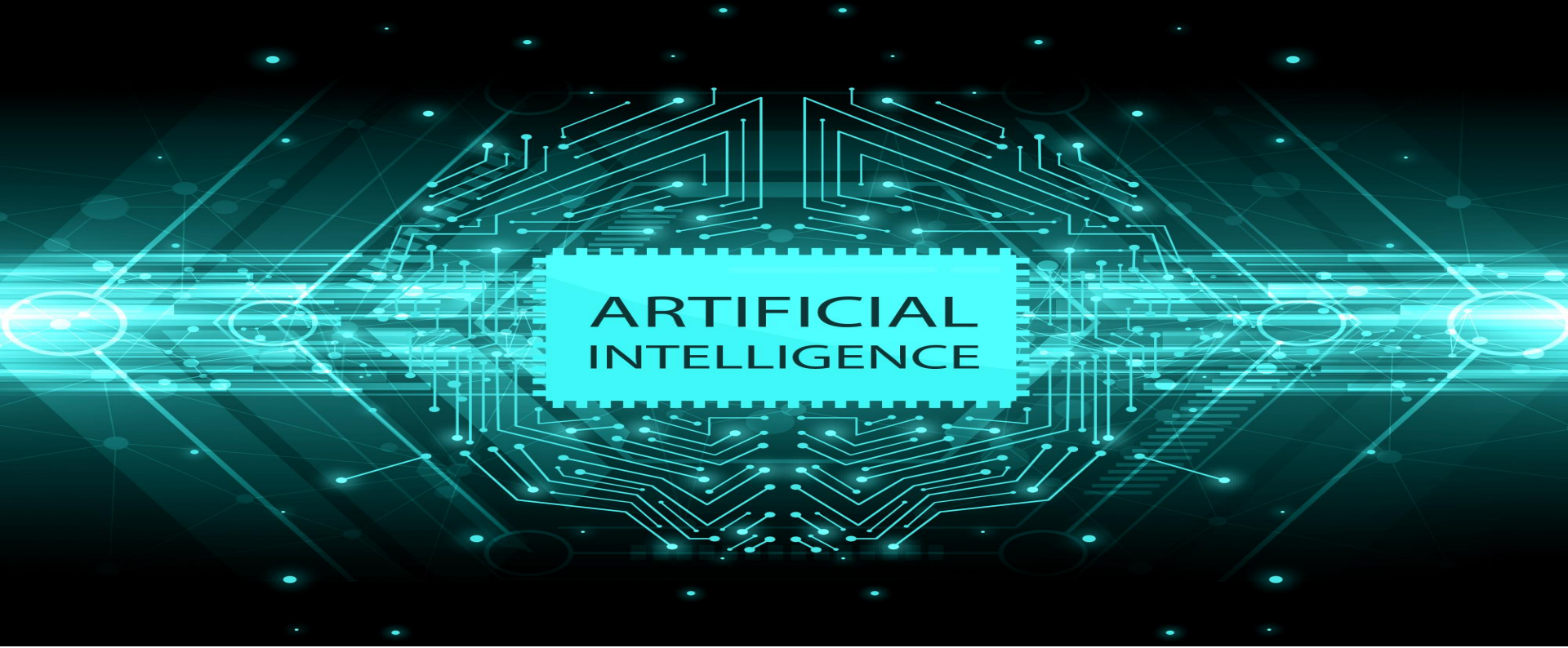
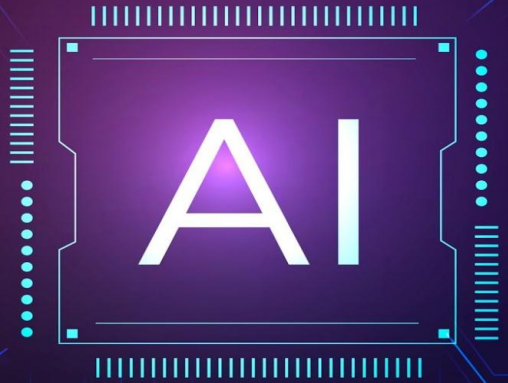


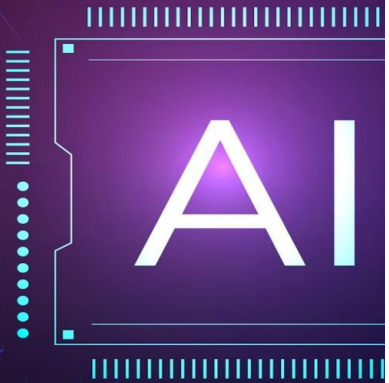

# Inteligencia Artificial

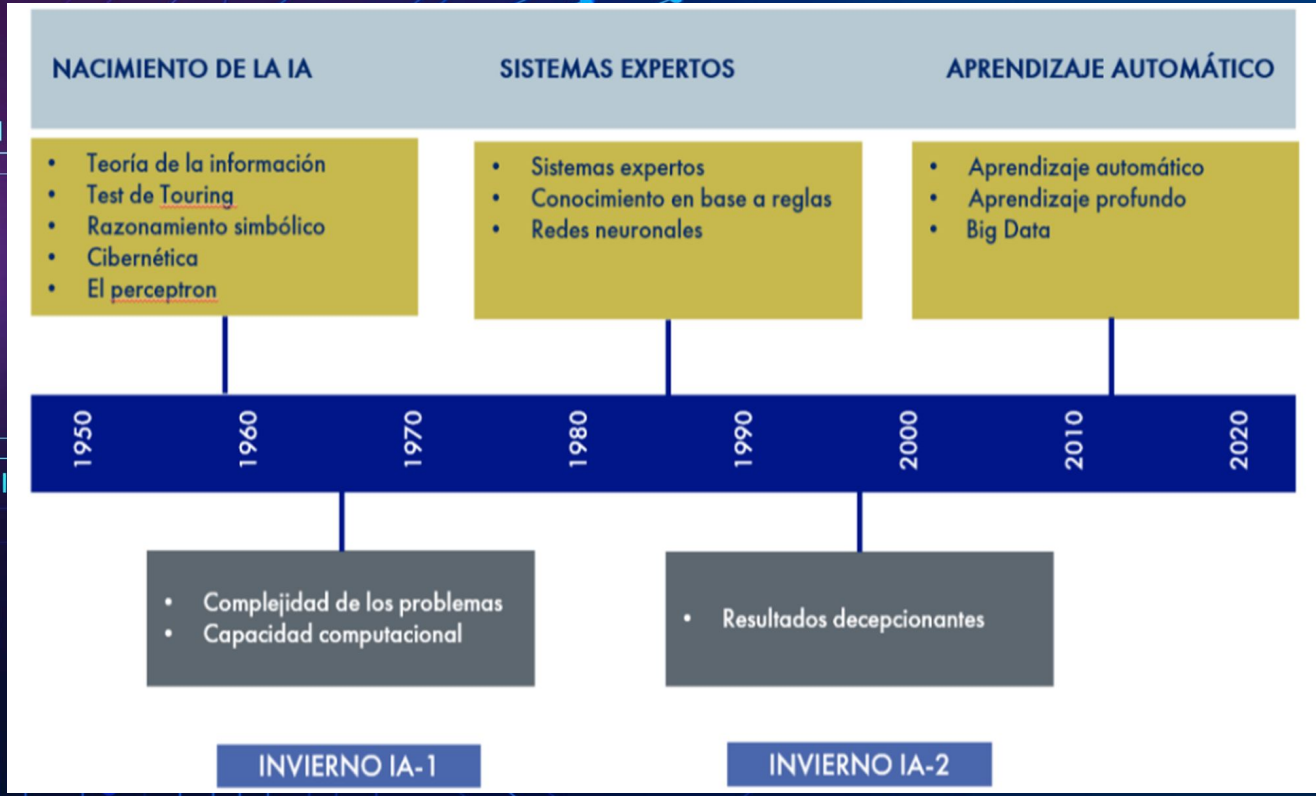
# Historia y Evolución de la Inteligencia Artificial



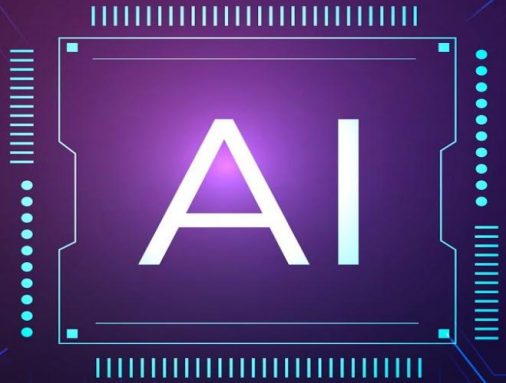



# Historia y Evolución de la Inteligencia Artificial



# Introducción a la inteligencia Artificial (AI)



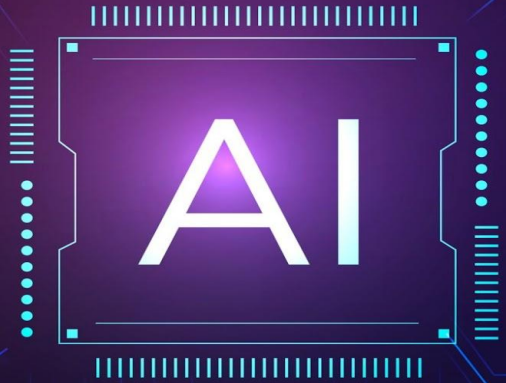

## Prueba de Turing

### Descripción

La prueba de Turing o test de Turing es una herramienta de evaluación de la capacidad de una máquina para exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este.

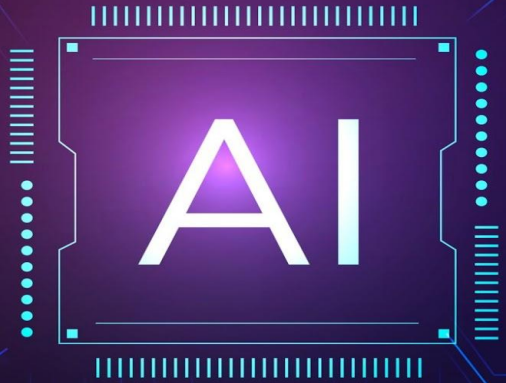
Wikipedia

# Introducción a la inteligencia Artificial (AI)




La inteligencia artificial (IA) es la capacidad de las computadoras para reemplazar tareas humanas con un poder y alcance impresionantes. Esta combina ingeniería y matemáticas para asegurar una labor más eficaz mediante sistemas informáticos sólidos.

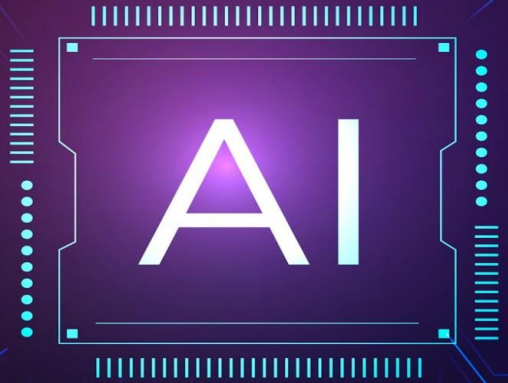
# Introducción a la inteligencia Artificial (AI)




“Es la ciencia e ingeniería para construir máquinas inteligentes, especialmente, programas de computación inteligentes. Así como, lo relativo a la tarea de usar computadoras para entender la inteligencia humana, pero no limitada a métodos observables biológicamente.”

**John McCarthy (2007)**

# Que puede y que no puede lograr la inteligencia Artificial (AI)



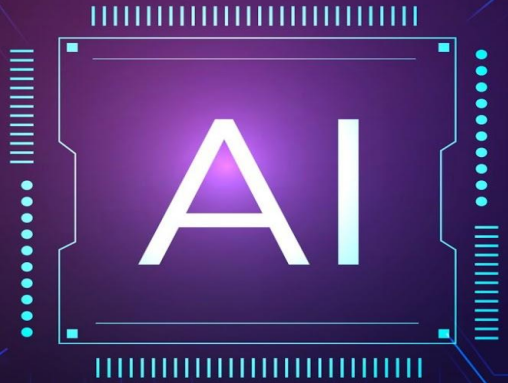

- **Puede:**

  - Predecir: a través de datos históricos.
  - Clasificar
  - Entender: puede resumir e interpretar textos por ejemplo
- - Puede aprender a decidir mejor que las personas en un campo en concreto
- - Generar contenido a través de los promps suministrados

**No puede:**

- Carece de la complejidad, creatividad y conciencia y la compasión humanas
- Crear de cero y ha voluntad.
- Innovar
- Opinar: mantiene ausencia de criterio.

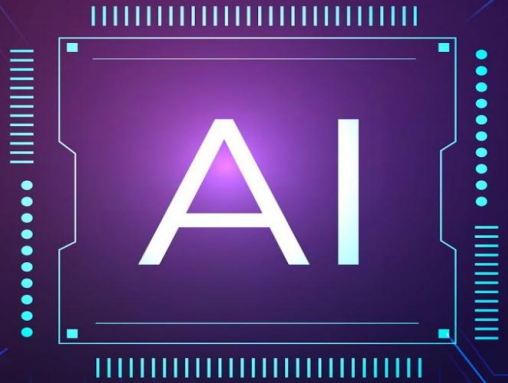
# Machine Learning y Deep Learning las bases de la AI




**Machine learning** es un subconjunto de la inteligencia artificial (IA). Se enfoca en enseñar a las computadoras para que aprendan de los datos y mejoren con la experiencia.

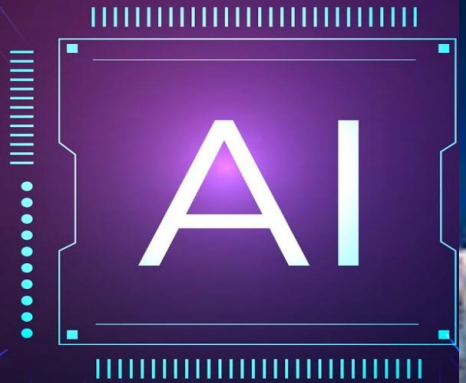
**Deep learning** trata de imitar el cerebro humano, aunque lejos de igualar su capacidad, lo que permite a los sistemas agrupar datos y realizar predicciones con una gran precisión

# Limitaciones en inteligencia Artificial (AI)



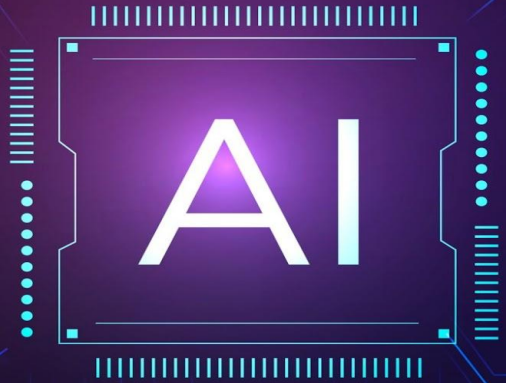

1. Falta de comprensión profunda
2. ChatGPT a veces escribe respuestas que suenan plausibles pero son incorrectas o sin sentido.
3. Sensibilidad a sesgo
4. Vulnerabilidad a la manipulación
5. Respuestas largas

# Machine Learning y Deep Learning las bases de la AI

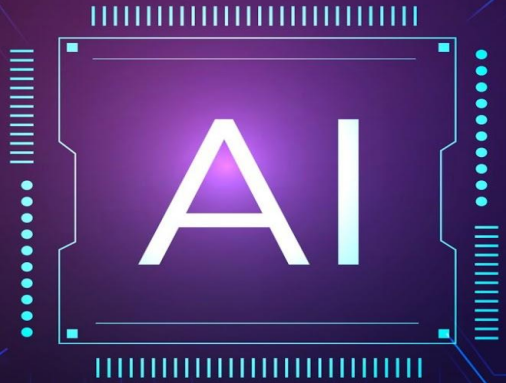





# Video Reflexión Pregunta al Doctor Kai-Fu Lee: Experto en AI.



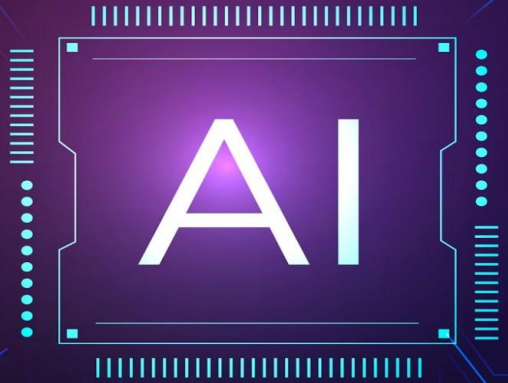
## El Álgebra Lineal es una herramienta fundamental en Machine Learning y Ciencia de Datos. Aquí tienes algunos puntos clave sobre su aplicación:




1. Representación de Modelos: Todos los modelos de Machine Learning se expresan mediante vectores y matrices. El Álgebra Lineal nos permite trabajar con estas estructuras de datos.
2. Procesamiento de Imágenes: Utilizando Python, puedes aplicar el Álgebra Lineal al procesamiento de imágenes.

