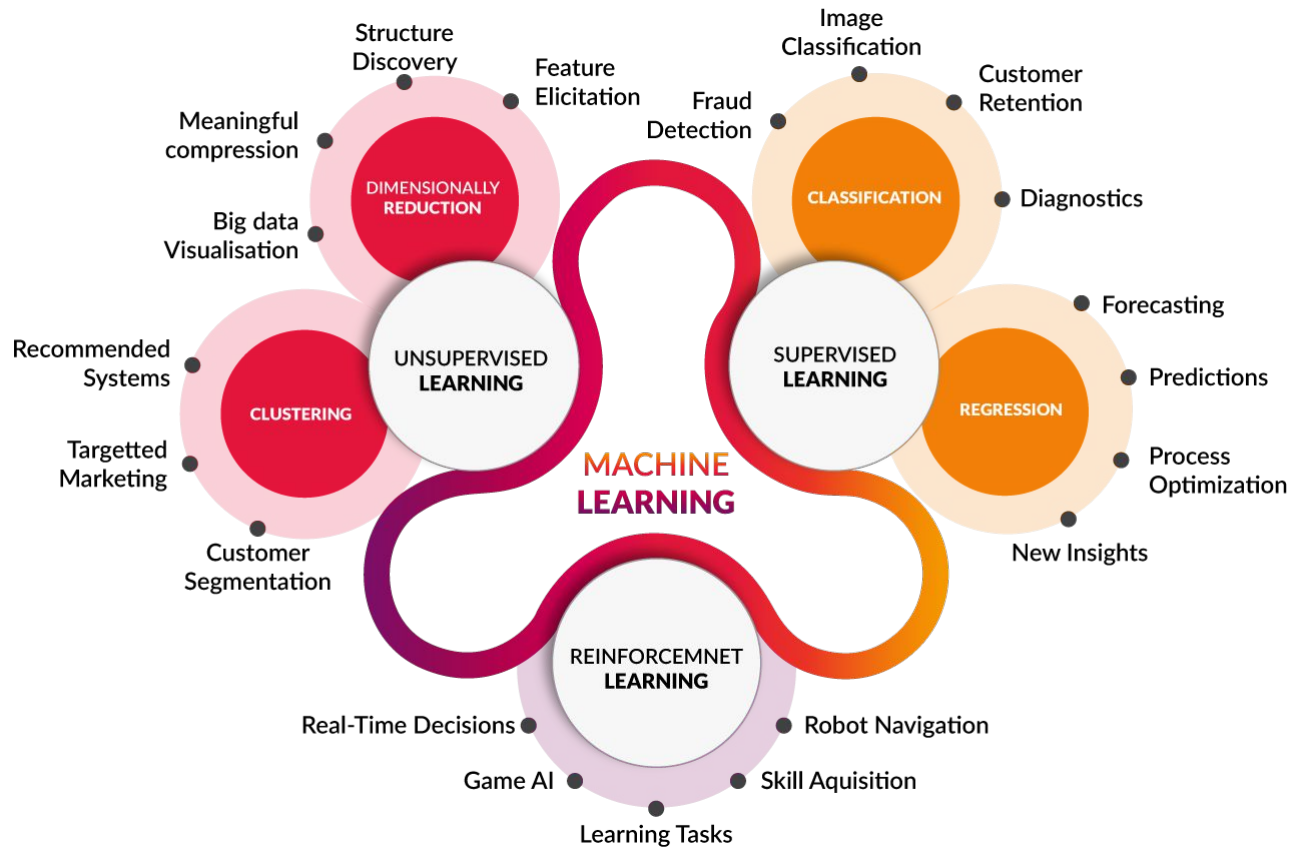


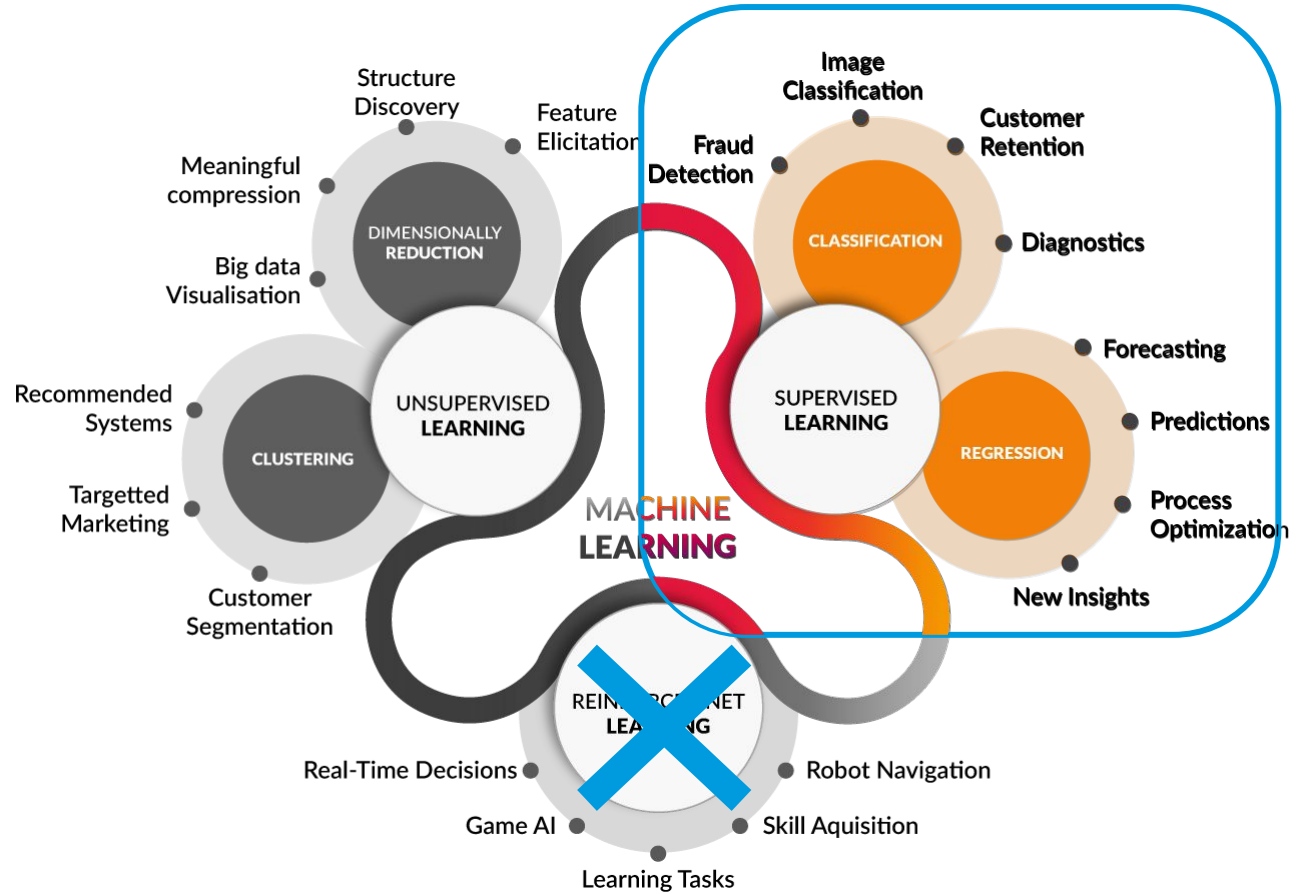
¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Regresión



Mapa





¡Vamos a ver que es
todo muy parecido!

