



Módulo V | Aprendizaje supervisado

Clase 12

Regresión y clasificación -
Algoritmos más comunes parte 1





¿Ponemos a grabar el
taller?

¿Qué vamos a ver hoy?



- Diferencia entre regresión y clasificación

- Algoritmos más comunes para regresión y clasificación parte 1

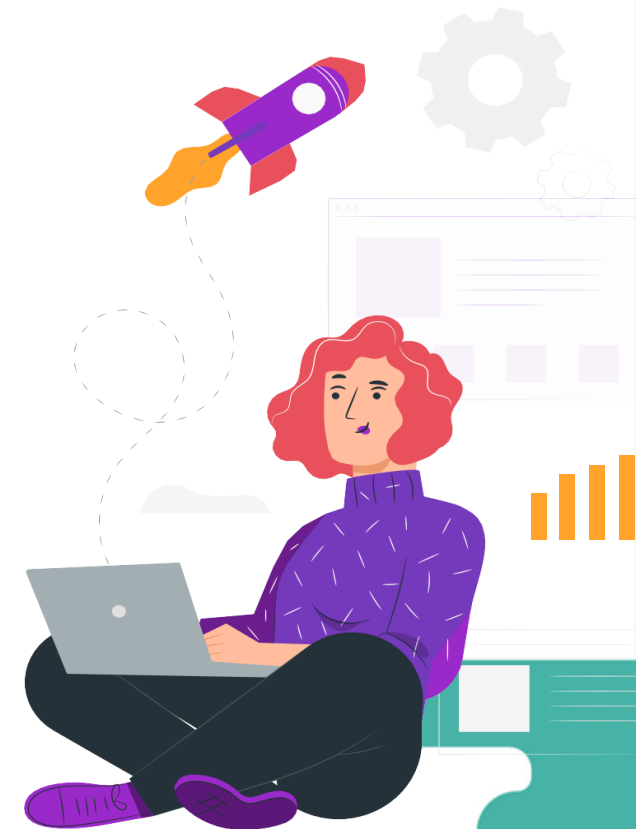
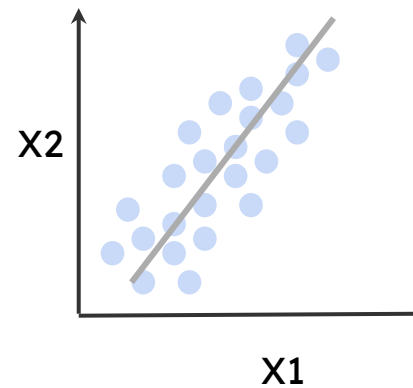


Regresión y Clasificación

Regresión

La variable dependiente es un **valor numérico**

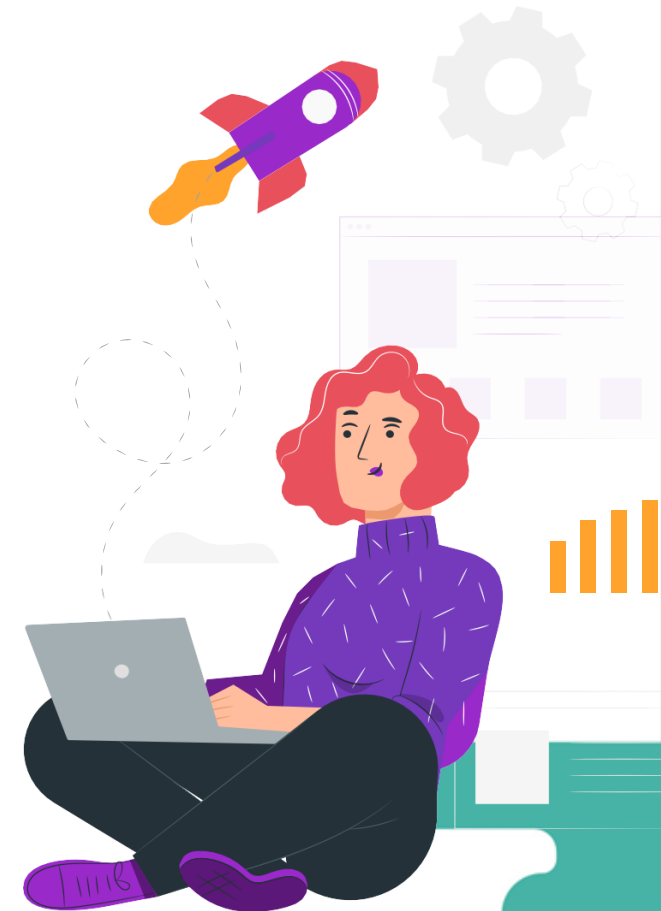
- Precio de una casa
- Valor de acciones