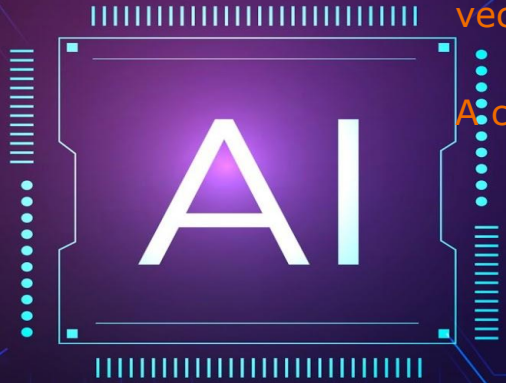
Ejemplos Específicos: En Ciencia de Datos y Machine Learning, el Álgebra Lineal se utiliza para definir operaciones sobre vectores y matrices. Esto es esencial para comprender y desarrollar modelos de manera efectiva

El Álgebra Lineal es una herramienta fundamental en Machine Learning y Ciencia de Datos. Aquí tienes algunos puntos clave sobre su aplicación:




La importancia que tiene el álgebra lineal dentro de Machine Learning es que convierte los problemas complicados en problemas sencillos, intuitivos y de cálculo eficiente.

# El Álgebra Lineal es una herramienta fundamental en Machine Learning y Ciencia de Datos. Aquí tienes algunos puntos clave sobre su aplicación:



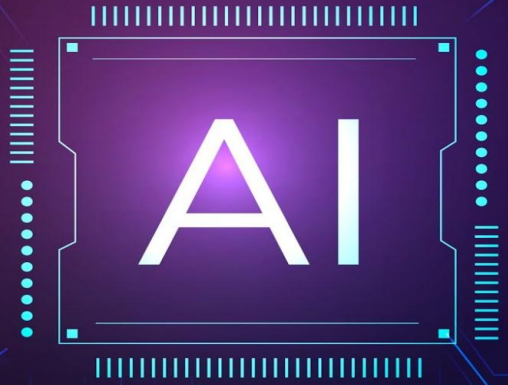

Los conjuntos de datos que se manejan suelen convertirse en forma de vectores o matrices, y las salidas también son escalares, vectores, matrices o tensores.

A continuación, se define cada una de estos términos.

- Escalares: son n único número o valor.
- Vectores: son una matriz de números, ya sea en una fila o en una columna, y se identifican con un solo índice.
- Matriz: es una matriz bidimensional de números, donde cada elemento se identifica con dos índices.
- Tensores: si llevamos una matriz un paso más allá, obtenemos un tensor. Se puede tener cualquier número de dimensiones en un tensor, aunque se hace más caro computacionalmente para manejar.



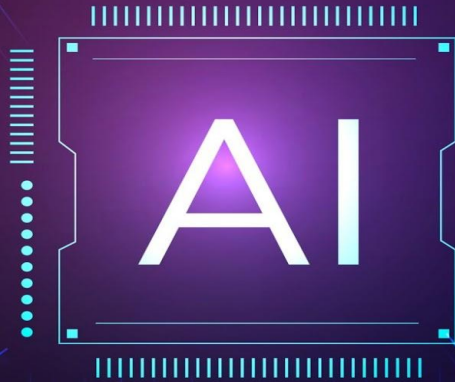
# Resuelva los siguietes ejercicios




- Con las siguientes matrices
- $A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -5 & 2 \end{bmatrix}$
- $B = \begin{bmatrix} 7 & -5 & 1 \\ 1 & -4 & 3 \end{bmatrix}$
- ✓ $C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$,
- ✓ $D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}$,
- ✓ $E = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$

- Determine las sgts operaciones
- *-2A, B-2A, 3C-E*

# Resuelva los siguietes operaciones con vectores




- Calcule u+v, u -2v

✓ $u = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, v = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$

✓ $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, v = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

# Sistemas de ecuaciones

- Resuelva el sistema de ecuaciones

✓
$$\begin{aligned} x_1 + 5x_2 &= 7 \\ -2x_1 - 7x_2 &= -5 \end{aligned}$$

✓
$$\begin{aligned} 2x_1 + 4x_2 &= -4 \\ 5x_1 + 7x_2 &= 11 \end{aligned}$$

- Resuelva el siguiente sistema de ecuación mediante el método de Eliminación de Gauss

✓
$$\begin{aligned} x_2 + x_3 &= -5 \\ x_1 + 3x_2 + 5x_3 &= -2 \\ 3x_1 + 7x_2 + 7x_3 &= 6 \end{aligned}$$

✓
$$\begin{aligned} x_2 + x_3 &= -5 \\ x_1 + 3x_2 + 5x_3 &= -2 \\ 3x_1 + 7x_2 + 7x_3 &= 6 \end{aligned}$$

# Funciones

- Calcule el valor de la función en los puntos x = 1, x = 3

✓ $f(x) = x^2$

✓ $f(x) = (2x - 1)^3$

✓ $h(t) = \frac{t}{t^2 - 1}$

# Matrices I

- Con las siguientes matrices
- $A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -5 & 2 \end{bmatrix}$
- $B = \begin{bmatrix} 7 & -5 & 1 \\ 1 & -4 & 3 \end{bmatrix}$
- ✓ $C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$,
- ✓ $D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}$,
- ✓ $E = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$

- *AC, CD, EB (multiplicación de matrices)*

# Ecuación de la línea recta

La ecuación de la recta es fundamental en geometría analítica y nos permite describir matemáticamente la posición de una recta en el plano cartesiano.

**Definición de recta**:

- Una recta es una sucesión infinita de puntos alineados en una misma dirección en el plano.
- Puede ser vertical, horizontal o diagonal según su orientación.

**Definición de ecuación de la recta**:

- La ecuación de la recta es una expresión algebraica que describe todos los puntos de la recta.
- Describe la posición tanto en el eje X como en el eje Y del plano cartesiano.

**Ecuación general de la recta**:

- La ecuación general de la recta se aplica a todas las rectas existentes en el plano cartesiano.
- Siempre cumple con la forma:
  $Ax+By+C=0$

# Ejercicio

Calcular la pendiente de la recta **2y = 4x + 12**

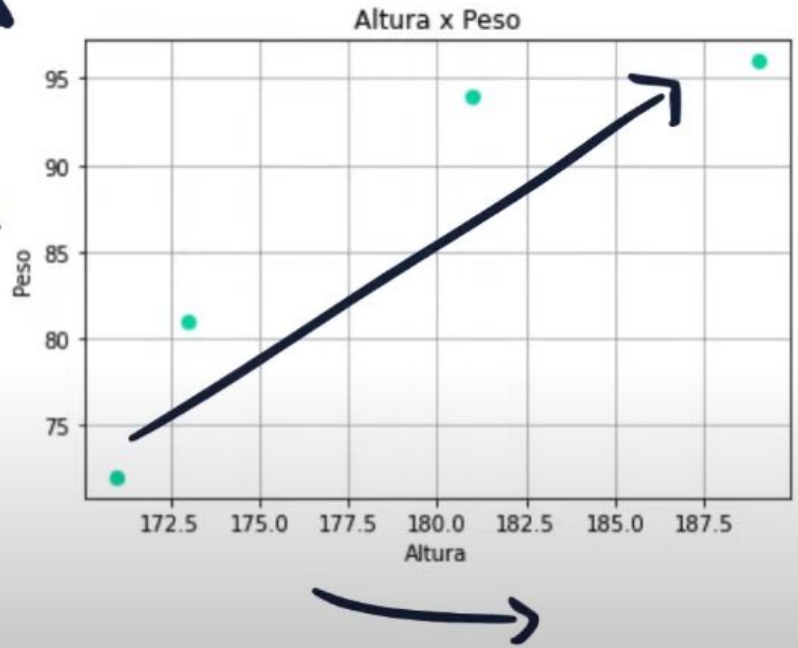
Encontrar la ecuación de la recta que pasa por los puntos **A = (2;3) y B = (6;5).**

# ¿Qué es la regresión lineal?

El análisis de regresión lineal se utiliza para predecir el valor de una variable basándose en el valor de otra variable. La variable que desea predecir se denomina variable dependiente. La variable que está utilizando para predecir el valor de la otra variable se denomina variable independiente.

$$y = \theta_0 + \theta_1 x$$

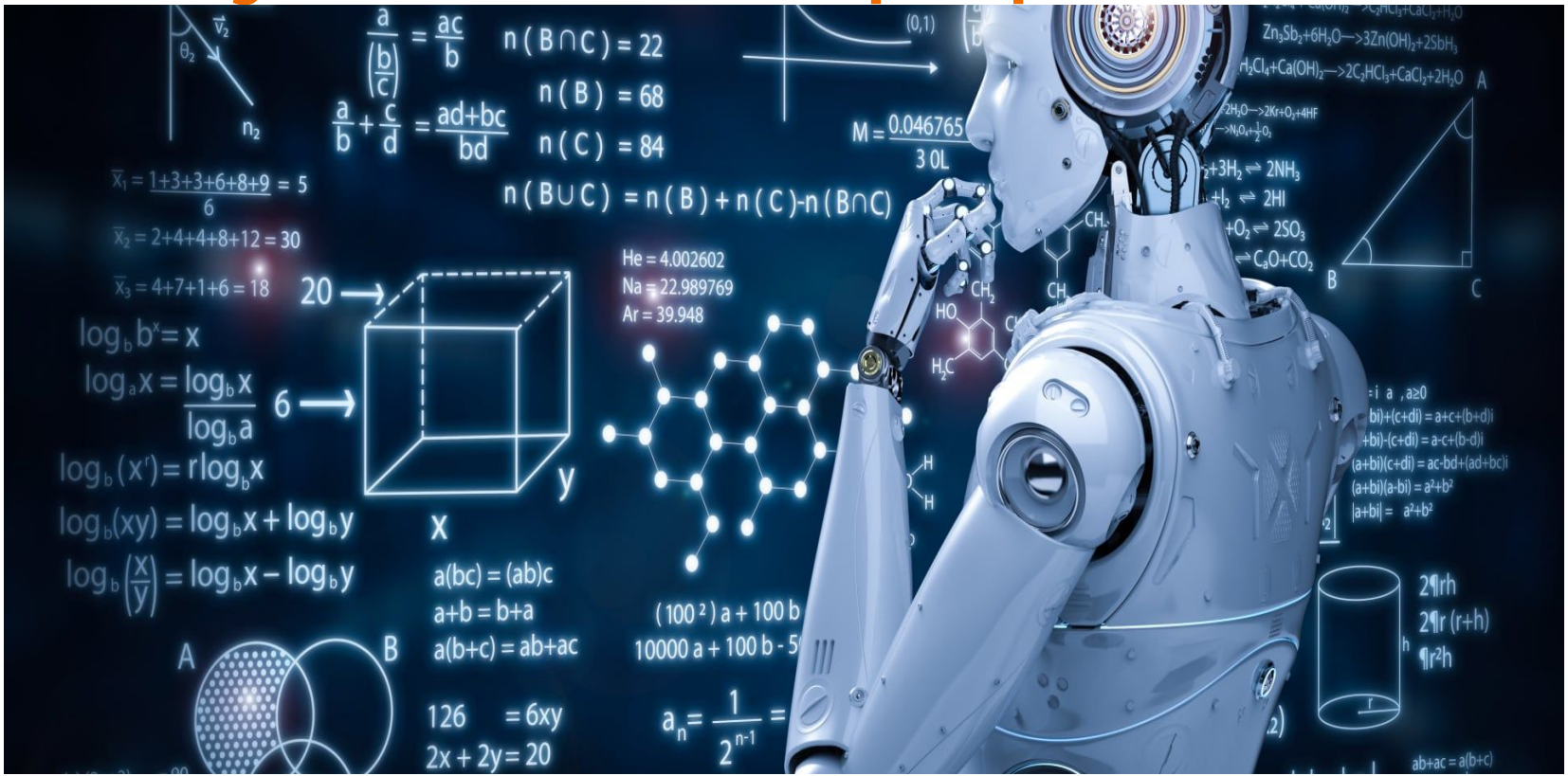
| Nombre | Altura (cm) | Peso (kg) |
|---|---|---|
| Antonio | 173 | 81 |
| José | 171 | 72 |
| Carlos | 189 | 96 |
| Enrique | 181 | 94 |
| David | 179 | ¿? |



| Nombre | Altura (cm) | Peso (kg) |
| --- | --- | --- |
| David | 179 | ¿? |



# Inteligencia artificial, conceptos previos




https://quizizz.com/join?gc=69908935

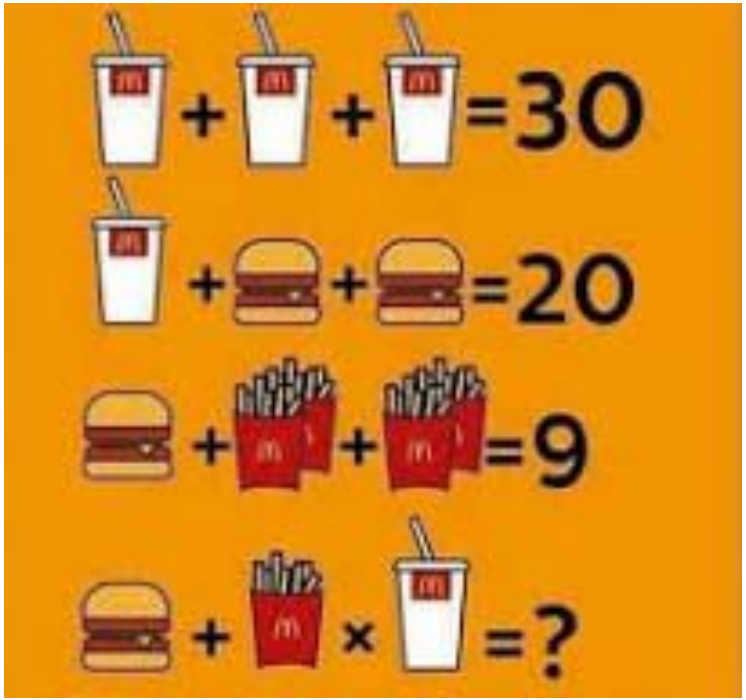
# ¿Qué es inteligencia artificial?

Es la capacidad que tienen los ordenadores o máquinas de imitar las funciones cognitivas humanas como el aprendizaje y la resolución de problemas, utilizando matemáticas, estadística e, incluso, relacionando otras disciplinas como la informática, mecánica y electrónica. En esta disciplina, existen campos como la visión artificial, el procesamiento del lenguaje natural, la robótica, los sistemas expertos, y el machine learning.