Problema de regresión

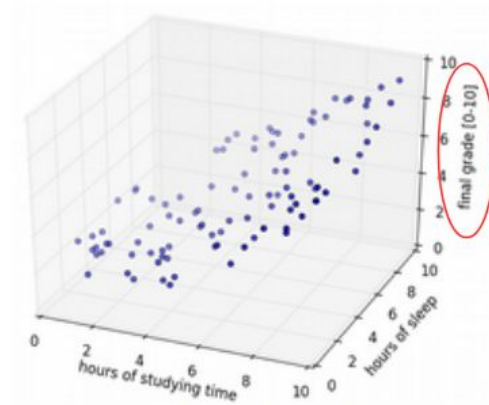
La variable dependiente Y es numérica.

Ejemplos:

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

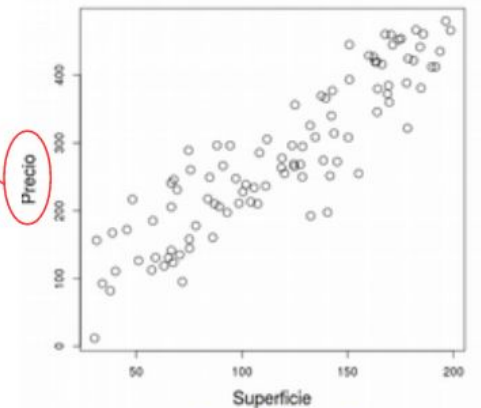
Problema de regresión

Nota en un examen



(2 atributos)

Precio de una propiedad



(1 atributo)