Regresión

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

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



Regresión

Algoritmos más comunes

A

Regresión Logística

B

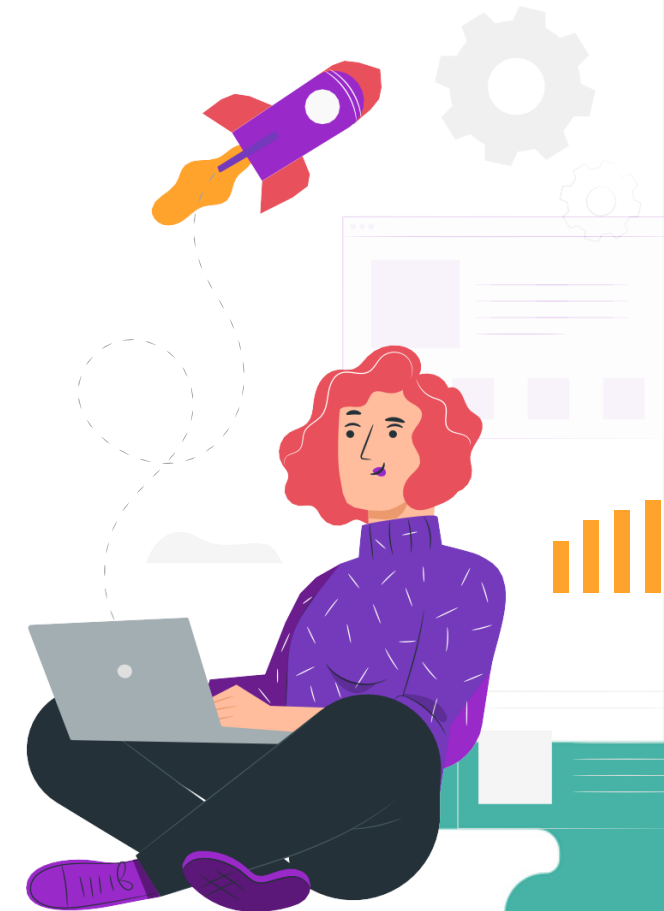
Árboles de Decisión

C

Random Forest

D

Support Vector Machine



Regresión

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

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



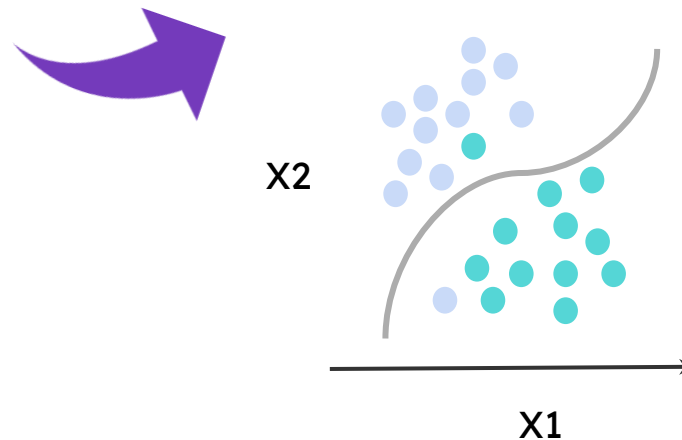


Clasificación