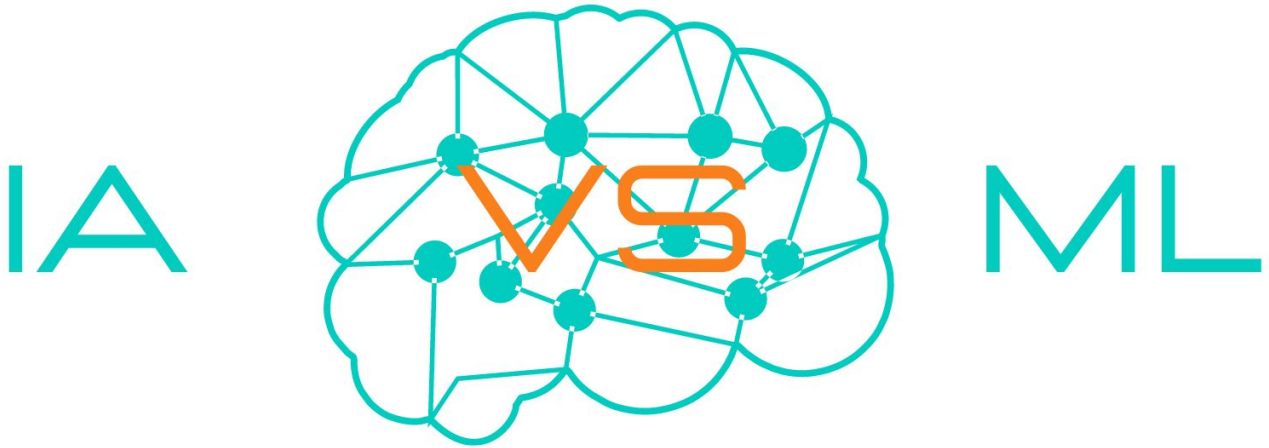
# IA



# MACHINE LEARNING

El ML examina, evalúa y compara datos para encontrar similitudes y patrones de comportamiento que finalmente permite hacer predicciones. Permite que el ordenador aprenda algo sin programarlo específicamente para ello, de esta forma va aprendiendo y mejorando por sí mismo a partir de los datos. Por ello, las empresas utilizan Machine Learning para tratar sus datos y poder tomar decisiones basadas en las predicciones, agrupaciones o clasificaciones proporcionadas por los algoritmos de ML, y así, obtener beneficios económicos.

# ¿Cuál es la diferencia entre la Inteligencia Artificial y el Machine Learning?



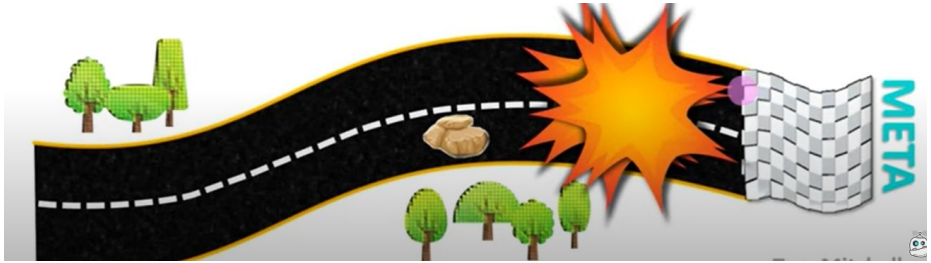
# IA vs ML

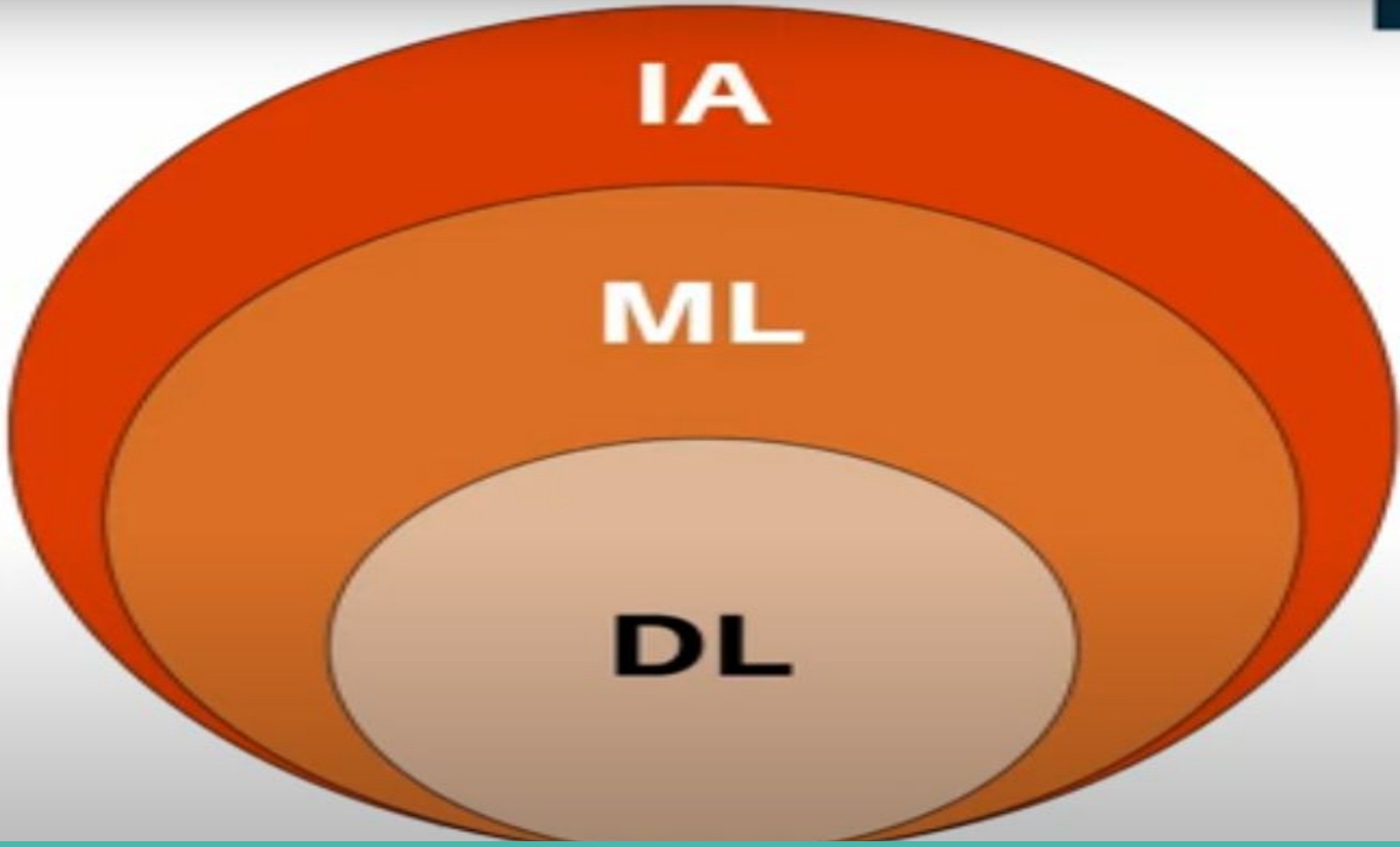









IA
ML
DL

# Qué hacer con los datos?



# ¿Cómo funciona?



¿Qué día y a que hora vende más?

¿Existe algún factor que incremente sus ventas?

- Análisis del problema
- Análisis de datos
- Preparación de datos
- Modelado
- Evaluación
- Implementación

# ¿Cómo funciona?



Fecha
Hora
Concepto
Cantidad
Precio Unitario
Atendió
Número de mesa
Promoción

Análisis del problema

Análisis de datos

Preparación de datos

Modelado

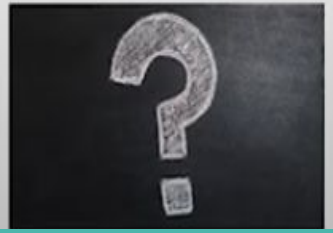
Evaluación

Implementación

# ¿Cómo funciona?







- Análisis del problema
- Análisis de datos
- Preparación de datos
- Modelado
- Evaluación
- Implementación

# ¿Cómo funciona?

| Fecha | Hora | Concepto | Cantidad | P. Unitario |
|---|---|---|---|---|
| 01 Agosto | 13:00 | Tacos de canasta | 5 | $10.00 |
| 01 Ag | 1:00 p.m. | Tacos al pastor | 10 | $5.00 |
| 2 Agosto 2019 | 13:30 | Tacos al pastor con piña | 3 | $7.00 |
| 02 Ag 19 | 13:40 | Tacos de canasta | ocho | $10.00 |
| 04 Ag 19 | 13:50 | Tacos de suadero | 1 | $9,900 |
| 10 agosto 2019 | 16:00 | | | $35.00 |

Análisis del problema

Análisis de datos

Preparación de datos

Modelado

Evaluación

Implementación

# ¿Cómo funciona?

| Fecha | Hora | Concepto | Cantidad | P. Unitario |
|---|---|---|---|---|
| 01-08-19 | 13:00 | Tacos_canasta | 5 | 10.00 |
| 01-08-19 | 13:00 | Tacos_pastor | 10 | 5.00 |
| 02-08-19 | 13:30 | Tacos_pastor_piña | 3 | 7.00 |
| 02-08-19 | 13:40 | Tacos_canasta | 8 | 10.00 |

- Análisis del problema
- Análisis de datos
- **Preparación de datos**
- Modelado
- Evaluación
- Implementación

# ¿Cómo funciona?



- Análisis del problema
- Análisis de datos
- Preparación de datos
- **Modelado**
- Evaluación
- Implementación

# ¿Cómo funciona?

1. Apropiado para el problema
2. Datos adecuados
3. Requerimientos



# ¿Cómo funciona?

No hay suficientes datos

Datos desactualizados

No comprendimos el problema

Modelo inadecuado



# Implementación




**Martes (14:00 p.m. – 16:00 p.m.)**
**Viernes ( 9:00 p.m – 11:00 p.m. )**
**Sábado ( 11:00p.m. – 2:00 a.m. )**

# Qué son los datos

Es una colección de ejemplos o muestras que se utilizan para entrenar y evaluar un modelo.

# 1. conjunto de datos

***dataset***

| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |
| nissan | negro | 2023 | $471,000 |
| kia | azul | 2020 | $340,000 |

automoviles.csv

| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |

**80%**
**(entrenamiento)**



| nissan | negro | 2023 | $471,000 |
|---|---|---|---|
| kia | azul | 2020 | $340,000 |

**20%**
**(validación)**

# Instancia: Se refiere a un solo elemento o muestra dentro de un conjunto de datos.




| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |
| nissan | negro | 2023 | $471,000 |
| kia | azul | 2020 | $340,000 |

# Atributo: Se refiere a una característica específica que describe una instancia.

| numérico |
|---|
| 45.8 |
| 19.2 |
| 87.5 |
| 38.2 |

| categórico |
|---|
| 0 |
| 1 |
| 1 |
| 0 |

| texto |
|---|
| rojo |
| azul |
| rojo |
| amarillo |

| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |
| nissan | negro | 2023 | $471,000 |
| kia | azul | 2020 | $340,000 |

# En la tabla se representan 3 atributos

| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |
| nissan | negro | 2023 | $471,000 |
| kia | azul | 2020 | $340,000 |

# Etiquetas ó Label: representan la variable objetivo que se intenta clasificar o predecir

| marca | color | año | precio |
|---|---|---|---|
| volkswagen | gris | 2018 | $219,000 |
| nissan | rojo | 2011 | $126,000 |
| toyota | azul | 2020 | $345,000 |
| toyota | amarillo | 2016 | $249,000 |
| volkswagen | blanco | 2009 | $98,000 |
| toyota | blanco | 2023 | $428,000 |
| nissan | azul | 2017 | $196,000 |
| kia | plata | 2024 | $567,000 |

atributos

etiquetas

# ANÁLISIS EXPLORATORIO DE LOS DATOS (EDA)

una de esas tareas fundamentales en esta etapa es el análisis exploratorio de los datos.

EDA ayuda a determinar la mejor manera de manipular las fuentes de datos para obtener las respuestas que necesita, lo que facilita a los científicos de datos descubrir patrones, detectar anomalías, poner a prueba una hipótesis o verificar suposiciones.

EDA se utiliza principalmente para ver qué datos se pueden revelar más allá de la tarea formal de modelado o de la prueba de hipótesis y proporciona una mejor comprensión de las variables del conjunto de datos y las relaciones entre ellas. También puede ayudar a determinar si las técnicas estadísticas que usted está considerando para el análisis de datos son adecuadas.

# Fases del análisis descriptivo



# Agenda



Explorar los conceptos básicos de datos



Explorar roles y responsabilidades en el mundo de los datos



Describir conceptos de datos relacionales.



Explorar conceptos de datos no relacionales

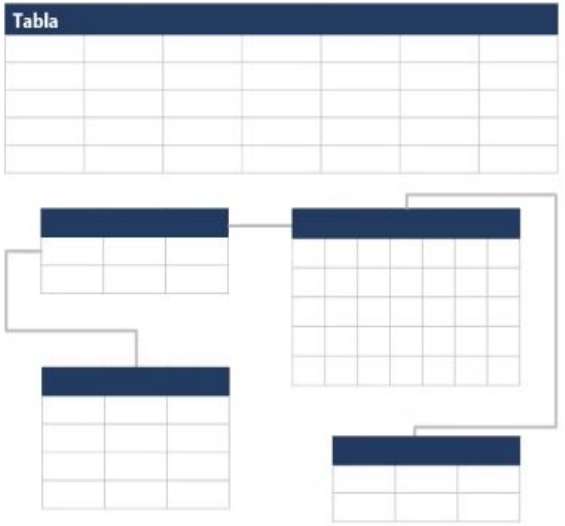


Explorar conceptos de análisis de datos

# ¿Qué son los datos?

Colección de hechos, números, descripciones, objetos, almacenados de forma estructurada, semiestructurada y no estructurada.

