



Módulo 5 – Aprendizaje de Máquina Supervisado

# Problemas de Clasificación

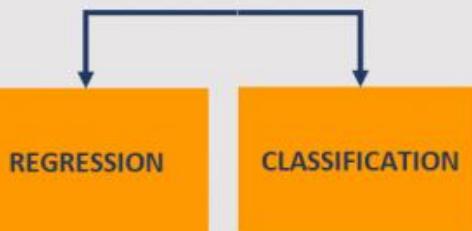
Especialización en Ciencia de Datos

# Tipos de ML

## MACHINE LEARNING

### SUPERVISED LEARNING

El algoritmo aprende a partir de la relación entre entradas y salidas en un set de entrenamiento



### UNSUPERVISED LEARNING

El algoritmo analiza los datos por tendencias y patrones sin conocer la salida



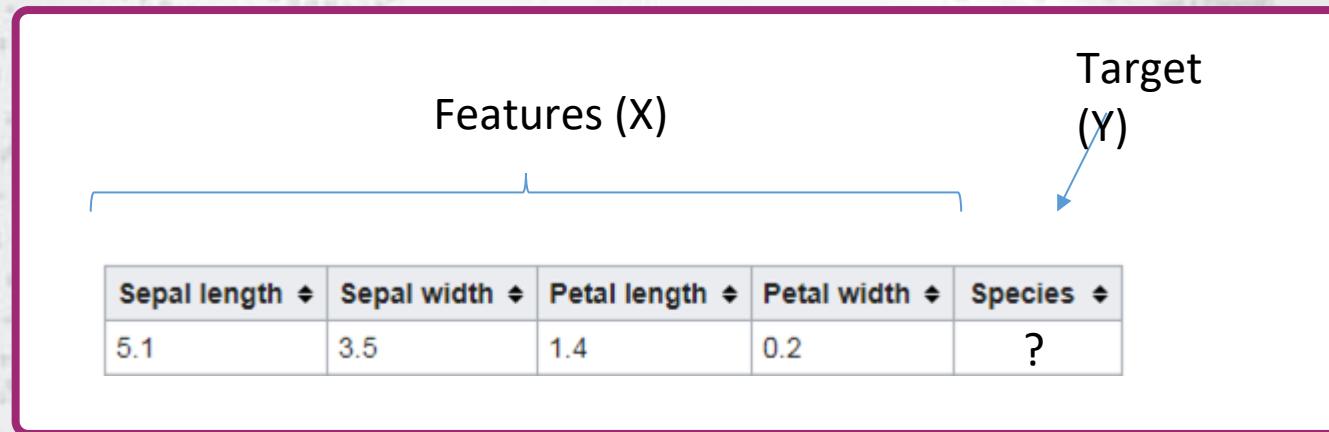
### REINFORCEMENT LEARNING

El algoritmo aprende sobre el tiempo para maximizar retorno basado en recompensas que recibe al realizar ciertas acciones



# ¿Qué es una tarea de clasificación?

Una tarea de clasificación es aquella que realiza el proceso de predicción de una clase o categoría (Y), a partir de sus features (X).



Entonces, para este caso, la tarea consiste en predecir qué especie de Iris (Y) correspondería clasificar esta instancia de datos (X).

Features (X)

Target (Y)

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	<i>I. setosa</i>
4.9	3.0	1.4	0.2	<i>I. setosa</i>
4.7	3.2	1.3	0.2	<i>I. setosa</i>
4.6	3.1	1.5	0.2	<i>I. setosa</i>
7.0	3.2	4.7	1.4	<i>I. versicolor</i>
6.4	3.2	4.5	1.5	<i>I. versicolor</i>
6.9	3.1	4.9	1.5	<i>I. versicolor</i>
5.5	2.3	4.0	1.3	<i>I. versicolor</i>
6.5	2.8	4.6	1.5	<i>I. versicolor</i>
7.2	3.0	5.8	1.6	<i>I. virginica</i>
7.4	2.8	6.1	1.9	<i>I. virginica</i>
7.9	3.8	6.4	2.0	<i>I. virginica</i>
6.4	2.8	5.6	2.2	<i>I. virginica</i>

## ¿Qué es una tarea de clasificación?

Un algoritmo de aprendizaje de máquina lleva a cabo la tarea de clasificación habiéndose entrenado previamente a partir de muchos ejemplos correctamente etiquetados.

X ? Y

# ¿Qué es una tarea de clasificación?

- ➊ Se habla de **clasificación binaria**, si la variable objetivo (Y) tiene solamente dos categorías posibles (por ejemplo: True/False, Si/No).
- ➋ Por otra parte, se habla de una **clasificación multi-clase** si la variable objetivo (Y) tiene más de dos categorías posibles (por ejemplo: Virginica, Versicolor, Setosa).
- ➌ Por último, se habla de una clasificación **multi-etiqueta** cuando una instancia pudiera estar asociada a más de una clase o categoría.

# Ejercicio 1 – Relacione el concepto

- 1** A partir de los textos de un correo electrónico, se desea entrenar un algoritmo que pueda determinar si un correo es spam.
  
- 2** A partir de la información demográfica y la cantidad de veces que un cliente ha sido expuesto a una publicidad, queremos saber si un cliente realizará una compra.
  
- 3** A partir de una fotografía del rostro de una persona, deseamos saber cuál es su nombre.
  
- 4** A partir de información de temperatura, humedad, presión atmosférica y velocidad del viento, queremos predecir cómo estará el clima el día de mañana.
  
- 5** A partir de un texto, queremos clasificar los tópicos al cual pertenece.
  
- 6** A partir de la información de cantidad de habitaciones, sector, metros cuadrados, cercanía al metro, dependencias, ubicación, año y material de construcción, queremos predecir el precio de un bien raíz.

Clasificación Binaria

Clasificación Multi-clase

Clasificación Multi-etiqueta

Ninguna de las anteriores

# Algoritmos de Clasificación

- ➊ Los algoritmos de clasificación son programas computacionales que están diseñados para encontrar modelos predictivos a partir de un set de datos de ejemplo. Dentro de los más populares, se pueden mencionar los siguientes:
  - Regresión Logística.
  - K-Nearest Neighbors.
  - Árboles de Decisión.
  - Máquinas de Soporte Vectorial.
  - Naive Bayes.
  - Redes Neuronales.
- ➋ Algunos de ellos, como la regresión logística y la máquina de soporte vectorial, están diseñados para clasificación binaria y de forma nativa no soportan clasificación multi-clases.

# Caso Social Network Ads

Features                      Variable Objetivo ↓

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

0: No compra  
1: Sí compra

Una empresa lanza una campaña en redes sociales para la promoción de un producto, se cuenta con un set de datos de las personas que fueron expuestas a dicha campaña y si compró o no el producto. Los datos son los siguientes:

- User ID: identificador del usuario en la red social.
- Gender: género.
- Age: edad.
- Estimated Salary: salario estimado.
- Purchased: si compró o no el producto.

Se busca construir un modelo predictivo que permita saber si una persona realizaría o no la compra, a partir de su edad y salario estimado.



# Dudas y consultas



KIBERNUM



# Fin Presentación



KIBERNUM