CLASIFICACIÓN

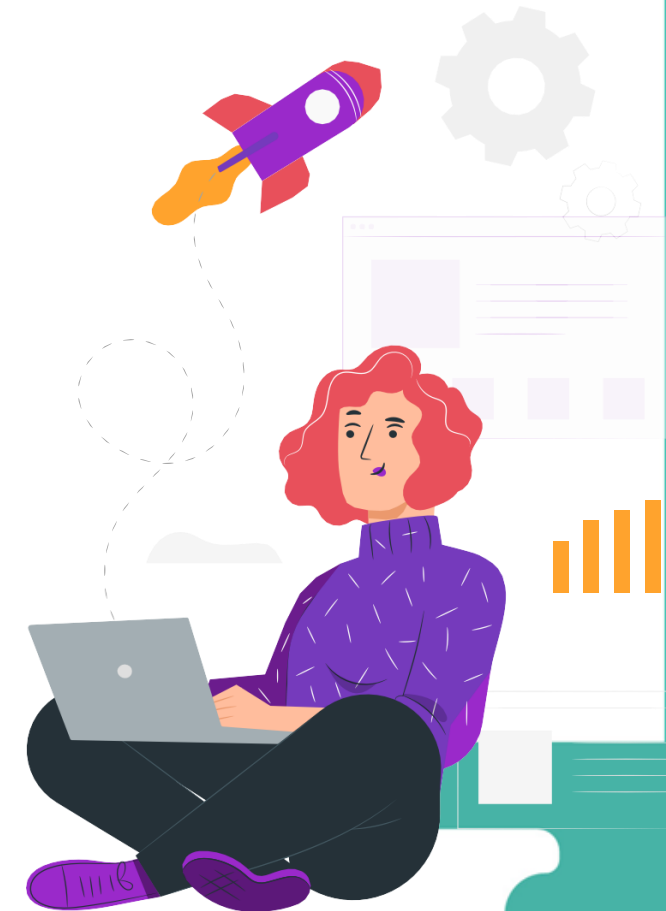
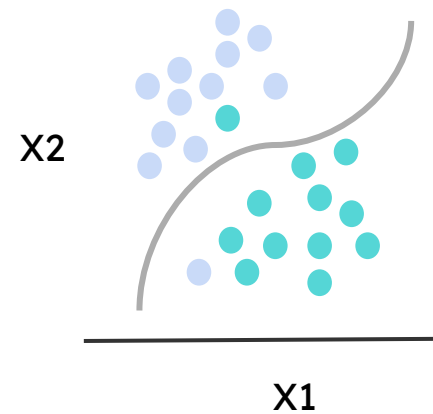
Cada instancia pertenece a **una clase**

- Email es spam o no
- Detección de Fraude
- Análisis de Sentimiento



CLASIFICACIÓN

Consiste en predecir la pertenencia de una observación a una clase Y en base a atributos X_1 , X_2



Clasificación

Algoritmos más comunes

A

Regresión Logística

B

Árboles de Decisión

C

Random Forest

D

Regresión de Máquinas de Soporte (SVR)