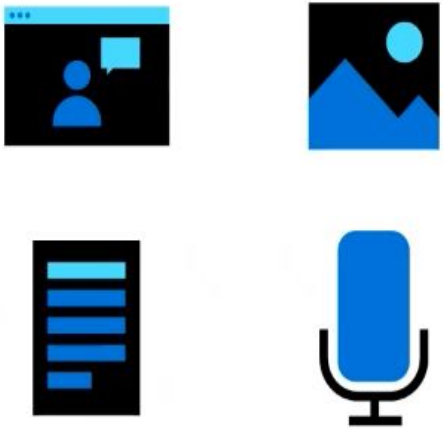
## Estructurados




## Semiestructurados

```
## Document 1 ## {
"customerID": "103248",
"name": { "first": "AAA",
"last": "BBB" }, "address": {
"street": "Main Street",
"number": "101", "city":
"Acity", "state": "NY" },
"ccOnFile": "yes",
"firstOrder": "02/28/2003" }
## Document 2 ## {
"customerID": "103249",
"name": { "title": "Mr",
"forename": "AAA",
"lastname": "BBB" },
"address": { "street":
"Another Street", "number":
"202", "city": "Bcity",
"county": "Gloucestershire",
"country-region": "UK" },
"ccOnFile": "yes" }
```

## No estructurados



# Roles en los datos

## Administrador de base de datos

Administración de base de datos

Implementa la seguridad de los datos

Copias de seguridad

Acceso de usuarios

Supervisa el rendimiento



## Ingeniero de datos

Procesos y canalizaciones de datos

Almacenamiento de ingesta de datos

Prepara datos para el análisis

Prepara datos para el procesamiento analítico



## Analista de datos

Proporciona conclusiones sobre los datos

Informes visuales

Modelado de datos para análisis

Combina datos para visualización y análisis



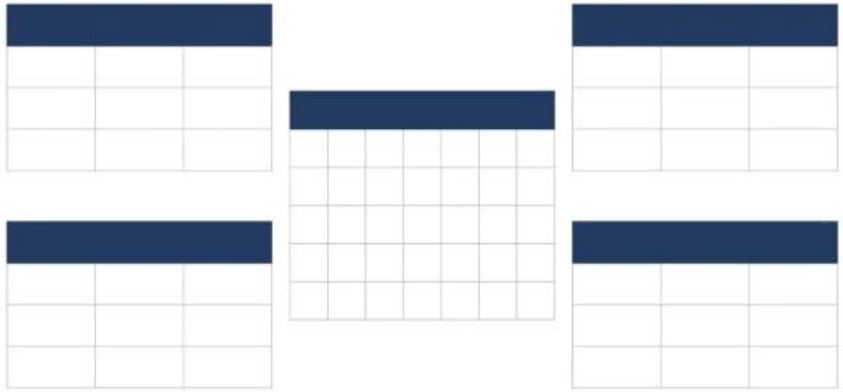
# Almacenes de datos transaccionales frente a analíticos

## Procesamiento transaccional en línea (OLTP)

| Cliente | | |
|---|---|---|
| IDdeCliente | CustomerName | CustomerPhone |
| | | |

| Pedidos | | |
|---|---|---|
| OrderID | IDdeCliente | OrderDate |
| | | |
| | | |

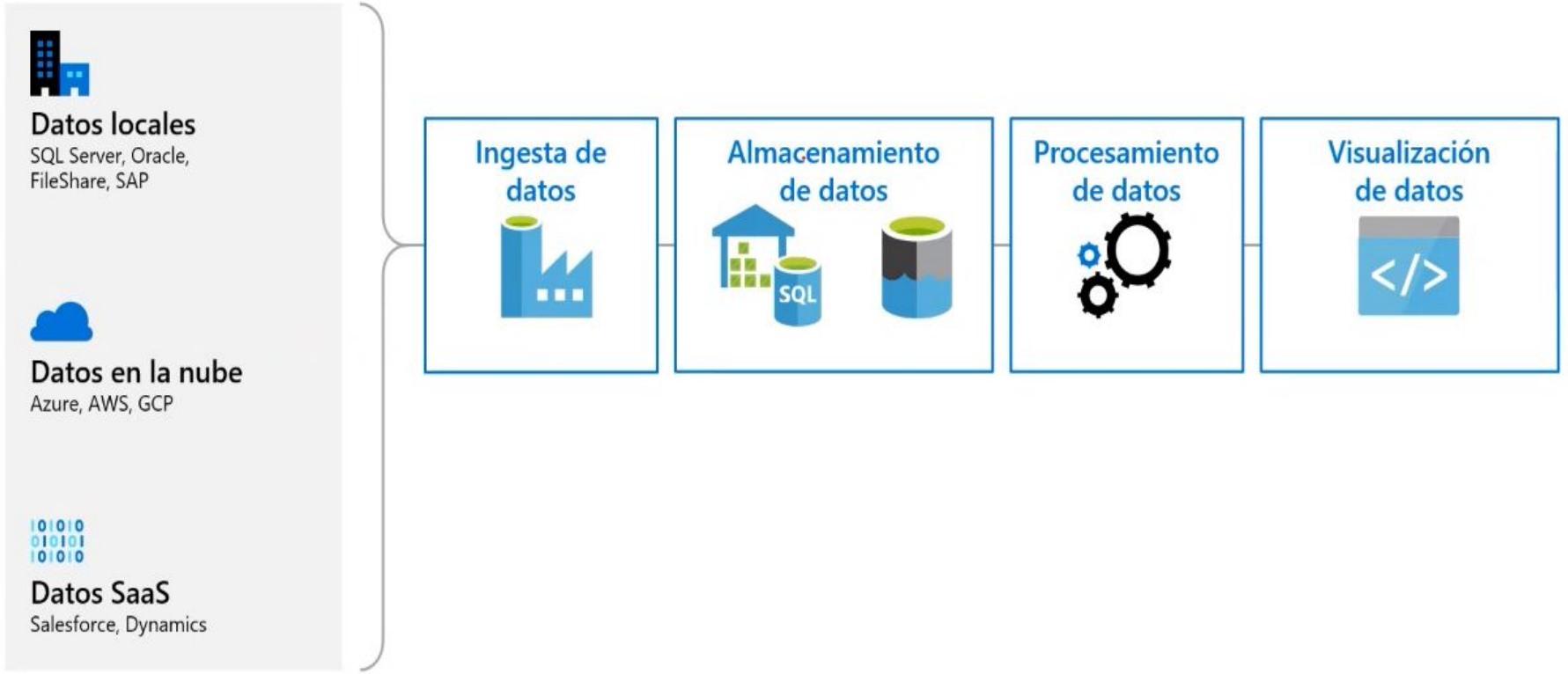
## Procesamiento analítico en línea (OLAP)







# Sistema de análisis



# Tablas

| Clientes | | |
|---|---|---|
| **IDdeCliente** | **CustomerName** | **CustomerPhone** |
| 100 | Mateo Lara | *XXX-XXX-XXXX* |
| 101 | Noam Maoz | *XXX-XXX-XXXX* |
| 102 | Vanja Matkovic | *XXX-XXX-XXXX* |
| 103 | Qamar Mounir | *XXX-XXX-XXXX* |
| 104 | Oscar Zamora | *XXX-XXX-XXXX* |
| 105 | Alexander Romero | *XXX-XXX-XXXX* |
| 106 | Eduardo Ponce | *XXX-XXX-XXXX* |
| 107 | Francisco Robles | *XXX-XXX-XXXX* |

Los datos se almacenan en una tabla

La tabla consta de filas y columnas

Todas las filas tienen el mismo número de columnas

Cada columna está definida por un tipo de dato

# Datos por lotes/datos de streaming



tablaProducto

| CódigoProducto | Producto | Precio |
|---|---|---|
| M12 | Monitor | Q 450,00 |
| CPU12 | CPU | Q 685,00 |
| M34 | Mouse | Q 35,00 |
| T12 | Teclado | Q 75,00 |
| A34 | Audífonos | Q 90,00 |

tablaMarca

| CódigoMarca | Marca |
|---|---|
| 34 | Bell |
| 23 | TXR |
| 12 | Bix |
| 87 | Bimox |

tablaCliente

| ID | nombre_cliente | Dirección | Teléfono |
|---|---|---|---|
| a12 | Eduardo López | Retalhuleu | 12345678 |
| b34 | Astrid Gómez | Mazatenango | 23568974 |

tablaPedido

| ID_pedido | Fecha | ID_cliente | CódigoProducto | CódigoMarca | Precio | Cantidad | Total |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |

| ID | nombre_cliente | Dirección | Teléfono | Producto | Precio | | CódigoProducto | Marca |
|---|---|---|---|---|---|---|---|---|
| a12 | Eduardo López | Retalhuleu | 12345678 | Monitor | Q | 450,00 | M12 | Bell |
| a12 | Eduardo López | Retalhuleu | 12345678 | CPU | Q | 685,00 | CPU12 | TXR |
| b34 | Astrid Gómez | Mazatenango | 23568974 | Mouse | Q | 35,00 | M34 | Bix |
| a12 | Eduardo López | Retalhuleu | 12345678 | Teclado | Q | 75,00 | T12 | Bix |
| b34 | Astrid Gómez | Mazatenango | 23568974 | Audífonos | Q | 90,00 | A34 | Bimox |

tablaCliente

| ID | nombre_cliente | Dirección | Teléfono |
|---|---|---|---|
| a12 | Eduardo López | Retalhuleu | 12345678 |
| b34 | Astrid Gómez | Mazatenango | 23568974 |

tablaProducto

| Marca | CódigoProducto | Producto | Precio | |
|---|---|---|---|---|
| Bell | M12 | Monitor | Q | 450,00 |
| TXR | CPU12 | CPU | Q | 685,00 |
| Bix | M34 | Mouse | Q | 35,00 |
| Bix | T12 | Teclado | Q | 75,00 |
| Bimox | A34 | Audífonos | Q | 90,00 |

# Relaciones

**Clientes**

| IDdeCliente | CustomerName | CustomerPhone |
|---|---|---|
| 100 | Mateo Lara | *XXX-XXX-XXXX* |
| 101 | Noam Maoz | *XXX-XXX-XXXX* |
| 102 | Vanja Matkovic | *XXX-XXX-XXXX* |
| 103 | Qamar Mounir | *XXX-XXX-XXXX* |
| 104 | Oscar Zamora | *XXX-XXX-XXXX* |
| 105 | Alexander Romero | *XXX-XXX-XXXX* |
| 106 | Eduardo Ponce | *XXX-XXX-XXXX* |

**Pedidos**

| OrderID | IDdeCliente | SalesPersonID |
|---|---|---|
| AD100 | 101 | *200* |
| AD101 | 101 | 200 |
| AD102 | 101 | 200 |
| AX103 | 10 | 201 |
| AS104 | 103 | 201 |
| AR105 | 105 | 200 |
| MK106 | 105 | 201 |



## En un esquema de base de datos normalizado:

- Las claves primarias y externas se utilizan para definir relaciones.
- No existe duplicación de datos (excepto los valores de clave en 3.ª forma normal (3NF).
- Los datos se recuperan uniendo tablas en una consulta

# Índices

| Clientes | | |
|---|---|---|
| IDdeCliente | CustomerName | CustomerPhone |
| 100 | Mateo Lara | *XXX-XXX-XXXX* |
| 101 | Noam Maoz | *XXX-XXX-XXXX* |
| 102 | Vanja Matkovic | *XXX-XXX-XXXX* |
| 103 | Qamar Mounir | *XXX-XXX-XXXX* |
| 104 | Oscar Zamora | *XXX-XXX-XXXX* |
| 105 | Alexander Romero | *XXX-XXX-XXXX* |
| 106 | Eduardo Ponce | *XXX-XXX-XXXX* |

| IDX-CustomerRegion | |
|---|---|
| IDdeCliente | Región |
| 100 | Francia |
| 101 | Brasil |
| 102 | Croacia |
| 103 | Jordán |
| 104 | España |
| 105 | Francia |
| 106 | EE. UU. |

**Un índice:**

Optimiza las consultas de búsqueda para una recuperación de datos más rápida.

Reduce la cantidad de páginas de datos que deben leerse para recuperar los datos en una instrucción SQL.

Los datos se recuperan uniendo tablas en una consulta

# Vista

| Clientes | | |
|---|---|---|
| IDdeCliente | CustomerName | CustomerPhone |
| 100 | Mateo Lara | *XXX-XXX-XXXX* |
| 101 | Noam Maoz | *XXX-XXX-XXXX* |
| 102 | Vanja Matkovic | *XXX-XXX-XXXX* |
| 103 | Qamar Mounir | *XXX-XXX-XXXX* |
| 104 | Oscar Zamora | *XXX-XXX-XXXX* |
| 105 | Alexander Romero | *XXX-XXX-XXXX* |
| 106 | Eduardo Ponce | *XXX-XXX-XXXX* |

| Pedidos | | |
|---|---|---|
| OrderID | IDdeCliente | SalesPersonID |
| AD100 | 101 | *200* |
| AD101 | 101 | 200 |
| AD102 | 101 | 200 |
| AX103 | 103 | 201 |
| AS104 | 103 | 201 |
| AR105 | 105 | 200 |
| MK106 | 105 | 201 |
| DB205 | 100 | 205 |

Crear la definición de una vista:

```
CREATE VIEW
vw_customerorders AS

SELECT Customers.CustomerID,
Customers.CustomerName,
Orders.OrderID FROM
Customers JOIN Orders on
Customers.CustomerID =
Orders.CustomerID
```

Recupere los pedidos realizados por el cliente 102 utilizando la vista:

```
SELECT CustomerName, OrderID
from vw_customerorders WHERE
CustomerID=102
```

**Una vista es una tabla virtual que se basa en el conjunto de resultados de la consulta:**

Las vistas se crean para simplificar la consulta.

Combinan datos relacionales en una vista única de panel.

# Comprobación de Conocimientos



¿Cuál de las siguientes afirmaciones es una característica de una base de datos relacional?

- ❑ Todos los datos deben almacenarse como cadenas de caracteres.
- ❑ Una fila en una tabla representa una sola entidad
- ❑ Diferentes filas en la misma tabla pueden contener diferentes columnas



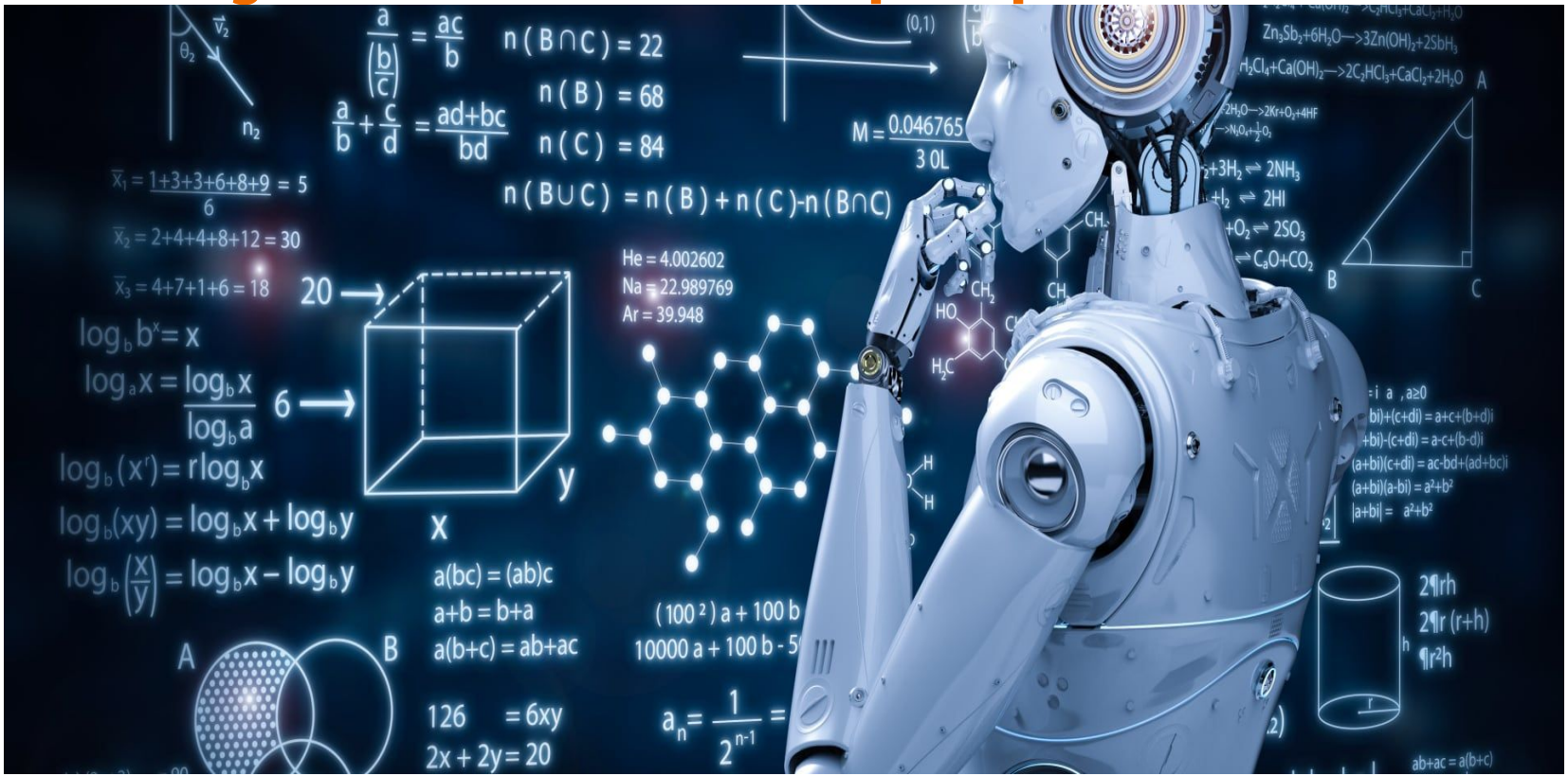
¿Qué es un índice?