Variable
dependiente

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



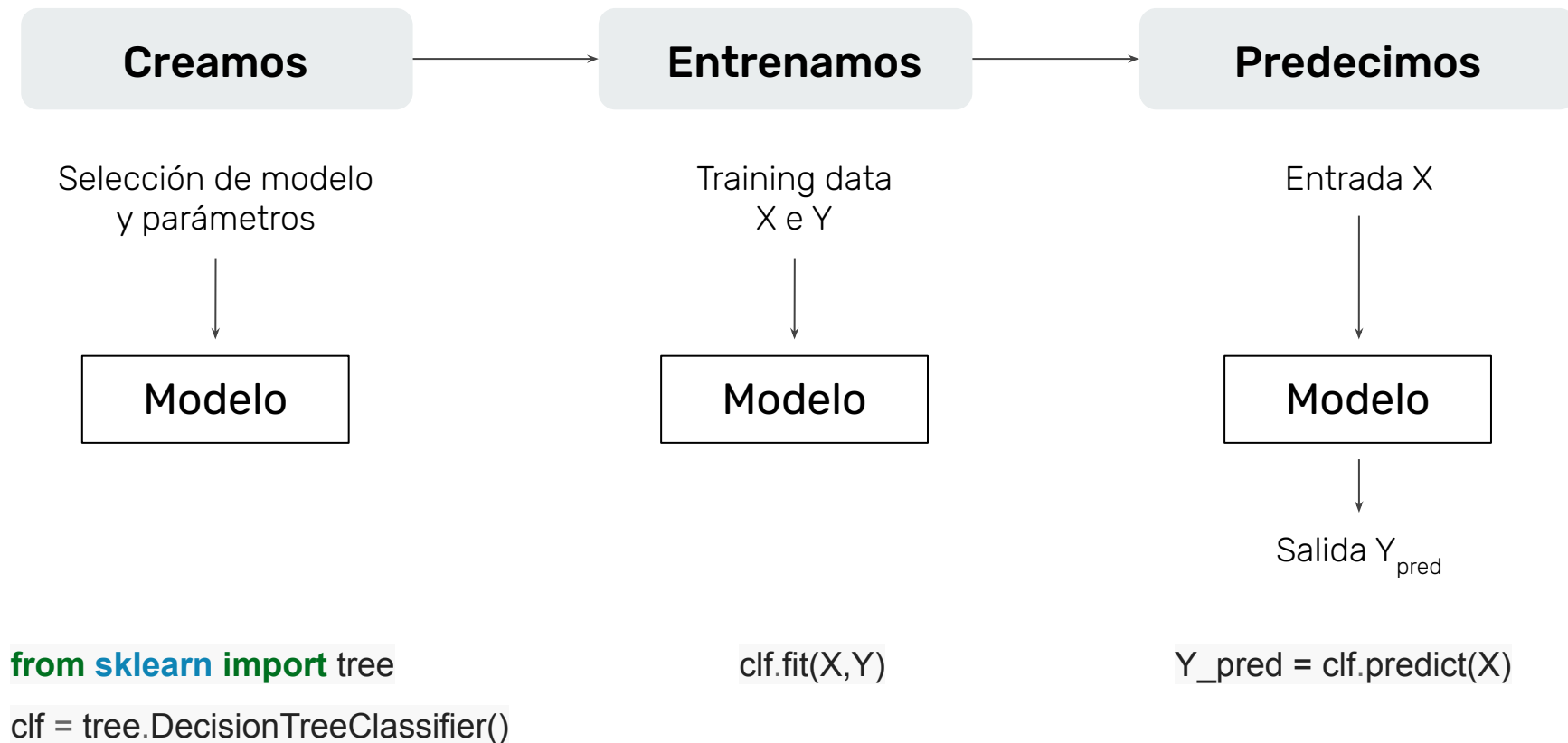
Regresión



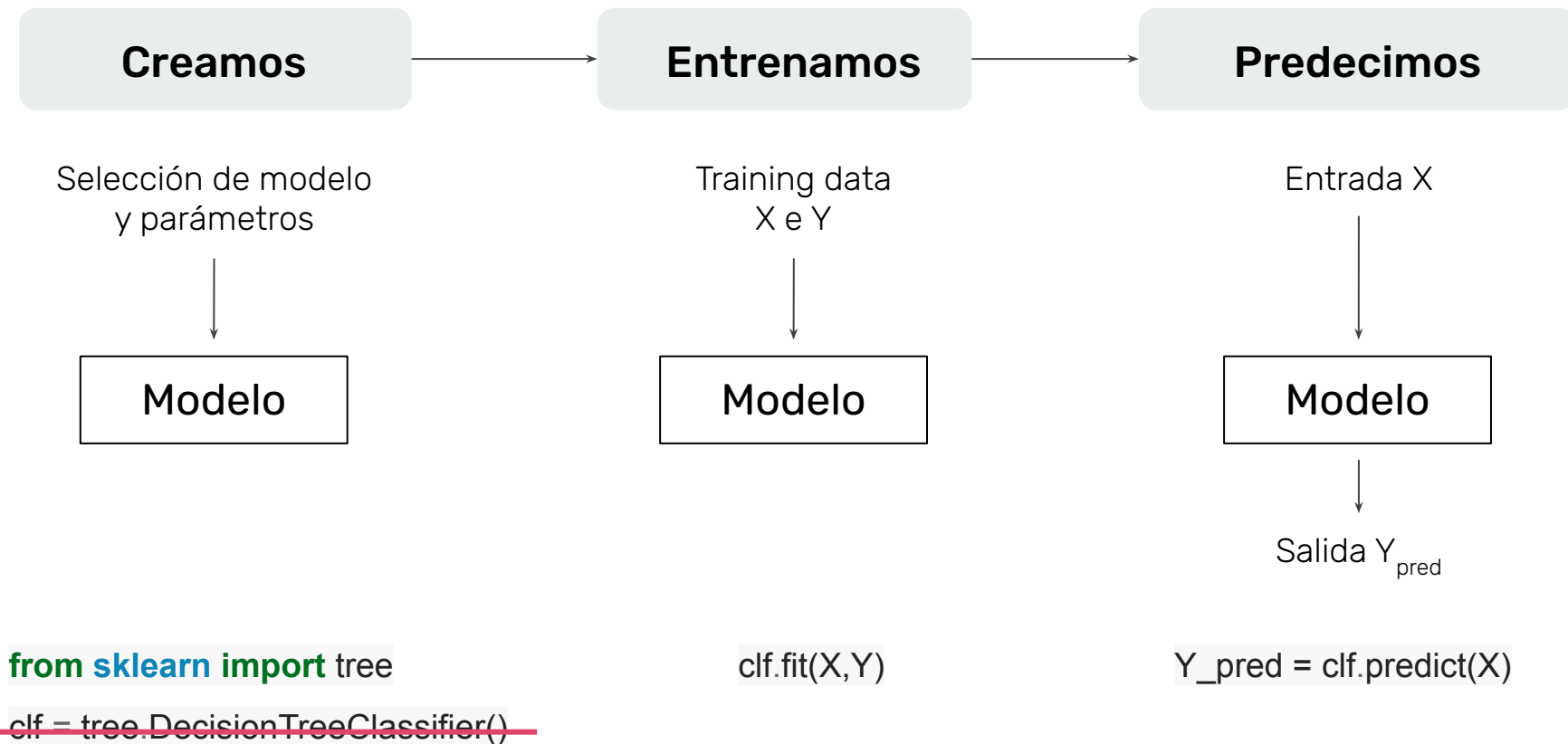
Modelos

- Regresión Lineal
- Árbol de Decisión
- k-nearest neighbors
- etc...

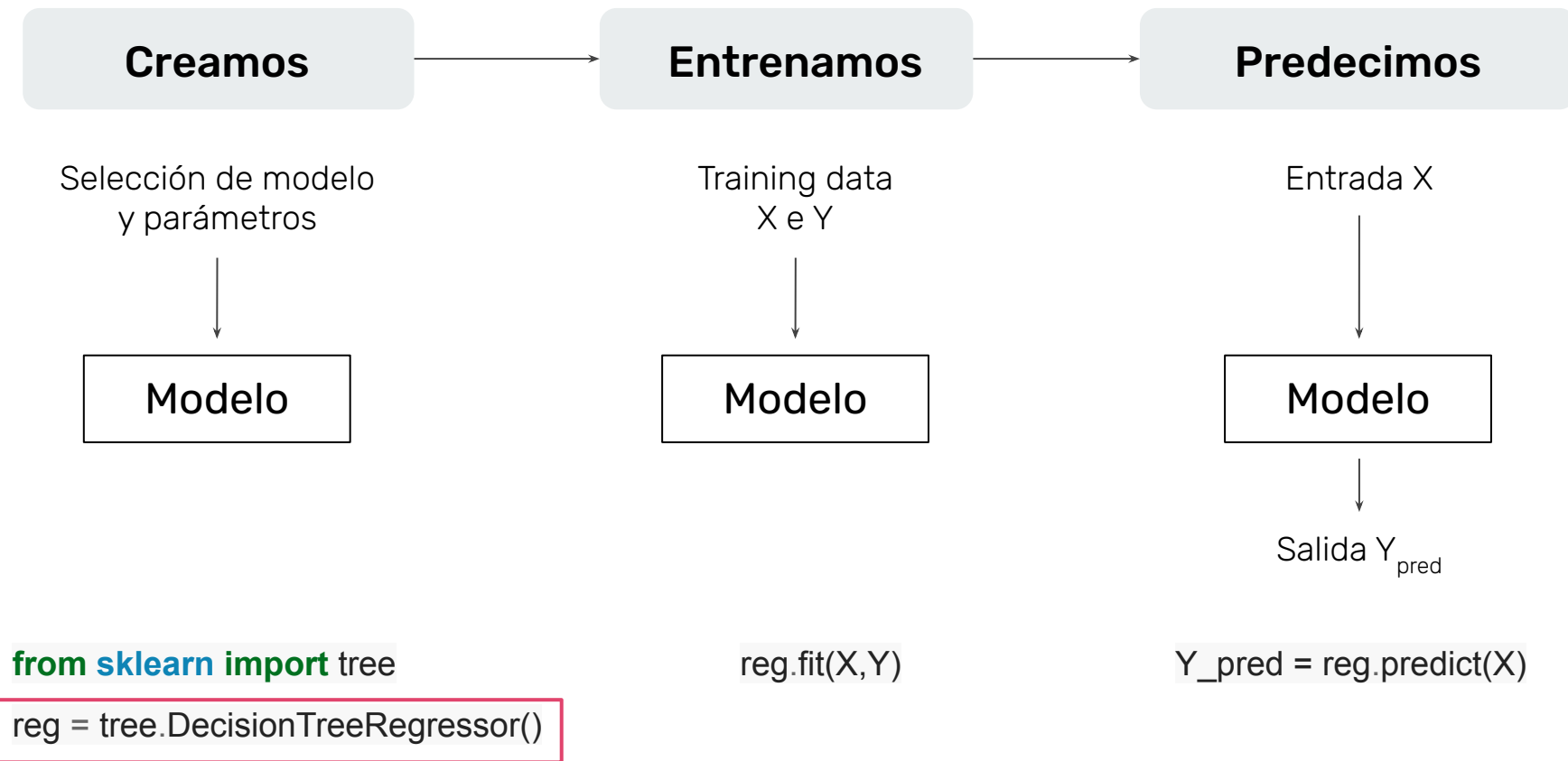
Flujo de trabajo **Scikit Learn**



Flujo de trabajo **Scikit Learn**



Flujo de trabajo **Scikit Learn**



Regresión



Machine Learning



Aprendizaje Supervisado

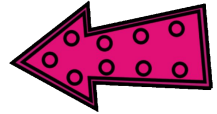


Regresión



Modelos

- **Regresión Lineal**
- **Árbol de Decisión**
- **k-nearest neighbors**
- etc...