Repasamos en Kahoot

Respondemos algunas
preguntas en Kahoot



Algoritmos más comunes



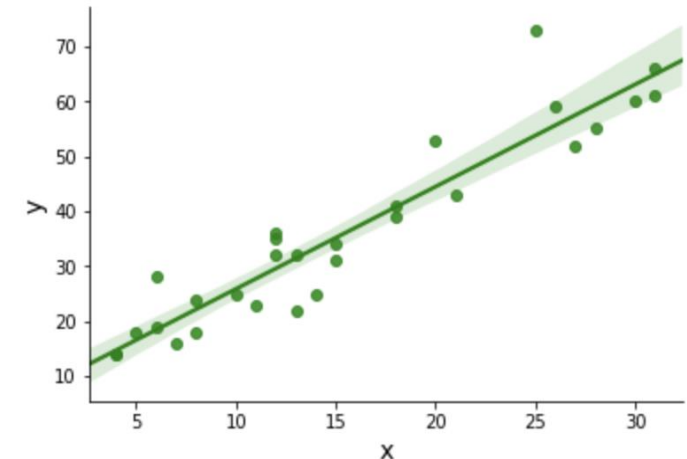
Regresión lineal

Regresión Lineal

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

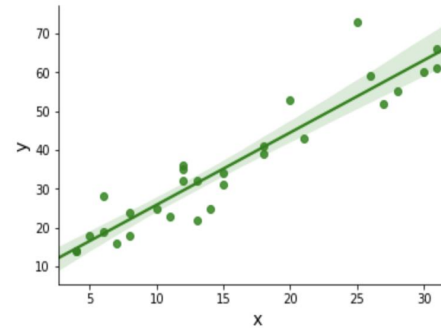
m: pendiente

b: ordenada al origen

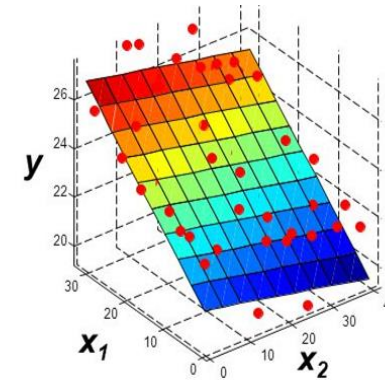
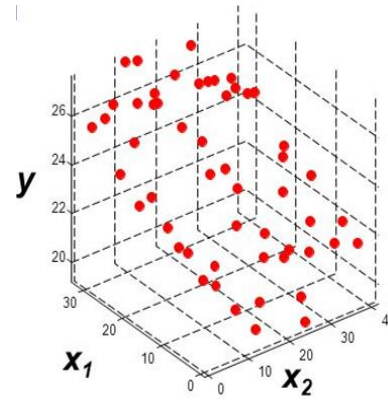


Regresión Lineal

Un atributo: x_1
 $Y = m_1x_1 + b$



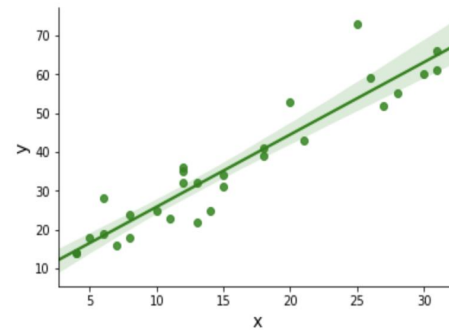
Dos atributos: x_1, x_2
 $Y = m_1x_1 + m_2x_2 + b$



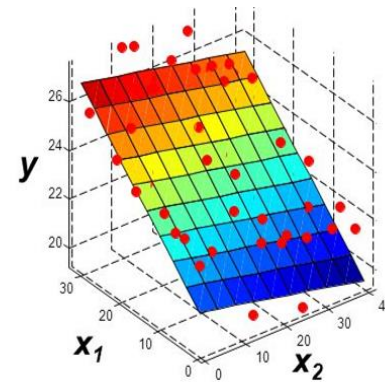
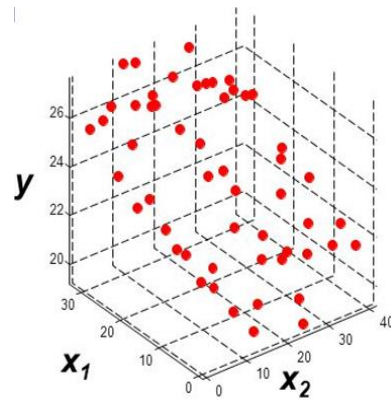
n atributos