- ❑ Una estructura que le permite ubicar filas en una tabla rápidamente, usando un valor indexado
- ❑ Una tabla virtual basada en el conjunto de resultados de una consulta
- ❑ Una estructura que comprende filas y columnas que usa para almacenar datos

# Algoritmos: serie de pasos o reglas para realizar tareas específicas como clasificación o regresión



# Inteligencia artificial, conceptos previos



# Aprendizaje Automático

Machine Learning




Observaciones
(datos)



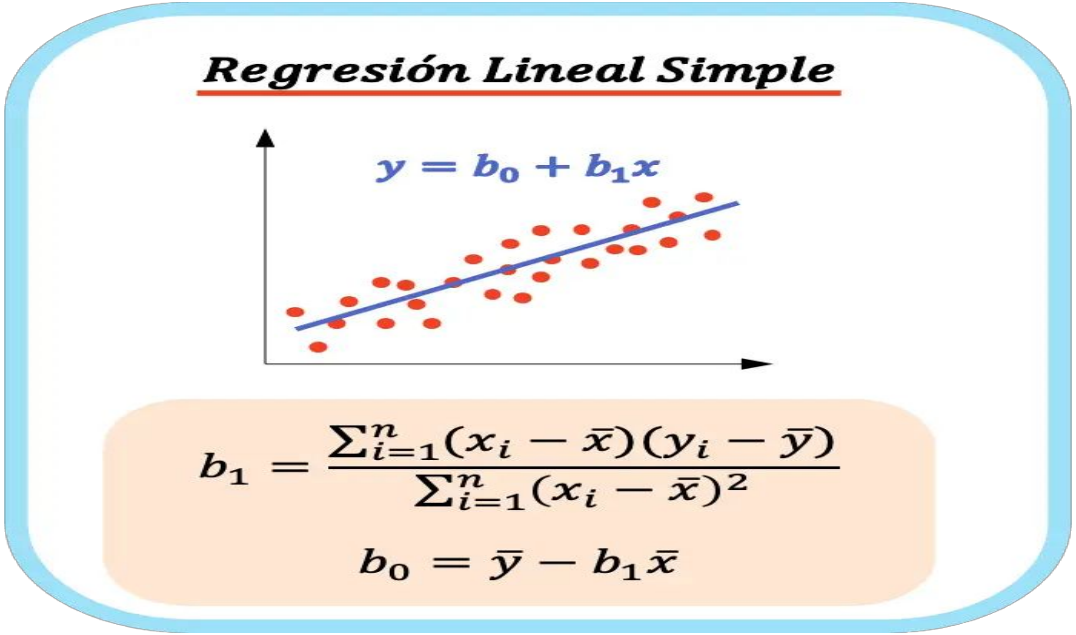
Algoritmos



Aprendizaje

# Regresión Lineal Simple

**En la regresión lineal simple, se predice el valor de la variable dependiente a partir del valor de la variable independiente.**

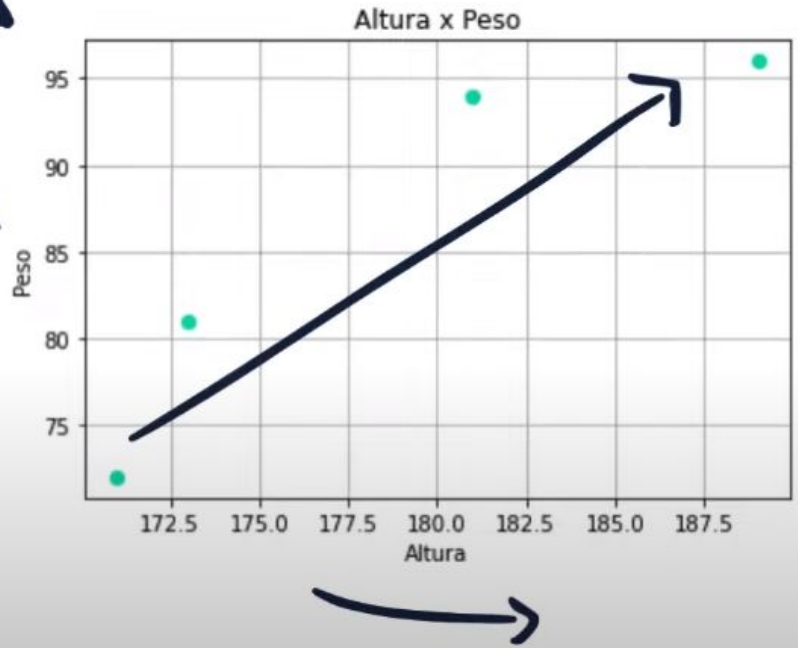


# ¿Qué es la regresión lineal?

El análisis de regresión lineal se utiliza para predecir el valor de una variable basándose en el valor de otra variable. La variable que desea predecir se denomina variable dependiente. La variable que está utilizando para predecir el valor de la otra variable se denomina variable independiente.

$$y = \theta_0 + \theta_1 x$$

| Nombre | Altura (cm) | Peso (kg) |
|---|---|---|
| Antonio | 173 | 81 |
| José | 171 | 72 |
| Carlos | 189 | 96 |
| Enrique | 181 | 94 |
| David | 179 | ¿? |



| Nombre | Altura (cm) | Peso (kg) |
| --- | --- | --- |
| David | 179 | ¿? |



## Regresión Lineal Simple

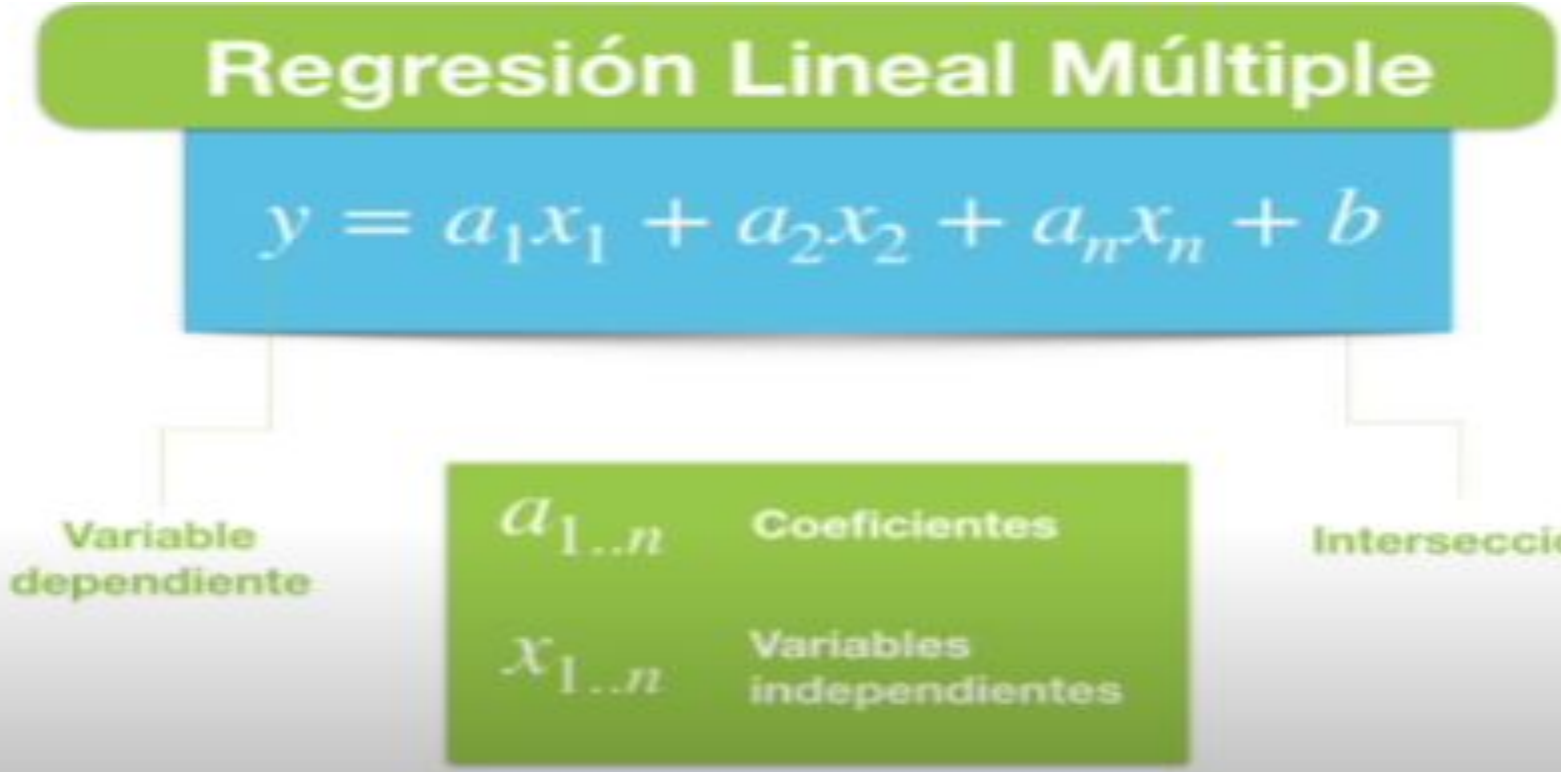
Utilizamos una variable independiente para predecir una variable dependiente

## Regresión Lineal Múltiple

Se manejan múltiples variables independientes que contribuyen a la variable dependiente

Se manejan múltiples coeficientes y es más compleja debido a las variables añadidas

# REGRESEION LINEAL MULTIPLE



# Regresión Lineal Múltiple

NO incluimos todas las variables independientes a la vez y posteriormente comenzamos a minimizar la función de error

Debemos construir una matriz de correlación para todas las variables independiente e incluimos la variable dependiente

Se debe enfocar en seleccionar las mejores variables independientes que puedan contribuir a la variable dependiente

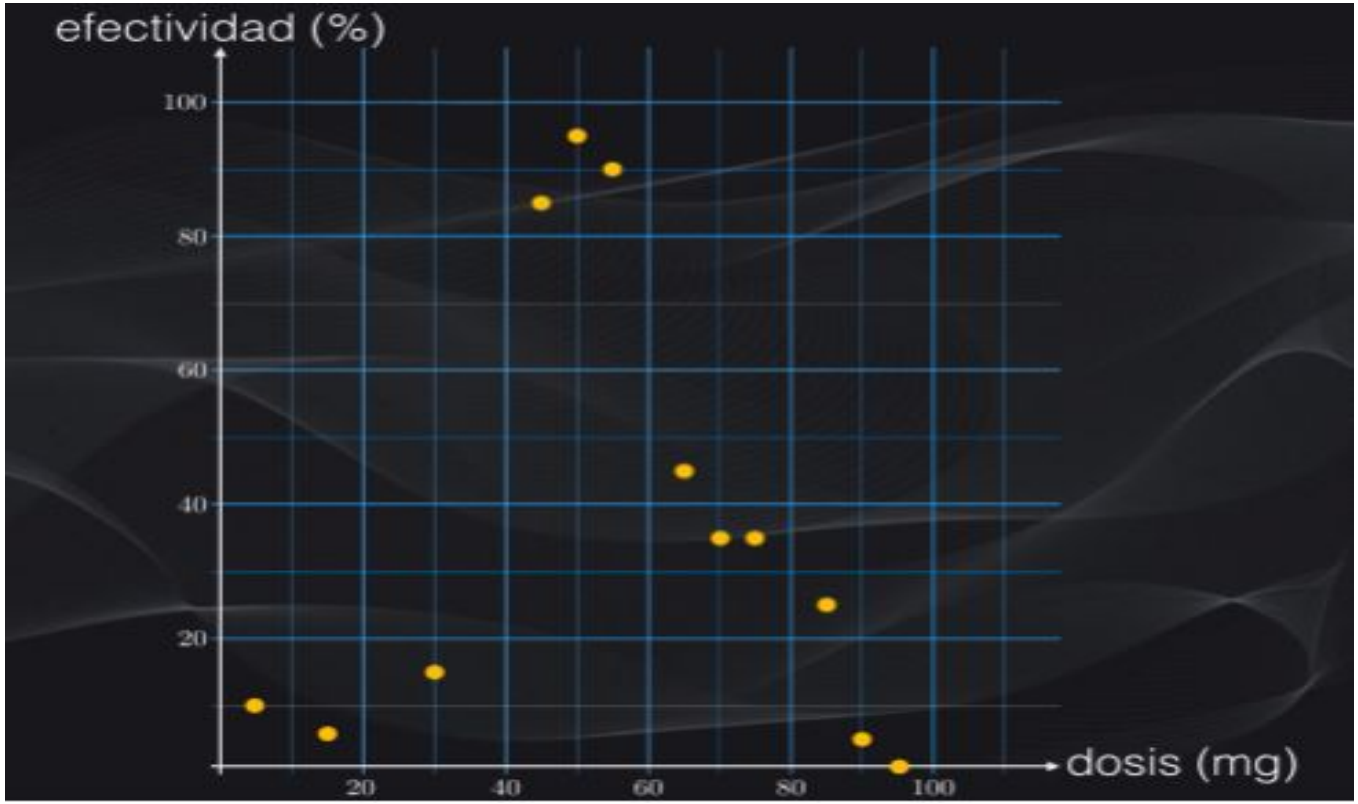
# Regresión Lineal Múltiple

Agregar más variables independientes no significa que la regresión sea mejor u ofrece mejores predicciones

Cuando se agrega más variables independientes se crean relaciones entre ellas

Agregar más variables independientes puede empeorar las cosas

# BOSQUES ALEATORIOS: Clasificación y Regresión



# Árbol de decisión

Un árbol de decisión es una representación gráfica de un proceso de toma de decisiones que se utiliza en una variedad de disciplinas, desde la inteligencia artificial hasta las elecciones empresariales y médicas.

**Para qué se utiliza un árbol de decisión**

Los árboles de decisión tienen un gran abanico de aplicaciones en diversas áreas. Su versatilidad radica en su capacidad para resolver problemas de clasificación y regresión.

# Características y componentes de un árbol de decisión

- **Nodo raíz:** es el punto de partida del árbol y representa la pregunta inicial o la condición que se debe evaluar para tomar una decisión.

- **Nodos internos:** estos nodos son preguntas adicionales o condiciones que se evalúan en el proceso de toma de decisiones.

- **Ramas:** las ramas conectan los nodos y hacen referencia a las posibles respuestas o resultados de las condiciones evaluadas en los nodos anteriores.