Problema de regresión

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

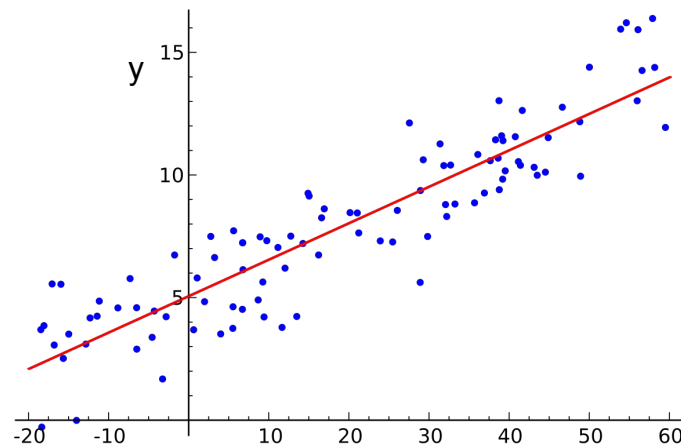
$$Y \approx f(X_1, X_2, \dots, X_p)$$

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

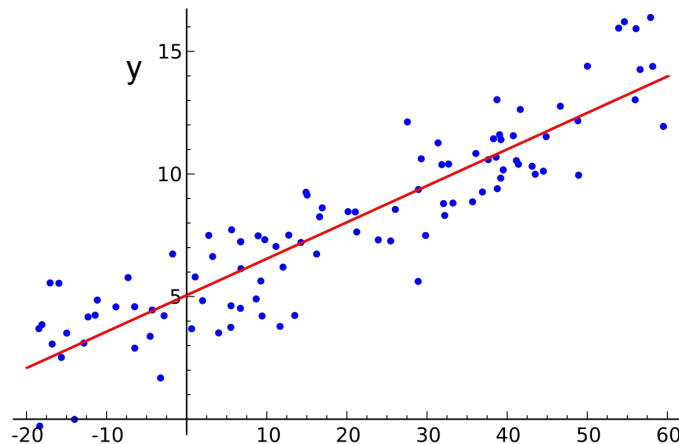
Consiste en predecir una respuesta numérica Y en base a atributos X_1, X_2, \dots, X_p .

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

El caso más sencillo es una regresión lineal.

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

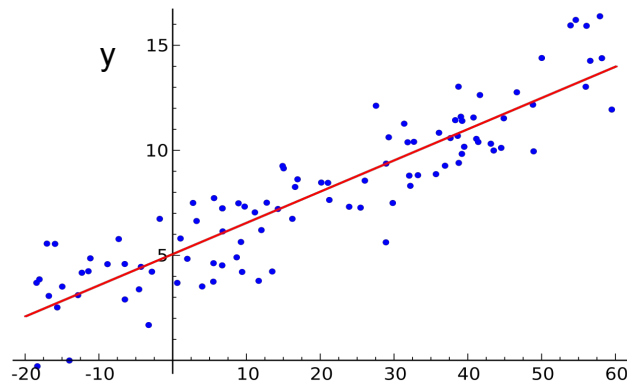
- **m: pendiente**
- **b: ordenada al origen**