Regresión Lineal

Un atributo: x_1
 $Y = m_1x_1 + b$

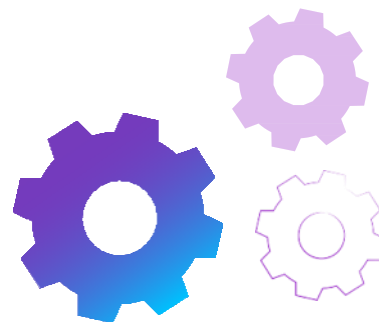


Dos atributos: x_1, x_2
 $Y = m_1x_1 + m_2x_2 + b$



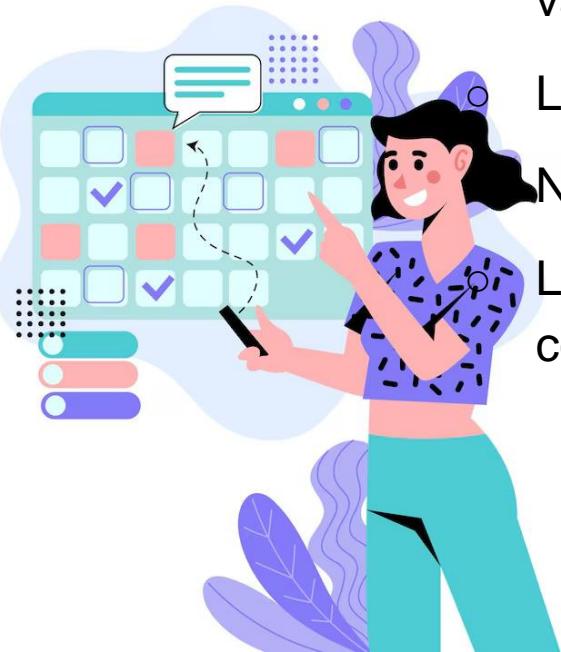
n atributos

$$\hat{y} = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x_3 \dots + \theta_n * x_n$$



Regresión Lineal

- Regresión
- **Supuestos**
 - Debe haber una relación lineal entre las variables independientes y la variable dependiente
 - Las variables tienen que tener una distribución normal
 - No debe haber colinealidad
 - La varianza del error condicional a las variables explicativas es constante a lo largo de las observaciones (Homocedasticidad)





Regresión logística

Regresión Logística

La regresión logística aplica el operador logístico a los datos para determinar la probabilidad de un punto a pertenecer a una determinada **clase**.

$$P(t) = \frac{1}{1 + e^{-t}}$$

Sigmoid Function

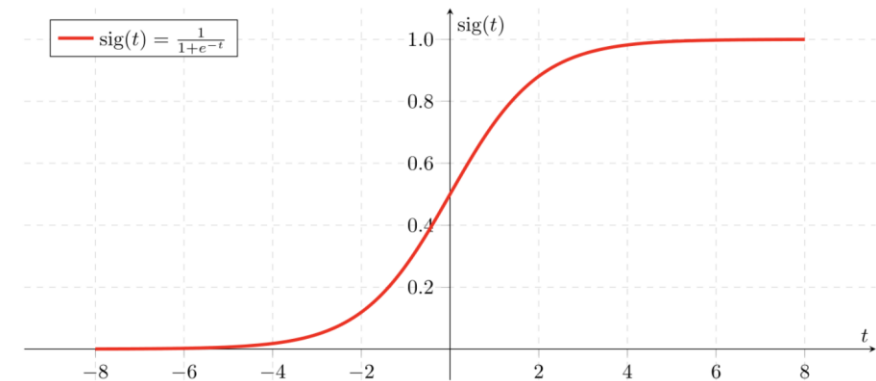
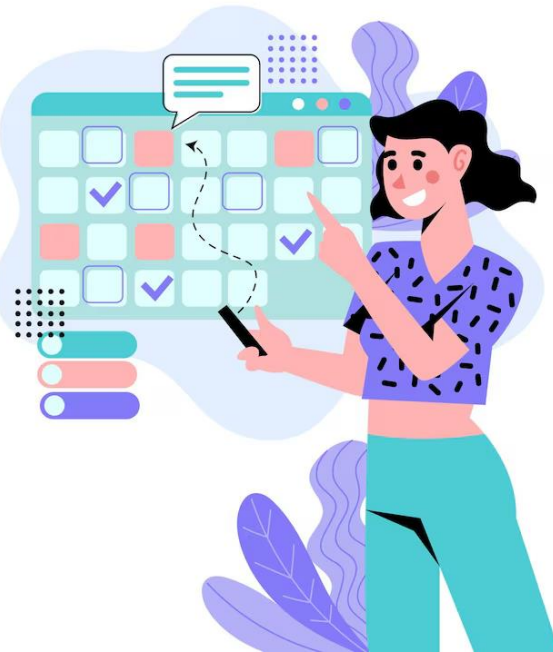