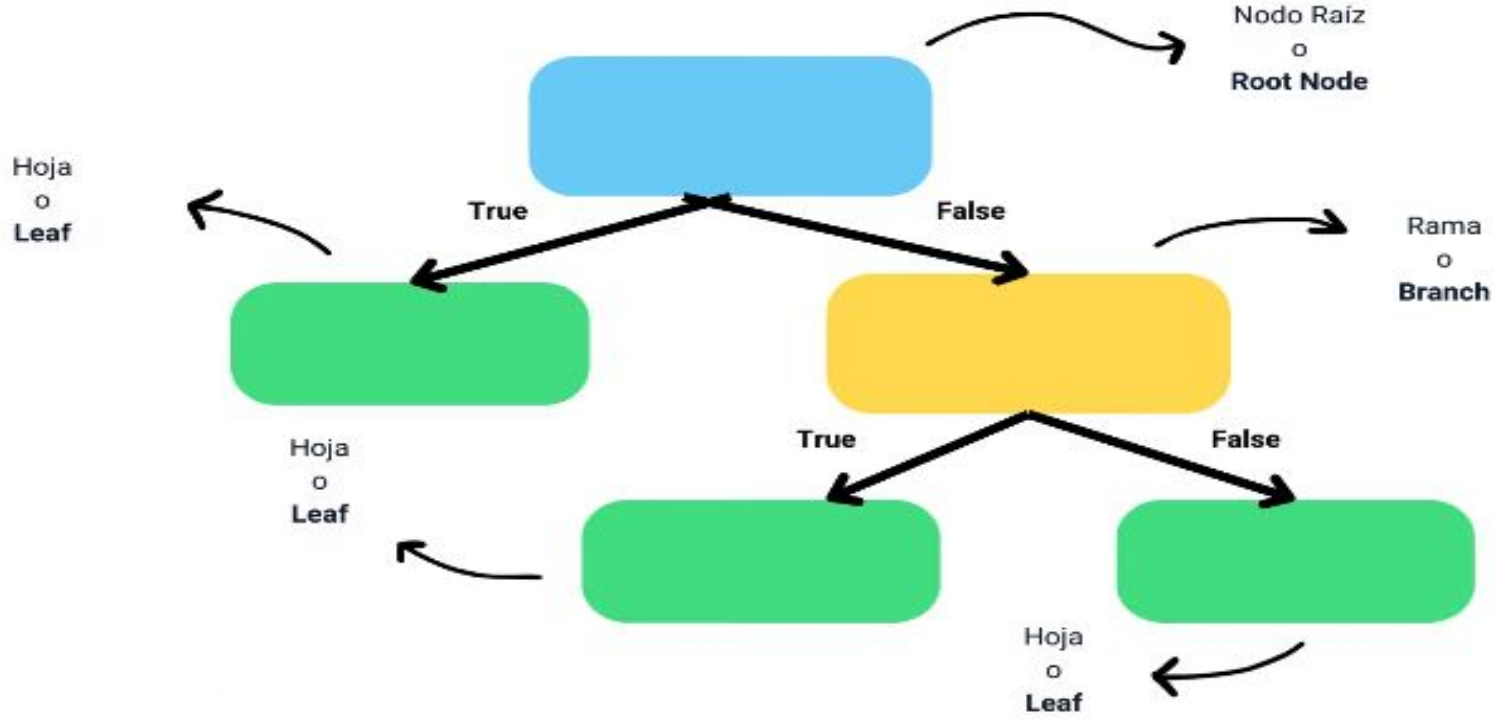
- **Hojas:** las hojas del árbol muestran las decisiones finales o las clasificaciones

# Cómo se construye un árbol de decisión

- **selección de la característica :** En cada nodo, se selecciona la característica que se utilizará para dividir los datos. Esta elección se basa en criterios como la ganancia de información o la reducción de la impureza.
- **División de los datos:** Los datos se dividen en subconjuntos en función de los valores de la característica seleccionada. Cada subconjunto se asigna a un nodo hijo.
- **Evaluación de la pureza:** Se calcula la pureza o impureza de cada nodo hijo utilizando una métrica como la entropía o el índice gini. El objetivo es maximizar la pureza en cada nodo.
- **Criterio de parada:** Se establece un criterio de parada para determinar cuándo detener la subdivisión del árbol, lo que puede incluir la profundidad máxima del árbol o el número mínimo de muestras en un nodo.

# Árbol de decisión

Un árbol de decisión, hace una afirmación y toma una decisión en base a si esa afirmación es verdadera o falsa.



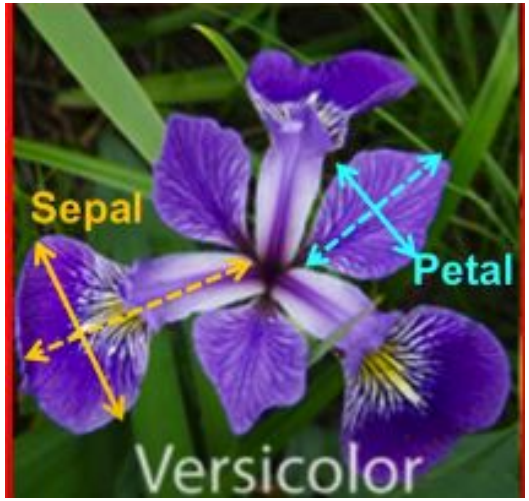
# Cómo se construye un árbol de decisión

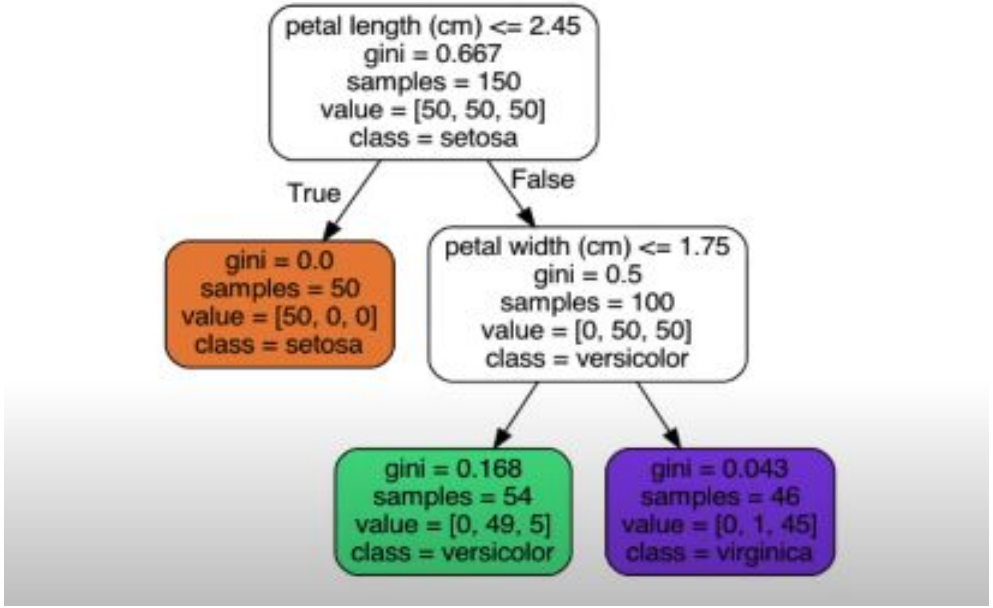
$$\text{Índice Gini} = 1 - (\text{probabilidad categoría 1})^2 - (\text{probabilidad categoría 2})^2$$


lema
liaco

# Ejemplo árbol de decisión

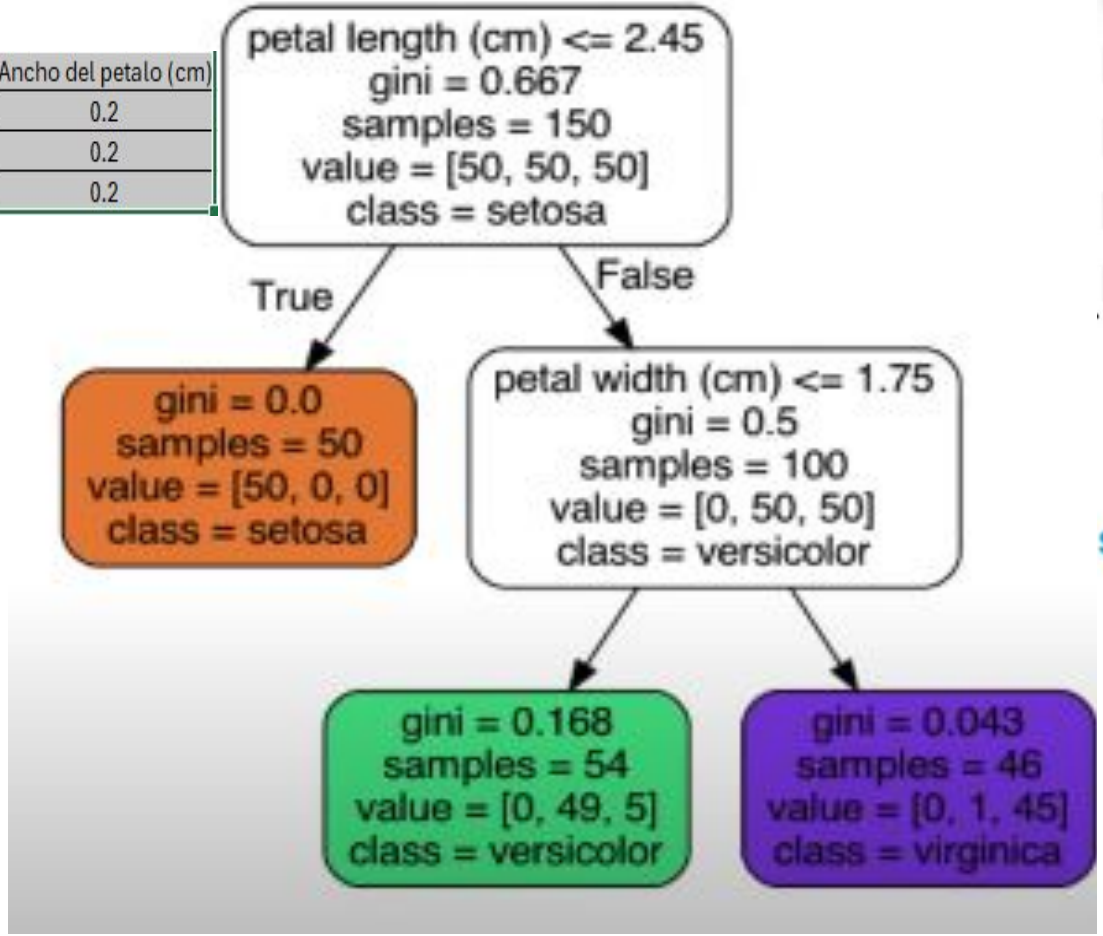
Arbol de decisión entrenado para clasificar flores deacuardo con las dimensiones de sus petalos.



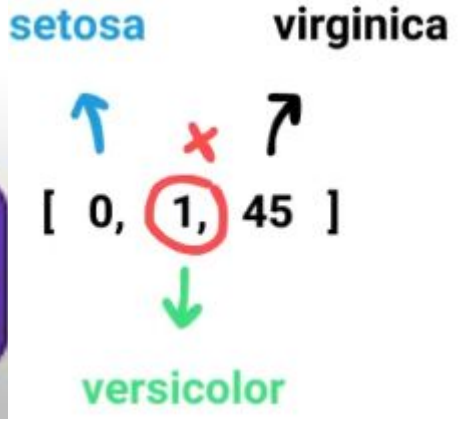



# Ejemplo árbol de decisión

| | Largo del petalo (cm) | Ancho del petalo (cm) |
|---|---|---|
| 0 | 1.4 | 0.2 |
| 1 | 1.4 | 0.2 |
| 2 | 3.7 | 0.2 |

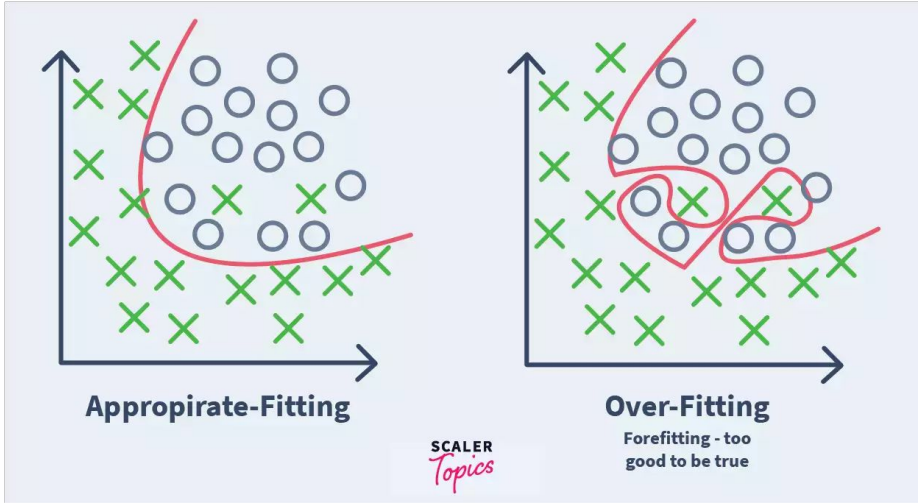



| | Petal length (cm) | Petal width (cm) |
|---|---|---|
| 0 | 1.4 | 0.2 |
| 1 | 1.4 | 0.2 |
| 2 | 1.3 | 0.2 |
| 3 | 1.5 | 0.2 |
| 4 | 1.4 | 0.2 |
| ... | ... | ... |
| 145 | 5.2 | 2.3 |
| 146 | 5.0 | 1.9 |
| 147 | 5.2 | 2.0 |
| 148 | 5.4 | 2.3 |
| 149 | 5.1 | 1.8 |

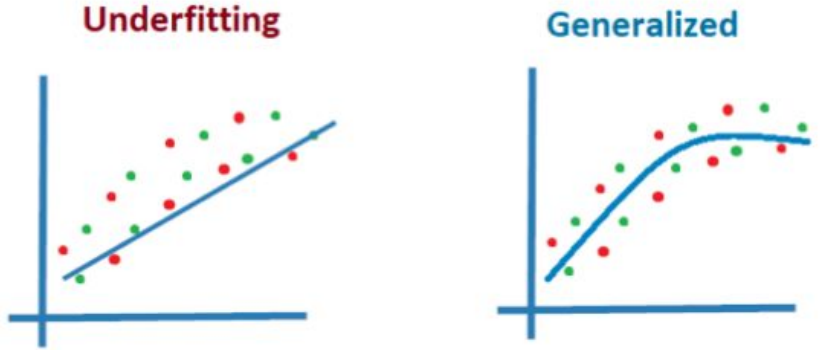


# Tipos de tendencias en los modelos de machine learning

“overfitting” o “sobreajuste”