Regresión **lineal**

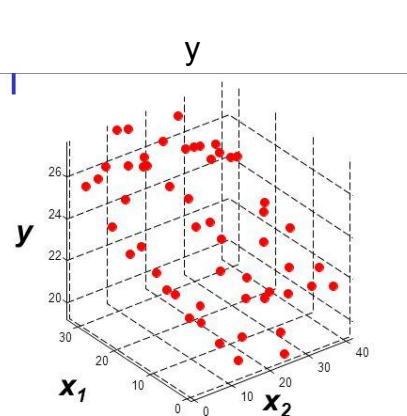
Un atributo: x_1

$$Y = m_1x_1 + b$$



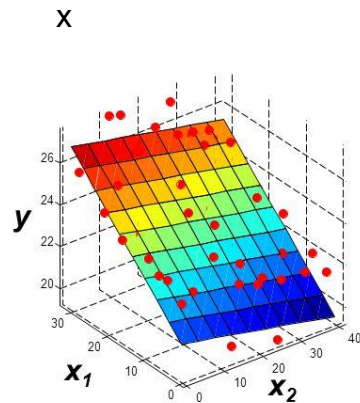
Dos atributos: x_1, x_2

$$Y = m_1x_1 + m_2x_2 + b$$



Tres atributos: x_1, x_2, x_3

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

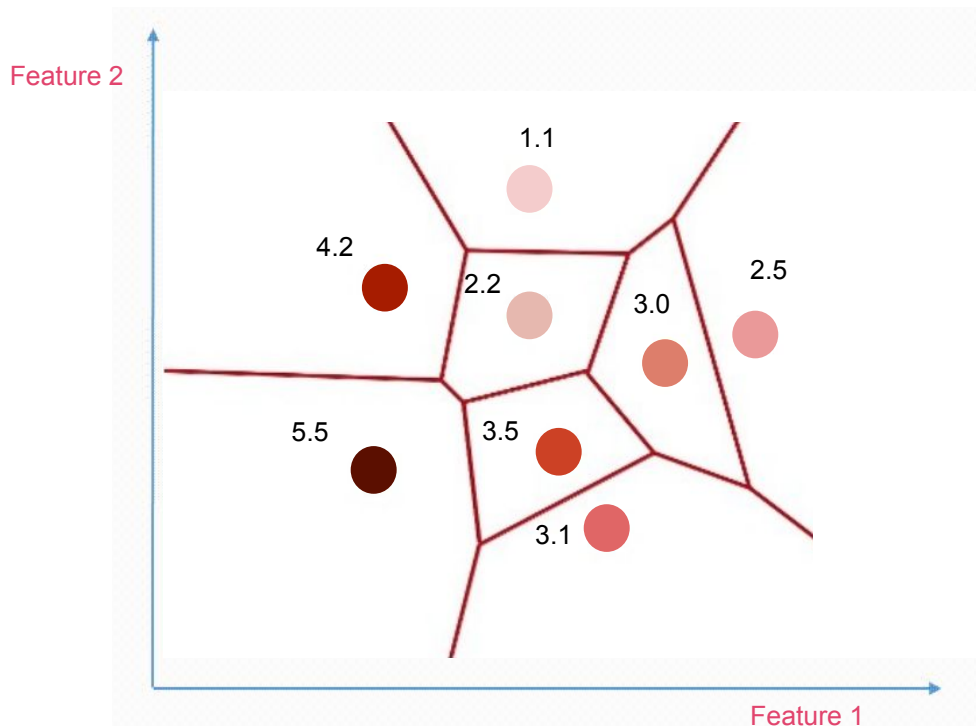


p atributos: $x_1, x_2, x_3, \dots, x_p$

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

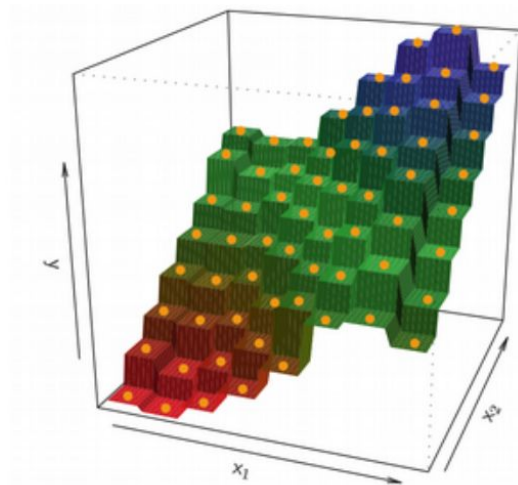
Regresión **con KNN**

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

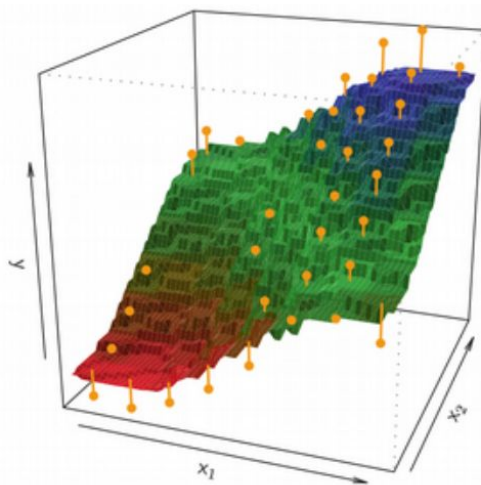


Regresión con KNN

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



K = 1



K = 9

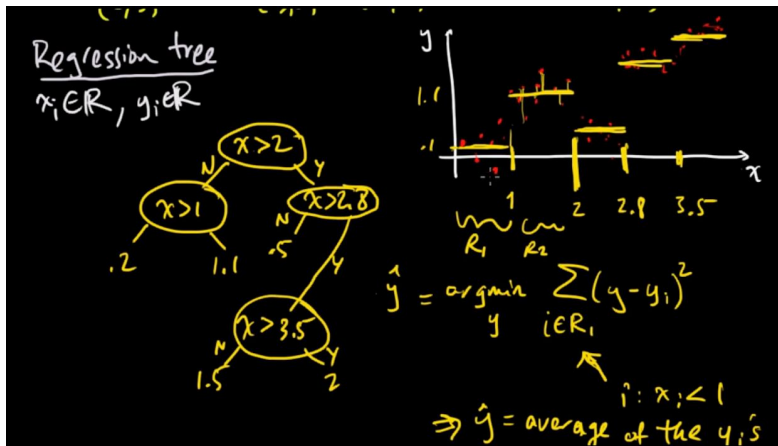
Regresión con Árboles

Construcción

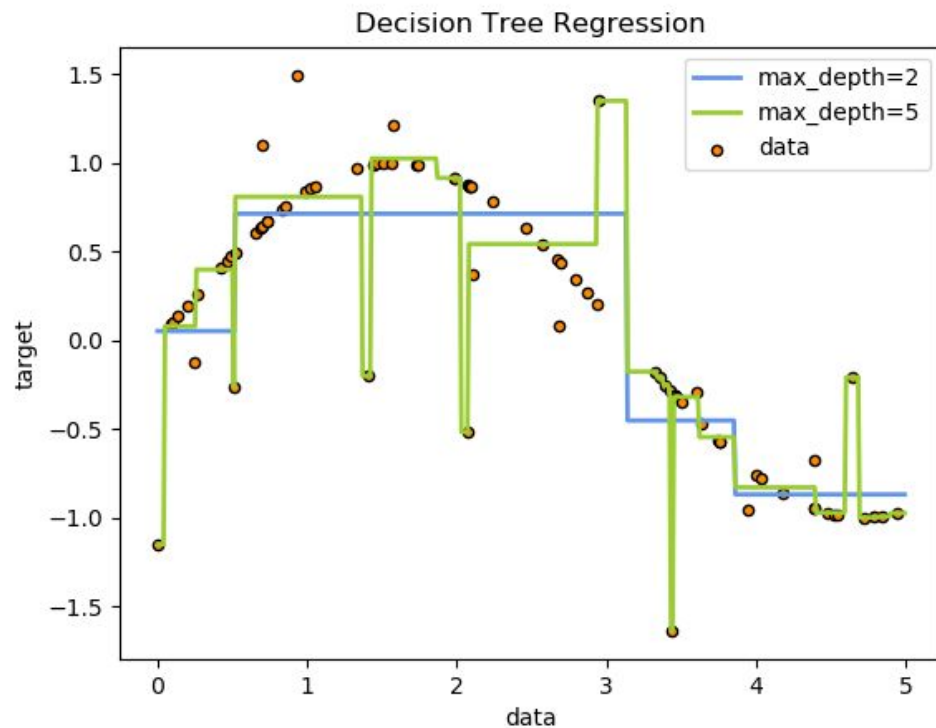
En cada nodo, usar reducción de desvío estándar de Y en lugar de gini/info gain.