Figure 2: Sigmoid Activation Function

Regresión Logística

- Clasificación
- Utiliza la misma estructura que la regresión lineal pero luego transforma la variable en un número entre 0 y 1.
- **Supuestos**
 - Debe haber una relación lineal entre las variables independientes y la variable dependiente
 - Las variables tienen que tener una distribución normal
 - No debe haber colinealidad
 - Las observaciones deben ser independientes entre sí





Descanso

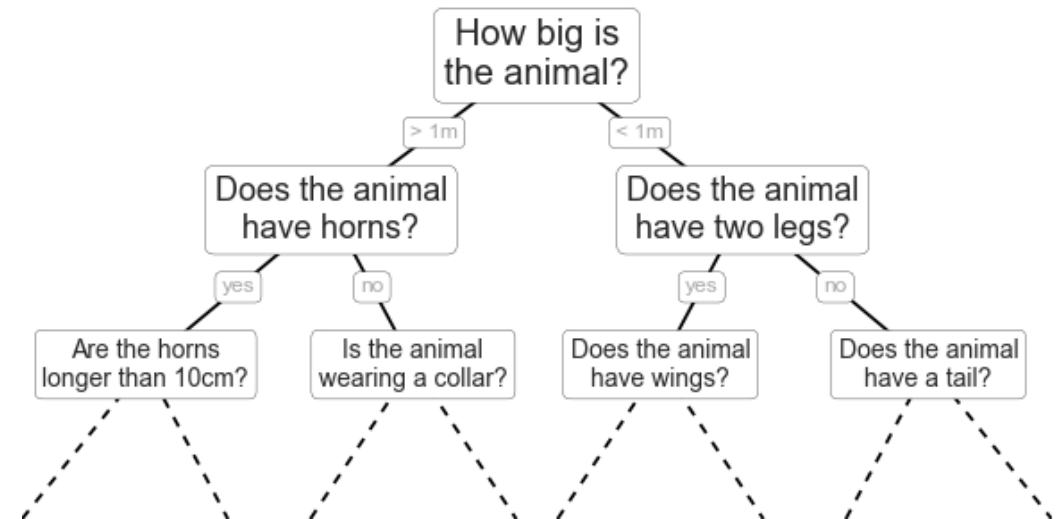
Nos vemos en 10 minutos



Árboles de decisión

Árboles de decisión

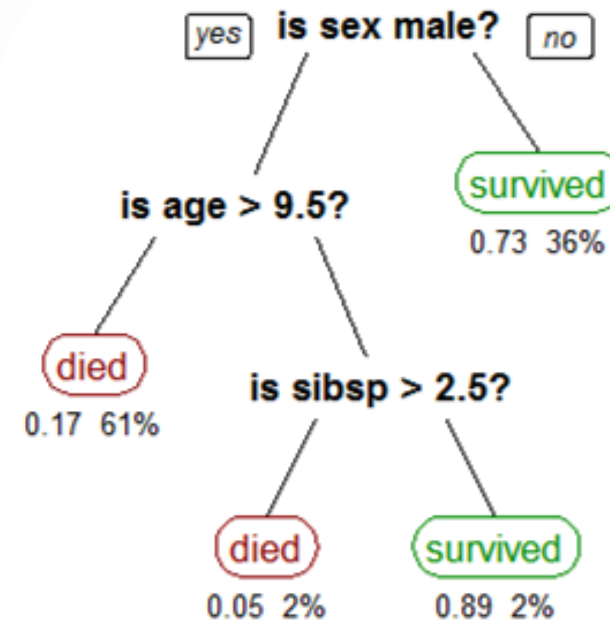
Un árbol de decisión **hace preguntas** y va clasificando de acuerdo a las respuestas.



Árboles de decisión

Un árbol de decisión **hace preguntas** y va clasificando de acuerdo a las respuestas.

¿Cómo decide preguntar?