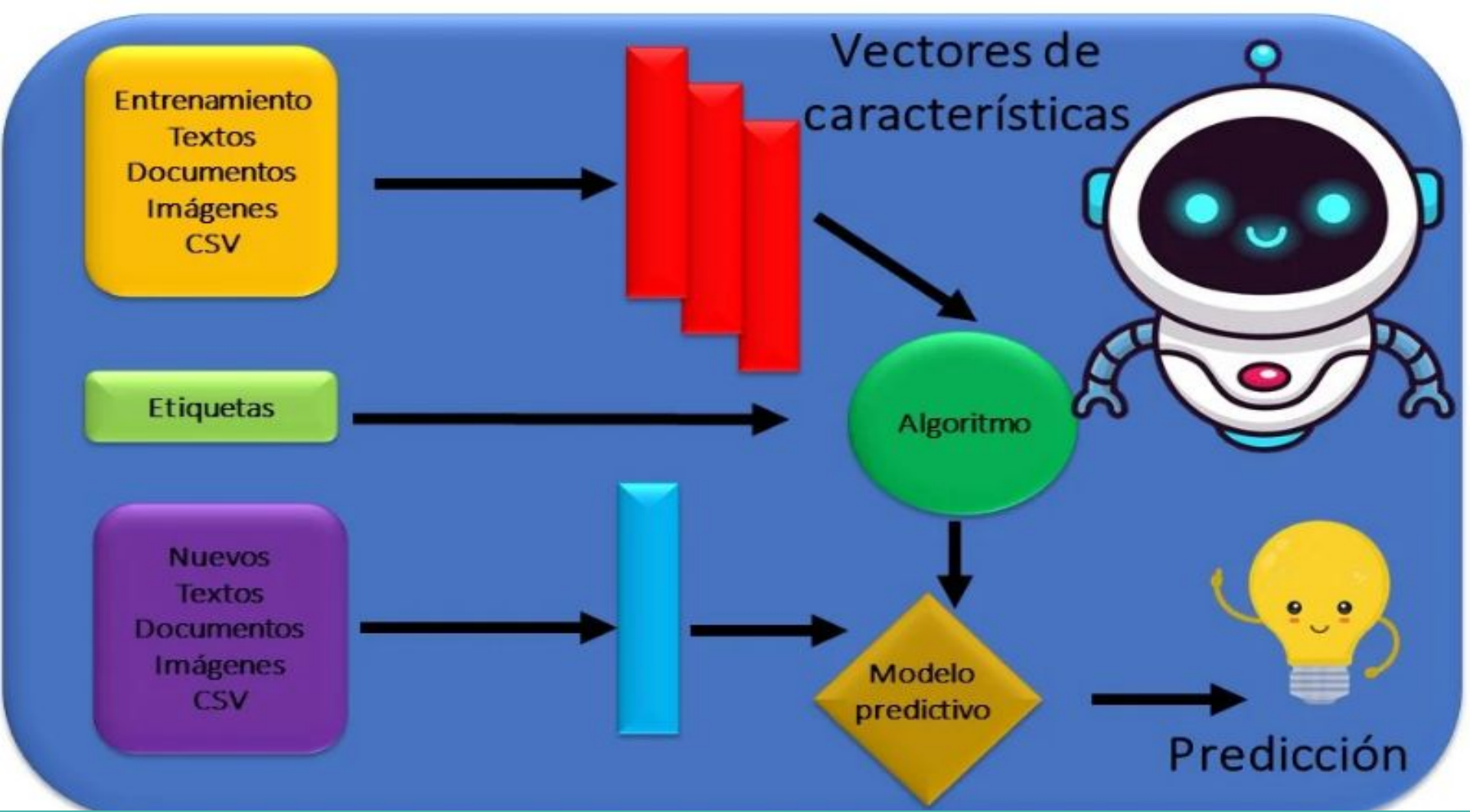

“underfitting” o “subajustado”



# Aprendizaje Supervisado.

El aprendizaje supervisado es un método de aprendizaje en cual se utilizan una serie de datos etiquetados o modelos que queremos que la máquina o el sistema aprenda. Para luego dar una respuesta acertada. Es una parte del campo de aplicación de la inteligencia artificial.




Entrenamiento
Textos
Documentos
Imágenes
CSV
Vectores de características
Etiquetas
Algoritmo
Nuevos
Textos
Documentos
Imágenes
CSV
Modelo predictivo
Predicción

| Ojo izquierdo | Ojo derecho | Orientación de boca | Lengua | Color | | Respuesta |
|---|---|---|---|---|---|---|
| Cerrado | Cerrado | Convexa | Visible | Amarilla |  | Positiva |
| Abierto | Abierto | Cóncava | Oculta | Amarilla |  | Negativa |
| Abierto | Abierto | Plana | Oculta | Verde |  | Negativa |
| Abierto | Cerrado | Plana | Visible | Amarilla |  | Positiva |
| Cerrado | Abierto | Cóncava | Visible | Amarilla |  | Positiva |
| Abierto | Abierto | Cóncava | Visible | Roja |  | Negativa |

# Tipos de aprendizaje supervisado

En general, los algoritmos que se emplean en el aprendizaje supervisado establecen modelos de relación y dependencia. Específicamente entre los resultados de la predicción y los datos de entrada.

1. **Por regresión:** Se basa fundamentalmente en la predicción de variables numéricas.

2. **Por clasificación :** Se basa fundamentalmente en la predicción de variables categóricas.

# Tipos de aprendizaje supervisado

## Por regresión

1

- Se basa fundamentalmente en la predicción de variables numéricas
- Estos algoritmos se emplean para hacer una predicción de:
  - -Precios
  - -La cantidad de la demanda
  - -El nivel de ventas
  - -El nivel de ingresos, entre algunos que se pueden mencionar.



## Por clasificación

2

- Se basa fundamentalmente en la predicción de variables categóricas
- Estos algoritmos ayudan a hacer predicciones sobre:
  - -El riesgo del crédito, si el cliente pagará o no pagará
  - -Predecir si un alumno aprobará un curso o no
  - -Para clasificar imágenes





# Aprendizaje supervisado

# Ejemplos

1. Detección de spam de correo electrónico
2. Reconocimiento de voz
3. Clasificación de noticias
6. Análisis de sentimiento

COLOMBIA VIDA

TIC

Este enfoque permite analizar datos complejos



permite a los sistemas

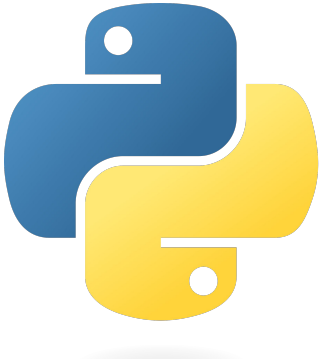
# ¿Qué es Python?

Python es un lenguaje de programación de propósito general muy poderoso y flexible, a la vez que sencillo y fácil de aprender.

Es un lenguaje de alto nivel, que permite procesar fácilmente todo tipo de estructuras de datos, tanto numéricas como de texto.

# Instalación y manejo de python







# Entrada y salida de datos con Python

La entrada y salida de datos en un programa, son conceptos esenciales para un desarrollador. Estos datos de entrada deben ser procesados para generar los resultados esperados, los cuales deben entregarse en un formato especificado por el usuario para que sean útiles en un sistema de información.

# Librerías utilizadas

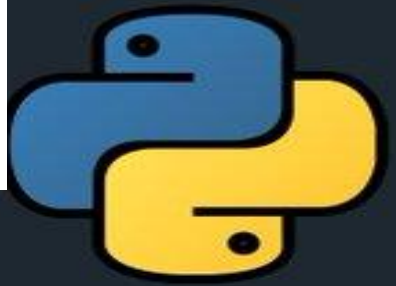
La entrada y salida de datos en un programa, son conceptos esenciales para un desarrollador. Estos datos de entrada deben ser procesados para generar los resultados esperados, los cuales deben entregarse en un formato especificado por el usuario para que sean útiles en un sistema de información.

# ciencia de datos

## LIBRERIAS DE PYTHON PARA CIENCIA DE DATOS




NEURAL CODERS





NumPy

NumPy es una extensión de Python, que le agrega mayor soporte para vectores y matrices, constituyendo una biblioteca de funciones matemáticas.



SciPy

Es una librería de computación científica que incorpora modelos de optimización, algebra lineal e ingeniería.



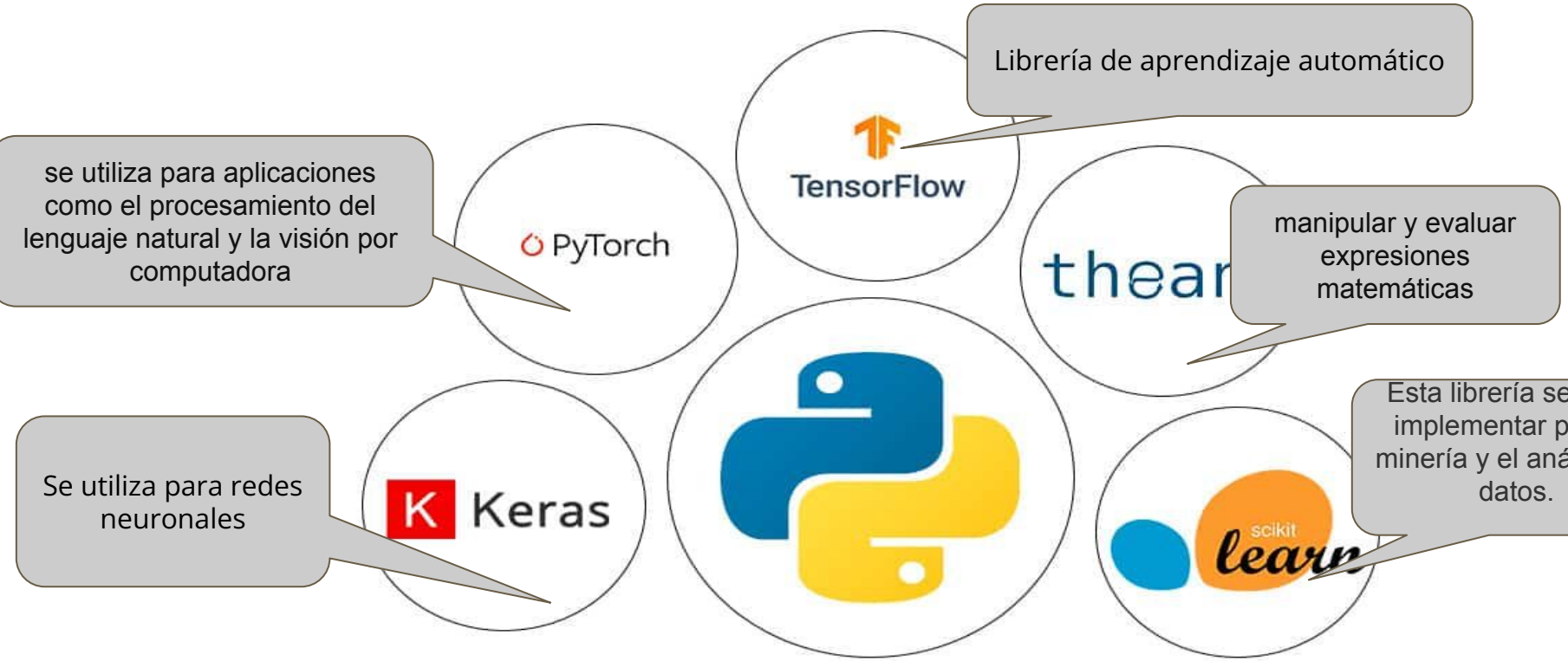
Pandas

Es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python.

# PLOTTING Y VISUALIZACION



# MACHINE LEARNING



# INPLACE

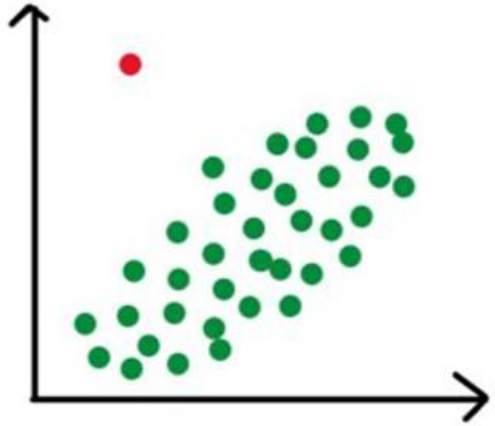
Inplace es un argumento utilizado en diferentes funciones. Algunas funciones en las que inplace se usa como atributos como set_index(), dropna() , fillna() , reset_index() , drop(), replace() y muchas más. El valor predeterminado de este atributo es False y devuelve la copia del objeto.

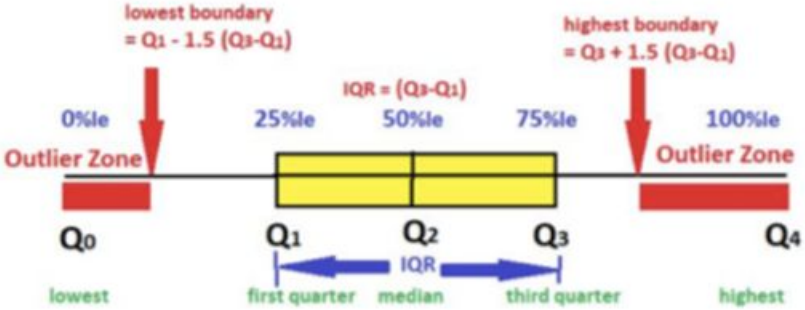
# FILLNA

permite sustituir los valores nulos de una estructura pandas por otro valor según ciertos criterios

# Outliers

- Son observaciones distantes del resto.
- Puede deberse a variabilidad propia de los datos o a errores de medida.
- No hay una definición matemática de que es un outlier pero hay 2 métodos para identificarlos:
  - 3* STD
  - + 1.5 IQR





# Missing Values




- Son atributos para los cuales no hay información en una observación.
- Cómo tratarlos?
  - Eliminarlos (Sencillo pero perdemos info)
  - Sustituir por valor estadístico (Media, mediana..)
  - Emplear un modelo para rellenar esos valores.

# ANALISIS DE REGRESION LINEAL

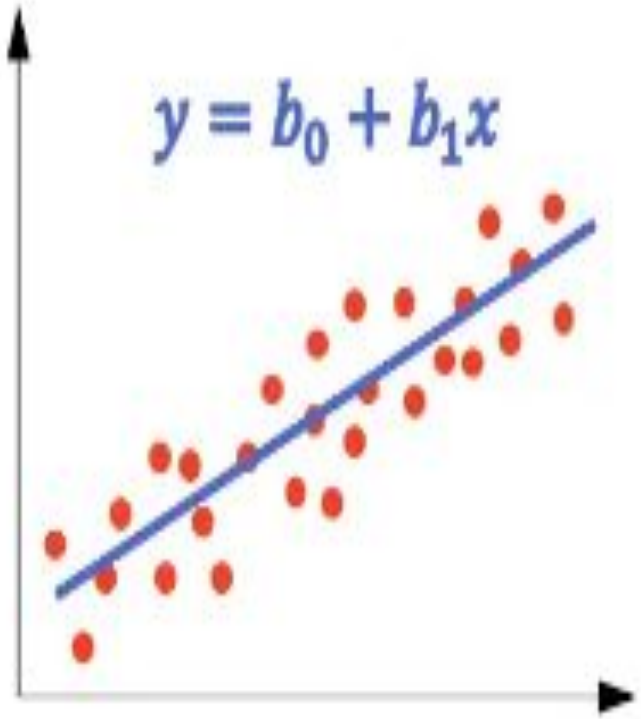
La regresión lineal es un modelo matemático que describe la relación entre varias variables.

La tarea de la regresión en el aprendizaje automático consiste en predecir un parámetro (Y) a partir de un parámetro conocido X.

Dependiendo de los objetivos del estudio, puedes elegir entre diversos tipos de análisis de regresión:

En una regresión lineal, se trata de establecer una relación entre una variable independiente y su correspondiente variable dependiente. Esta relación se expresa como una línea recta. No es posible trazar una línea recta que pase por todos los puntos de un gráfico si estos se encuentran ordenados de manera caótica. Por lo tanto, sólo se determina la ubicación óptima de esta línea mediante una regresión lineal.

# Regresión Lineal Simple




## ¿bo y b1?

$$\hat{y} = b_0 + b_1 X$$

variable

variable

coeficiente constante no varía un solo valor

INTERCEPTO

coeficiente constante no varía un solo valor

PENDIENTE

# Módulo matplotlib.pyplot

matplotlib.pyplot es un módulo de la librería Python Matplotlib que permite crear gráficos de forma sencilla e interactiva. Este módulo ofrece una interfaz fácil de usar que automatiza la creación de ejes y figuras al hacer un gráfico.

**Gráfica de dispersión :** Una gráfica de dispersión es un tipo de diagrama matemático que muestra la relación entre dos variables numéricas, colocando una variable en el eje x y la otra en el eje y.

Ejemplo.

Ejercicio.

# Funciones para trabajar graficos de dispersion

`scatter()` Este código utiliza la biblioteca matplotlib de python para crear un gráfico de dispersión donde el primer argumento es la variable del eje x y el segundo argumento es la variable del eje y.

`plot(x, y)` Dibuja un polígono con los vértices dados por las coordenadas de la lista `x` en el eje X y las coordenadas de la lista `y` en el eje Y.