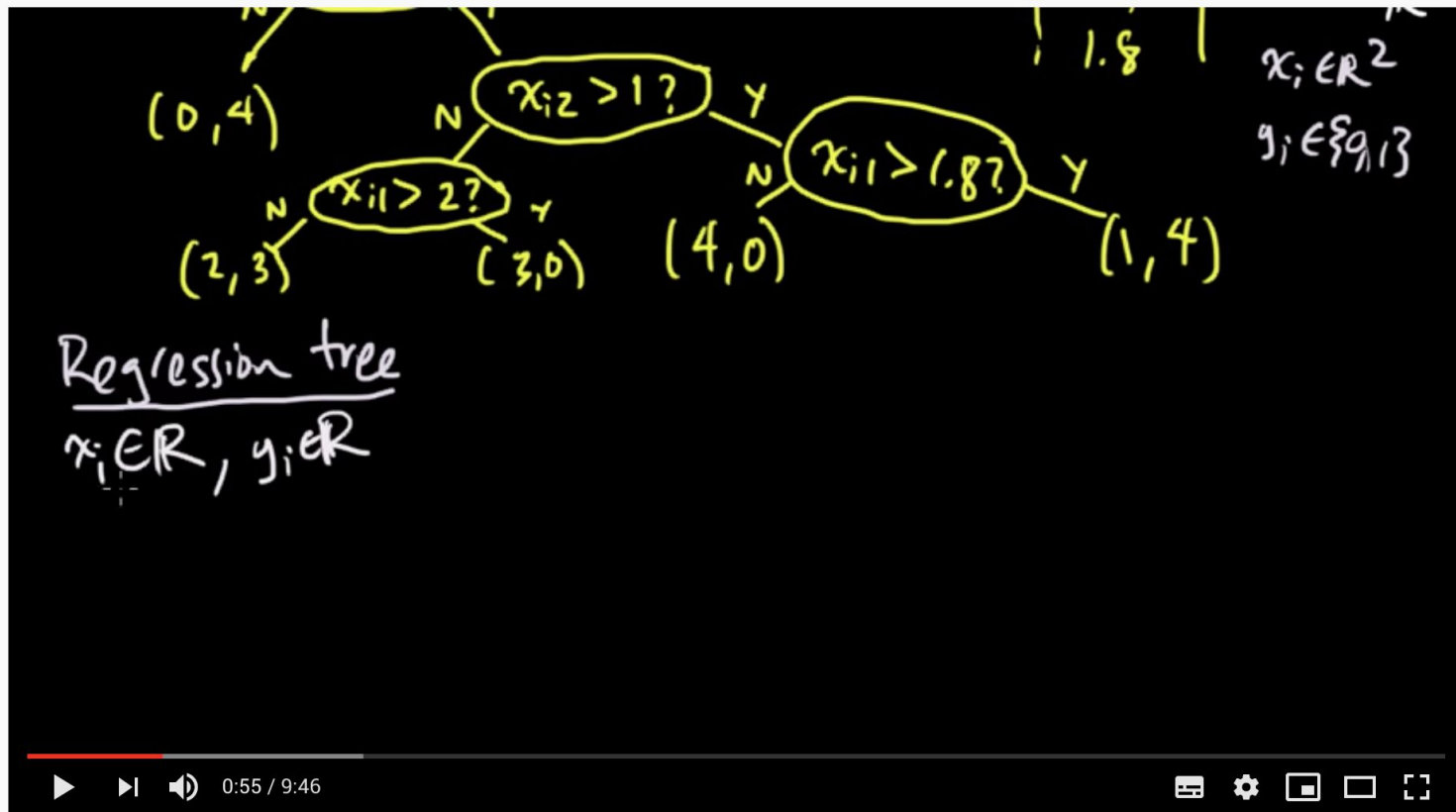
Consulta

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



Regresión con Árboles



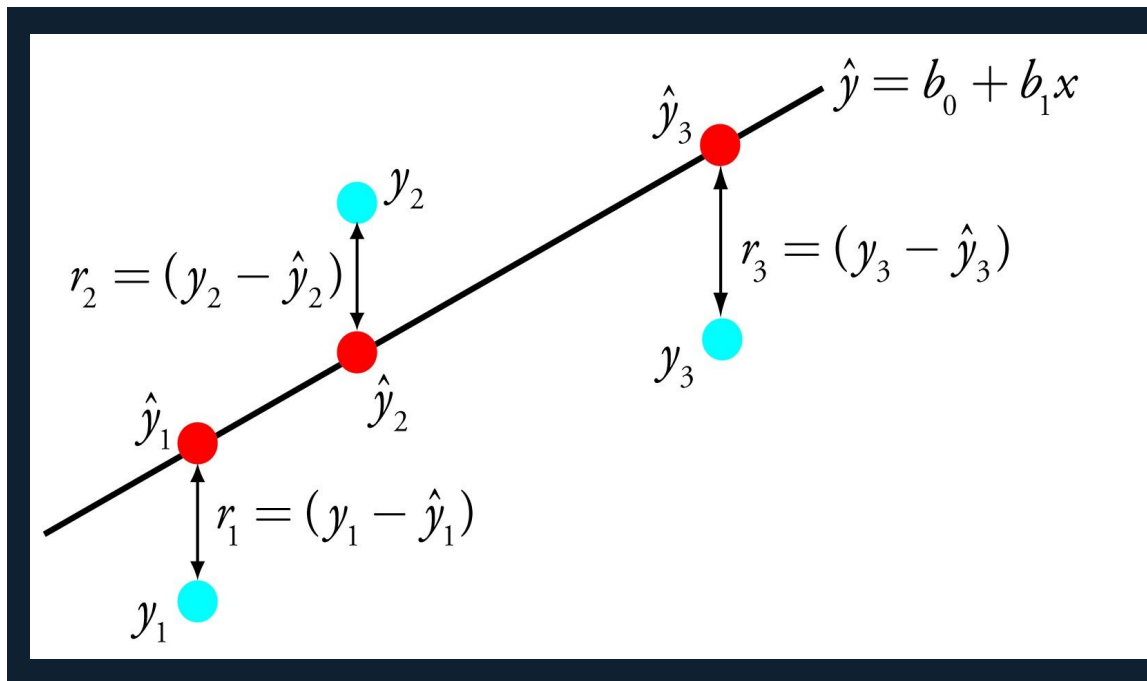


Métricas de Evaluación para Regresión



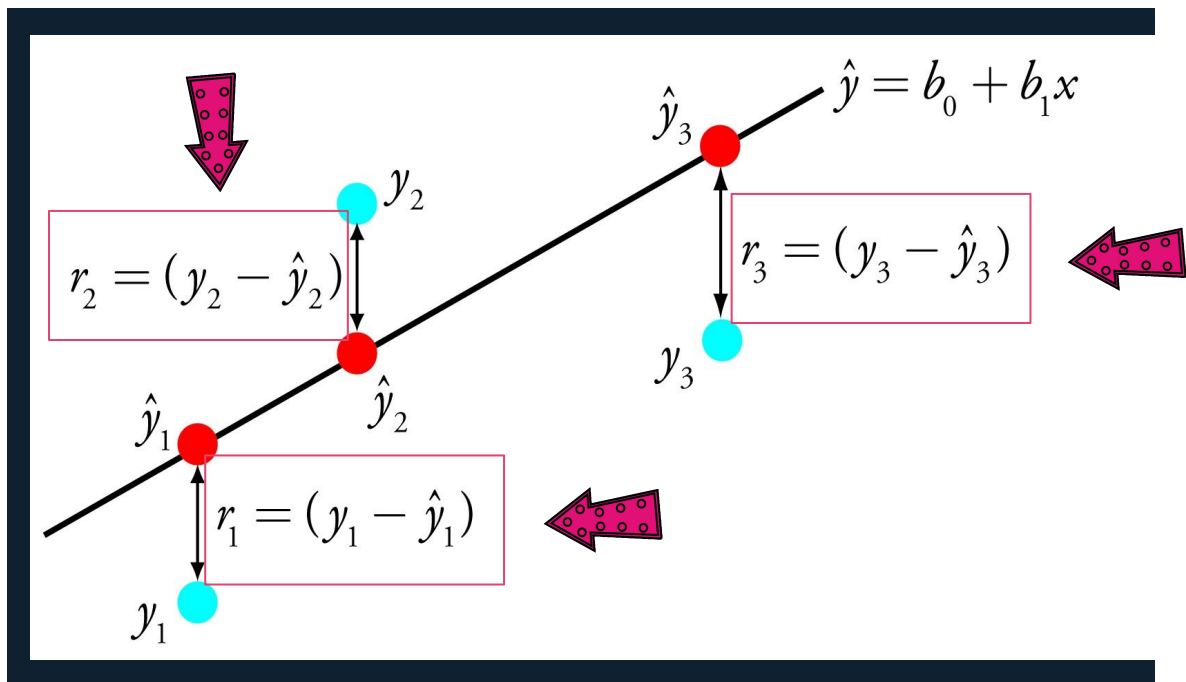
Errores

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



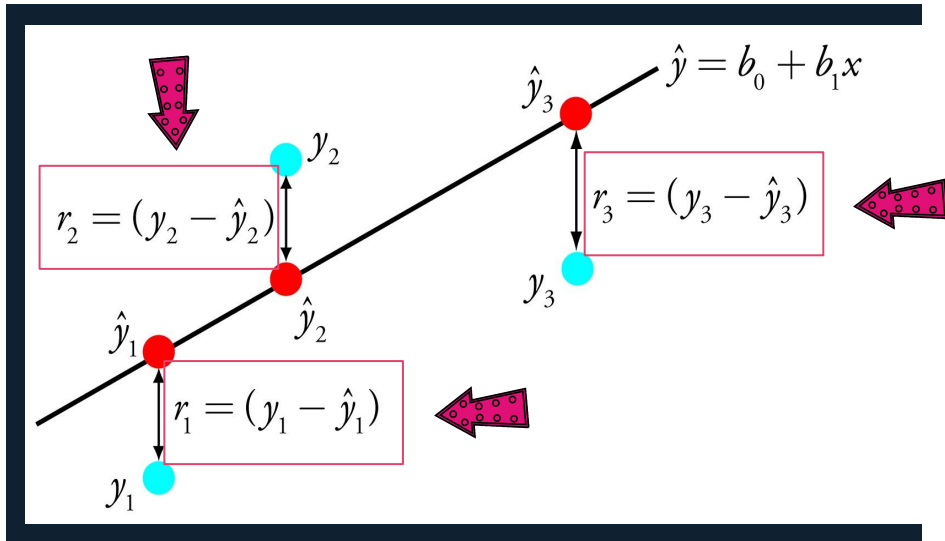
Errores

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



Errores

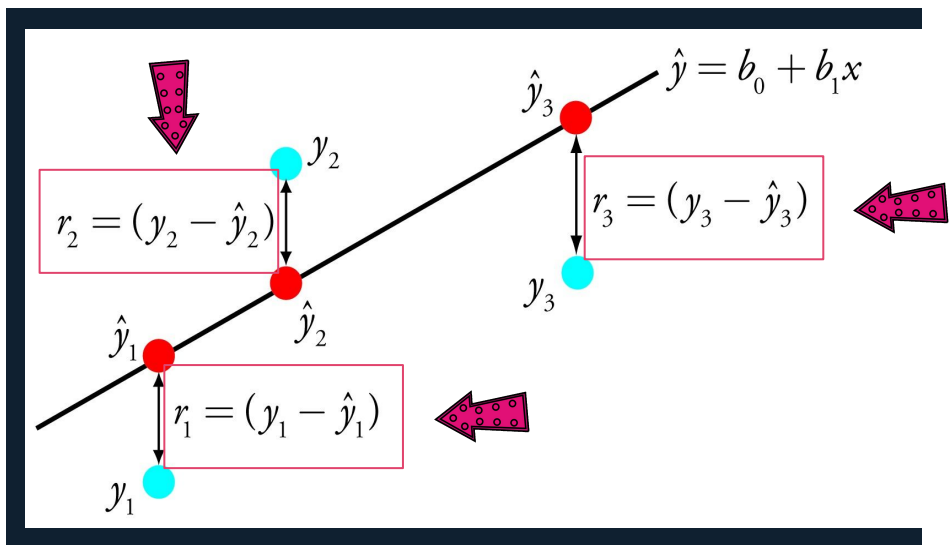
Los errores tienen distinto signo.