Árboles de decisión

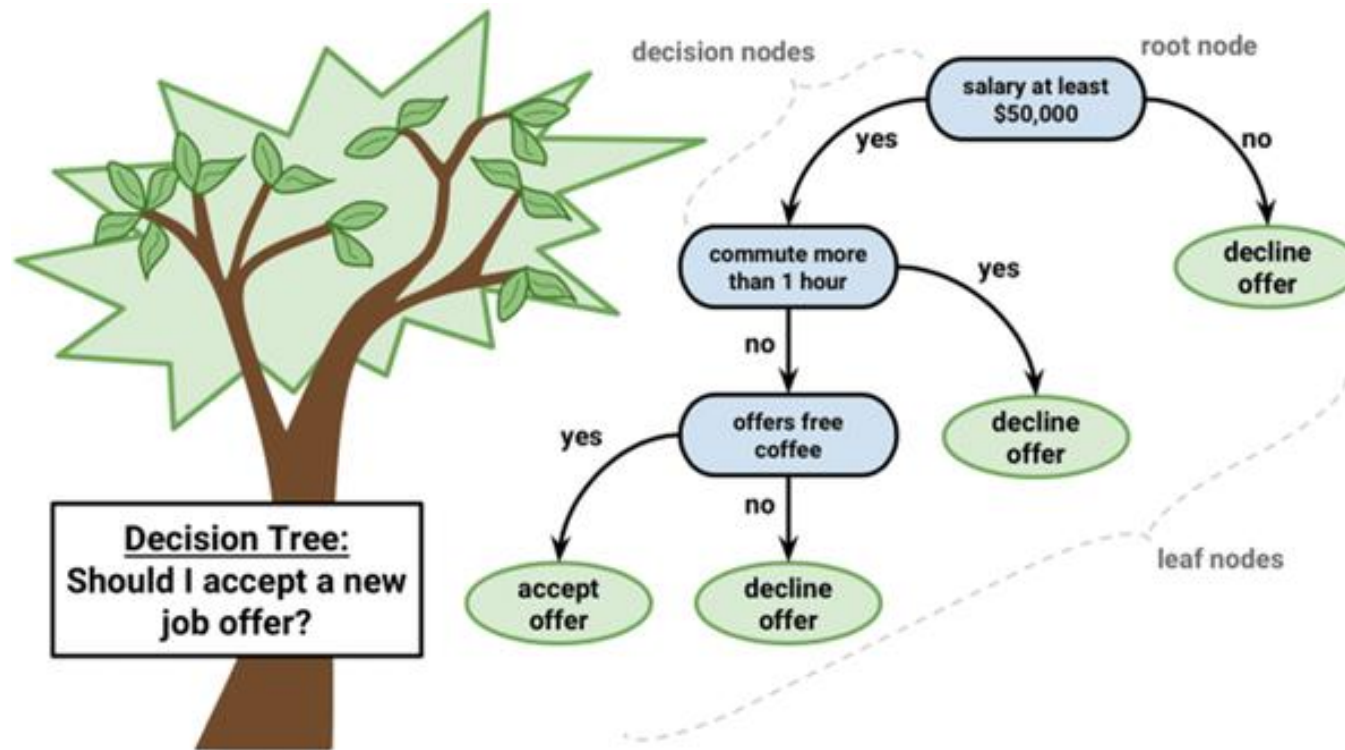
¿Cómo decide qué preguntar?

- 1. Impureza Gini
- 1. Ganancia de información

Son cálculos que se realizan sobre los datos. Le indican al modelo que tanta **relevancia** tiene un **feature** para **separar** las instancias por sus etiquetas.



Árboles de decisión



Árboles de decisión



- Clasificación y Regresión
- Simple de entender, interpretar y visualizar
- Modelo base para modelos más complejos (Random Forest, etc)
- **Parámetros para nodos:**
 - Impureza de Gini
 - Ganancia de información
- Solo toman atributos numéricos
- Tendencia a ajustarse a los datos de entrenamiento



Sección práctica:

Trabajamos con la Notebook 14
ajustando los primeros modelos

En la sala general Aprendizaje Supervisado - Parte 1

Trabajamos con la Notebook 14

Demostraremos cómo ajustar los modelos de regresión y clasificación vistos



Desafío 10

- Para la siguiente clase, repasar y ejercitar la notebook 14.



FUNDACIÓN
YPF

¡Muchas gracias!