Si sumamos sin considerar eso,
podría suceder que se cancelen.



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!

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.

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver fork are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!



Hands-on training



¡PRIMER PROYECTO!

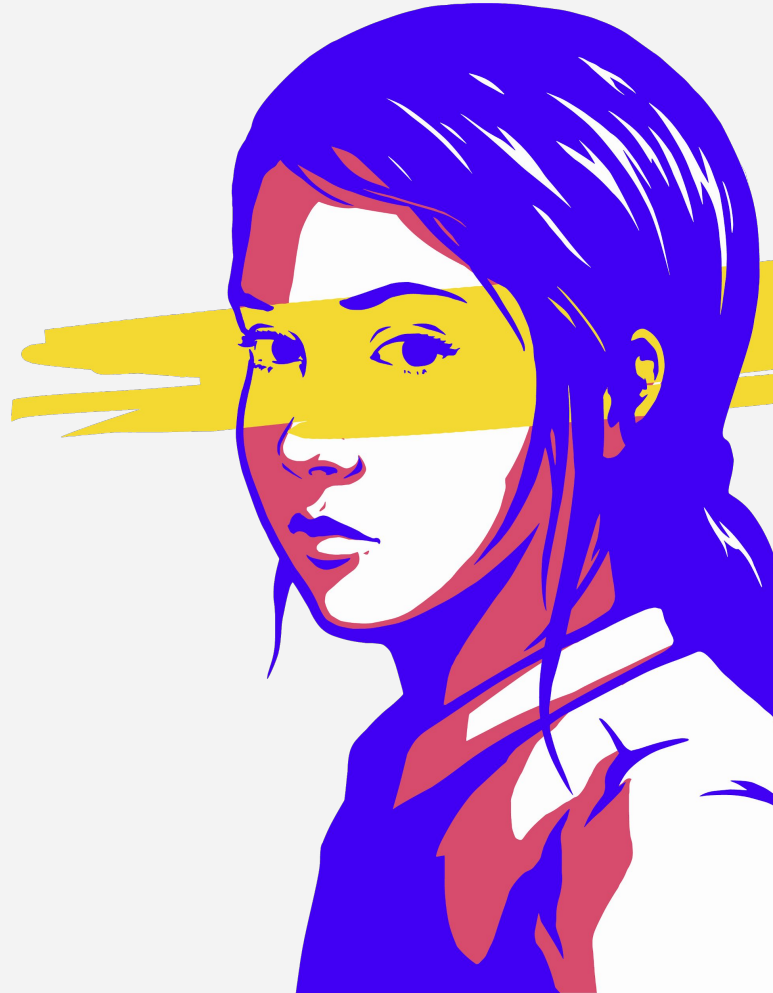
Si miraste el notebook de la Toolbox y no tienes dudas, empieza a aplicar lo aprendido en el primer proyecto.

Si aún no lo has visto, dedícale un tiempo al notebook.



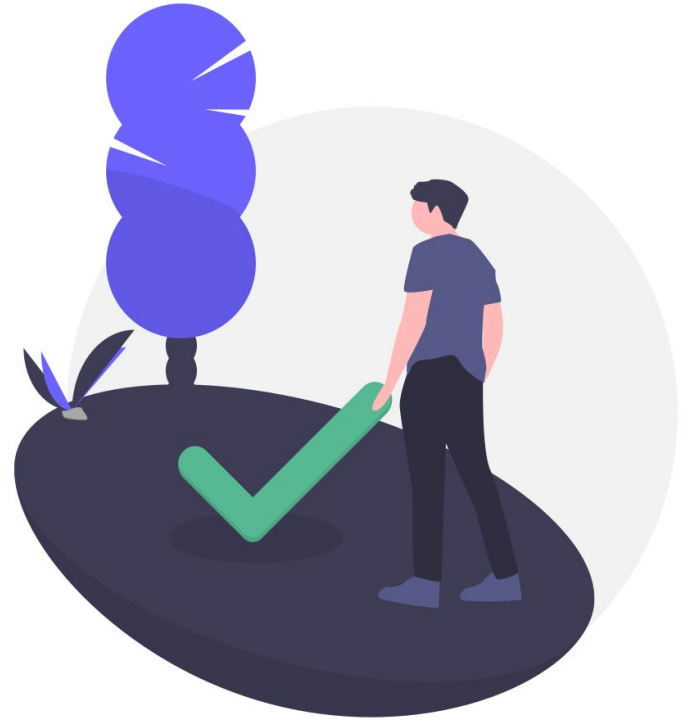
Tus deadlines

Cómo te certificarás



Durante el último encuentro del sprint, se te habilitará la opción de entrega de proyecto en la plataforma.

Una vez habilitada, quedará en ese estado durante todo el programa.



Tus deadlines

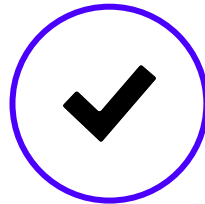
Al final de cada Sprint deberás entregar tu proyecto*. En un plazo máximo de una semana recibirás la devolución de tu evaluador/a:



*Entre un Sprint y otro contarás con una semana de “entretiempo”. Puedes usarla para terminar de ajustar tu proyecto y entregarlo, ayudar a tus compañeros/as a terminar su proyecto, o incluso descansar.

Tus deadlines

Al final de cada Sprint deberás entregar tu proyecto*. En un plazo máximo de una semana recibirás la devolución de tu evaluador/a:



Aprobado

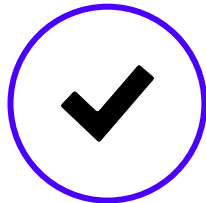
Cuando hayas cumplido satisfactoriamente con el 100% de los puntos del checklist.



*Entre un Sprint y otro contarás con una semana de “entretiempo”. Puedes usarla para terminar de ajustar tu proyecto y entregarlo, ayudar a tus compañeros/as a terminar su proyecto, o incluso descansar.

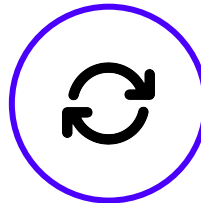
Tus deadlines

Al final de cada Sprint deberás entregar tu proyecto*. En un plazo máximo de una semana recibirás la devolución de tu evaluador/a:



Aprobado

Cuando hayas cumplido satisfactoriamente con el 100% de los puntos del checklist.



Para rehacer

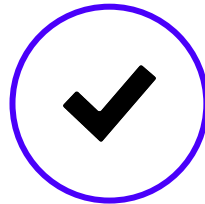
Cuando haya puntos del checklist a mejorar. ¡Tendrás una cantidad ilimitada de iteraciones!



*Entre un Sprint y otro contarás con una semana de “entretiempo”. Puedes usarla para terminar de ajustar tu proyecto y entregarlo, ayudar a tus compañeros/as a terminar su proyecto, o incluso descansar.

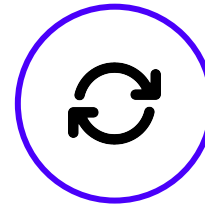
Tus deadlines

Al final de cada Sprint deberás entregar tu proyecto*. En un plazo máximo de una semana recibirás la devolución de tu evaluador/a:



Aprobado

Cuando hayas cumplido satisfactoriamente con el 100% de los puntos del checklist.



Para rehacer

Cuando haya puntos del checklist a mejorar. ¡Tendrás una cantidad ilimitada de iteraciones!

Obtendrás tu certificación cuando cuentes con **todos los proyectos aprobados.**



*Entre un Sprint y otro contarás con una semana de “entretiempo”. Puedes usarla para terminar de ajustar tu proyecto y entregarlo, ayudar a tus compañeros/as a terminar su proyecto, o incluso descansar.

Si necesitas ayuda



¿Tienes dudas sobre el feedback?

¡Escríbele a tu evaluador/a por Slack para consultarle sobre su feedback si lo necesitas!

¿Esperas tu devolución hace más de 7 días?

Envía un correo a hola@acamica.com solicitando la revisión de tu caso.



**¡Preparen sus proyectos
para mostrar en la DEMO
del encuentro que viene!**



Para la próxima

- Continúa trabajando el proyecto.
- Lee [la Toolbox 15](#) cuanto antes.
- ¡Prepárate para la Demo! Presta atención a la Toolbox para traer la presentación sobre tu proyecto. No te preocupes si aún no lo terminaste. Es una buena práctica como Data Scientist que puedas presentar lo que tienes hecho, lo que te falta, lo que harías distinto la próxima vez, y los aprendizajes que tuviste.

ACÀMICA