# Gráficos de dispersión con una línea de tendencia

Este método nos dará un ajuste polinómico por mínimos cuadrados donde el primer argumento es la variable x, la segunda variable es la variable y, y la tercera variable son los grados del ajuste (1 para lineal).

Mínimos cuadrados: Este método consiste en calcular la recta que más se aproxime a los puntos de un plano,determinando cual es la pendiente de la recta y el intercepto con el eje Y permitiendo así predecir puntos en el fenómeno que se está estudiando.

$$y = mx + b$$

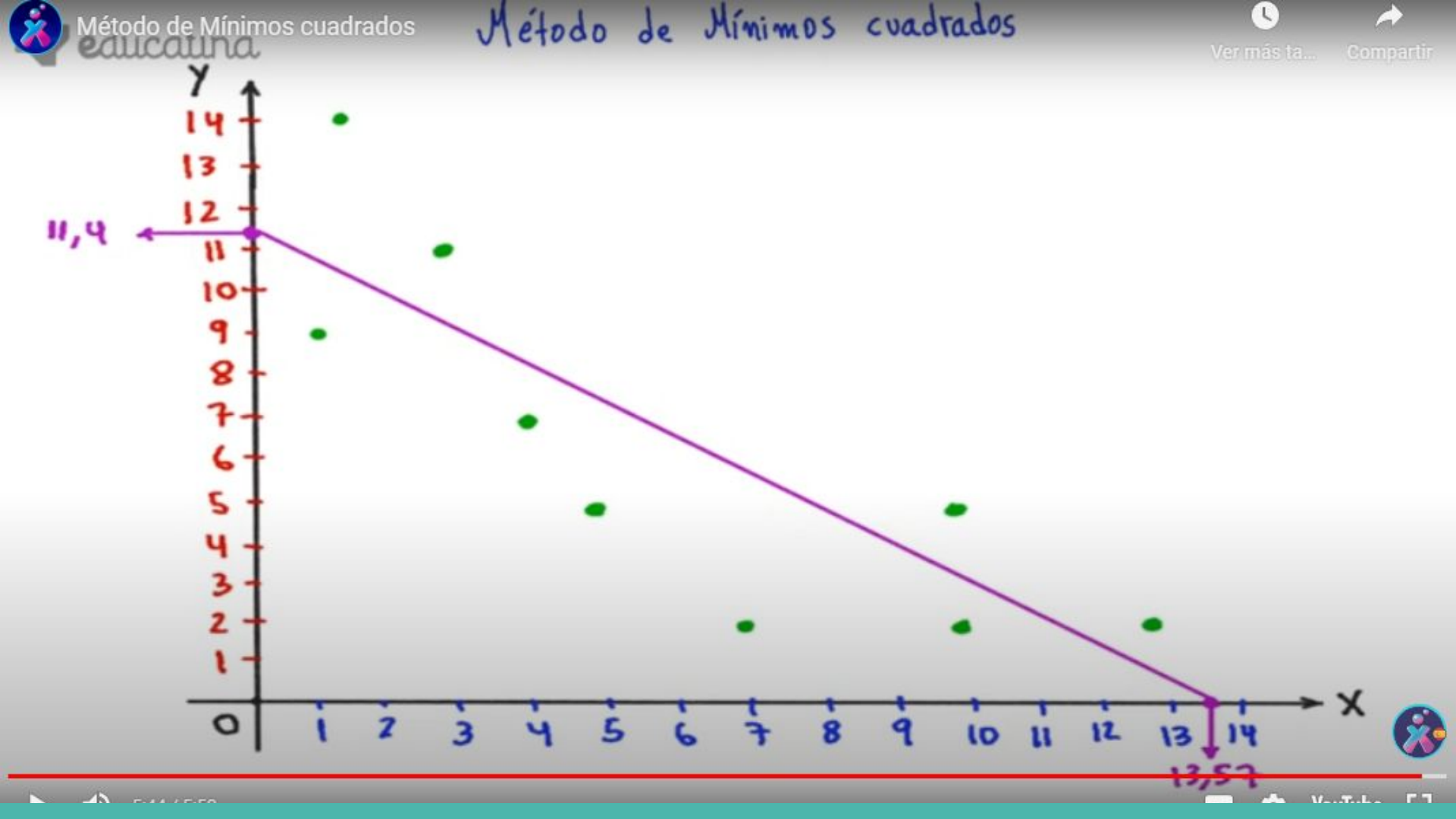
# Ejercicio: Encontrar la recta que mejor se ajuste a los siguientes datos

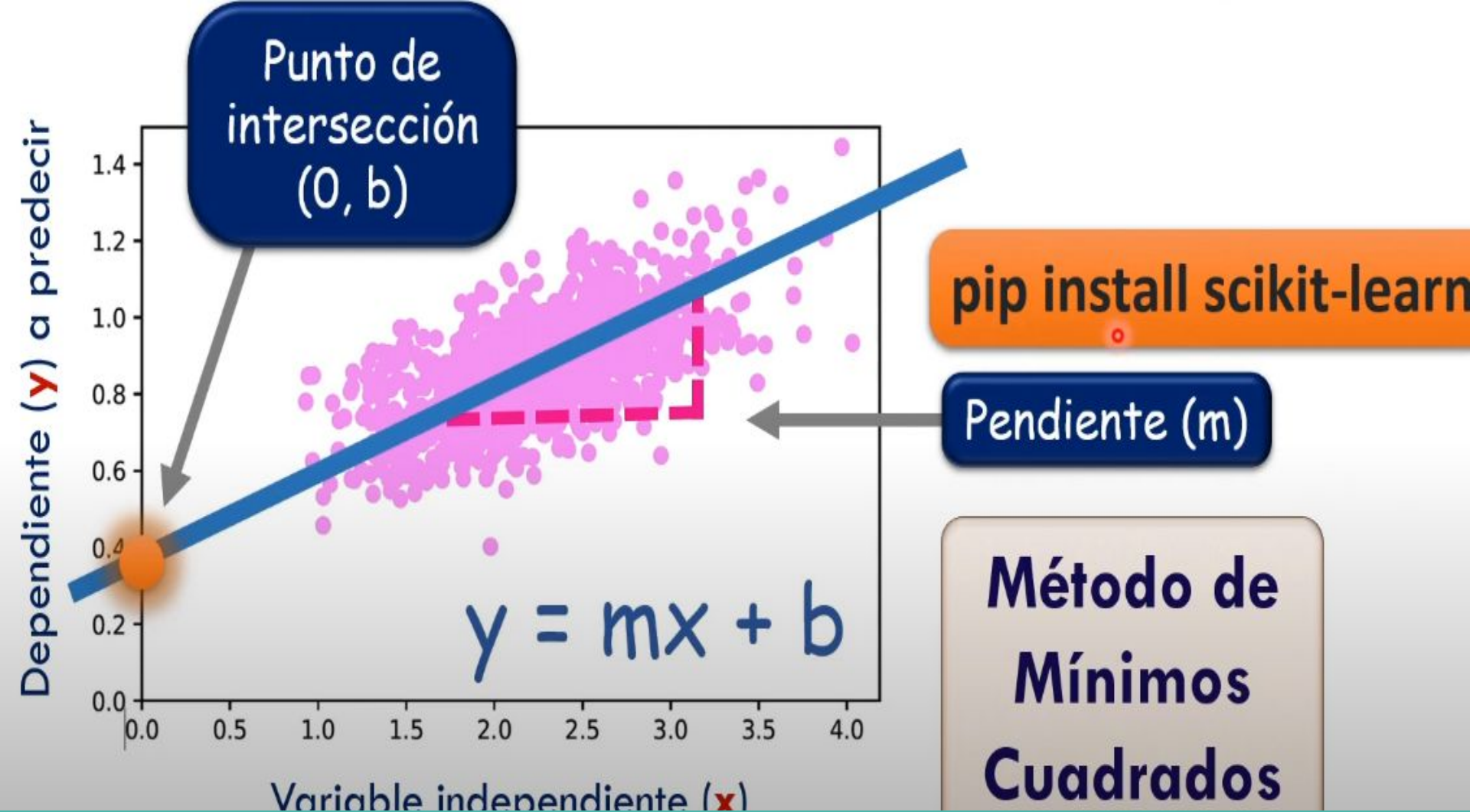
| x | y | x.y | $x^2$ |
|---|---|---|---|
| 7 | 2 | | |
| 1 | 9 | | |
| 10 | 2 | | |
| 5 | 5 | | |
| 4 | 7 | | |
| 3 | 11 | | |
| 13 | 2 | | |
| 10 | 5 | | |
| 2 | 14 | | |
| | | | |

## Ecuaciones

$$m = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$b = \bar{y} - m\bar{x}$$


Método de Mínimos cuadrados
Método de Mínimos cuadrados
Ver más ta...
Compartir
Y
X
0
11,4
13,57


Punto de intersección (0, b)
pip install scikit-learn
Pendiente (m)
Método de Mínimos Cuadrados
y = mx + b
Dependiente (y) a predecir
Variable independiente (x)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

# LIBRERIA SCIKIT LEARN

scikit-learn es uno de los open-source y bibliotecas de aprendizaje automático más populares en Python. La biblioteca scikit-learn contiene muchas herramientas eficientes para aprendizaje automático y modelado estadístico, incluyendo clasificación, regresión, agrupación, y reducción de dimensionalidad

Instalación librerías scikitlear

```
pip install --upgrade scikit-learn
```

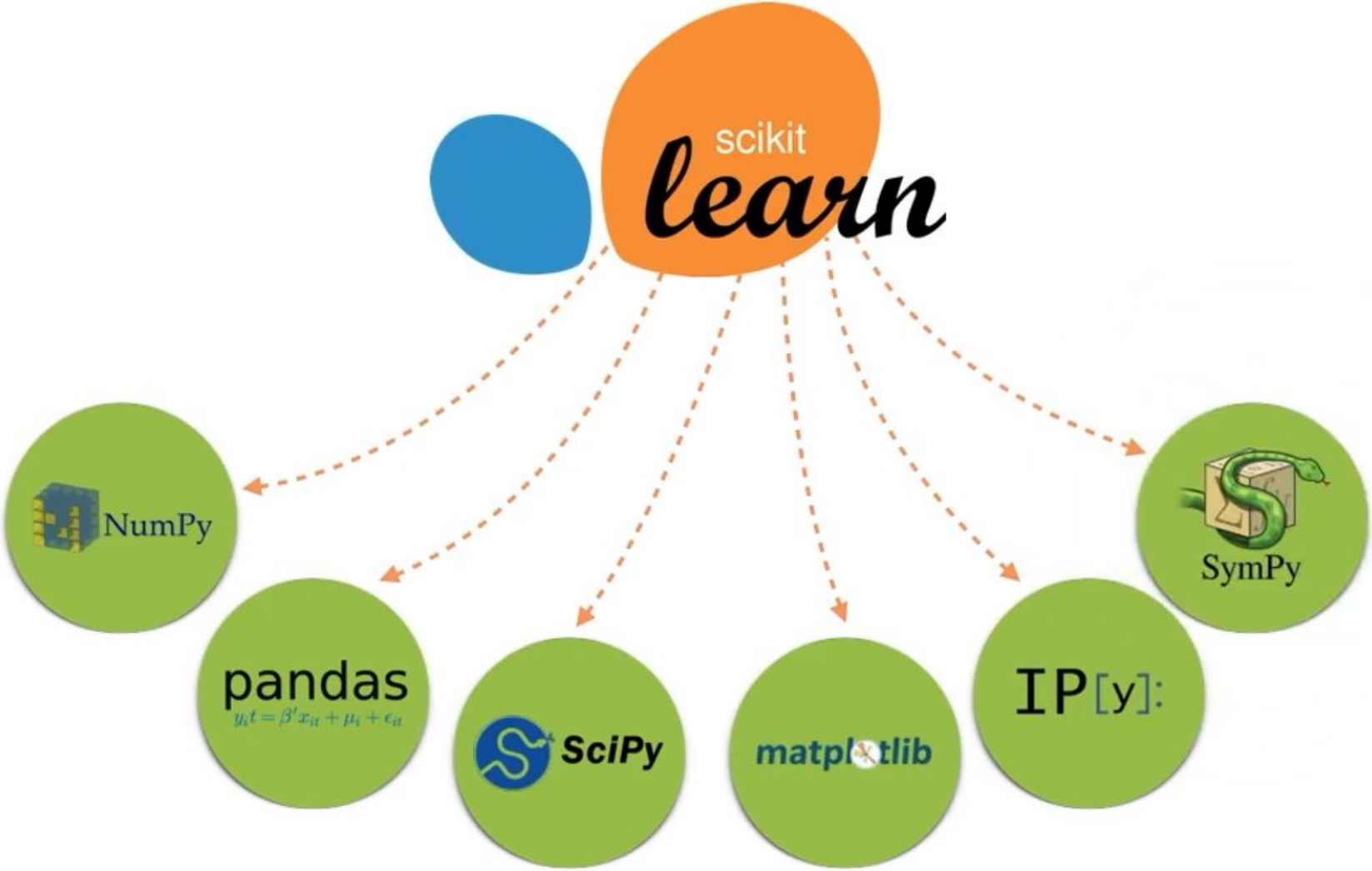
```
linearRegresion : Crea un modelo de regresión lineal
```

El método `reg.fit` en aprendizaje automático se utiliza para entrenar un modelo con datos de entrenamiento. Específicamente, ajusta el modelo a los datos




Librería de código abierto

Proporciona una gama de algoritmos de aprendizaje supervisados y no supervisados


scikit
learn
NumPy
pandas
$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$
SciPy
matplotlib
IP[y]:
SymPy

# Error absoluto máximo (M)

$y$ : valor verdadero

$\hat{y}$ : valor predicho

$$\mathrm{M}(y, \hat{y}) = max(|y_i - \hat{y}_i|)$$

# Error absoluto medio (mean absolute error - MAE)

$y$ : valor verdadero

$\hat{y}$ : valor predicho

$n$ : tamaño de la muestra

$$\mathrm{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

# Error cuadrático medio (mean squared error - MSE)

$y$ : valor verdadero

$\hat{y}$ : valor predicho

$n$ : tamaño de la muestra

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

# R^2 (Coeficiente de determinación)

$y$ : valor verdadero

$\hat{y}$ : valor predicho

$\bar{y}$ : promedio de los valores verdaderos

$n$ : tamaño de la muestra

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

# Regresión lineal múltiple

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_1^{\,2}$$

# ENTRENAMIENTO, VALIDACIÓN Y PRUEBA

## Introducción

Cuando abordamos un proyecto de Machine Learning usualmente debemos elegir entre múltiples modelos y determinar cuál de ellos generará las mejores predicciones.

Y generalmente lo que tenemos que hacer es entrenar múltiples modelos y ponerlos a prueba para determinar cuál de ellos tiene el mejor desempeño, garantizando a la vez que el modelo seleccionado funcione adecuadamente con datos que nunca antes ha visto.

Y para lograr todo esto usualmente debemos usar tres sets de datos: entrenamiento, validación y prueba.

# ENTRENAMIENTO, VALIDACIÓN Y PRUEBA

- set de entrenamiento: Nos permite obtener los parámetros de cada modelo candidato.

- Set de  validación: Nos permite obtener los hiper-parametros más adecuados de cada uno de los modelos candidatos y elegir de todos los modelos que estamos considerando cuál es el de mejor desempeño  para el problema que estamos resolviendo.

- Set de prueba: Colocamos nuestro modelo a prueba con datos nunca antes vistos.

# Evaluación "Tradicional"







# set de validación






set de entrenamiento
set de prueba
set de validación

peso



glucosa
en sangre



MODELO

MODELO



hipertenso

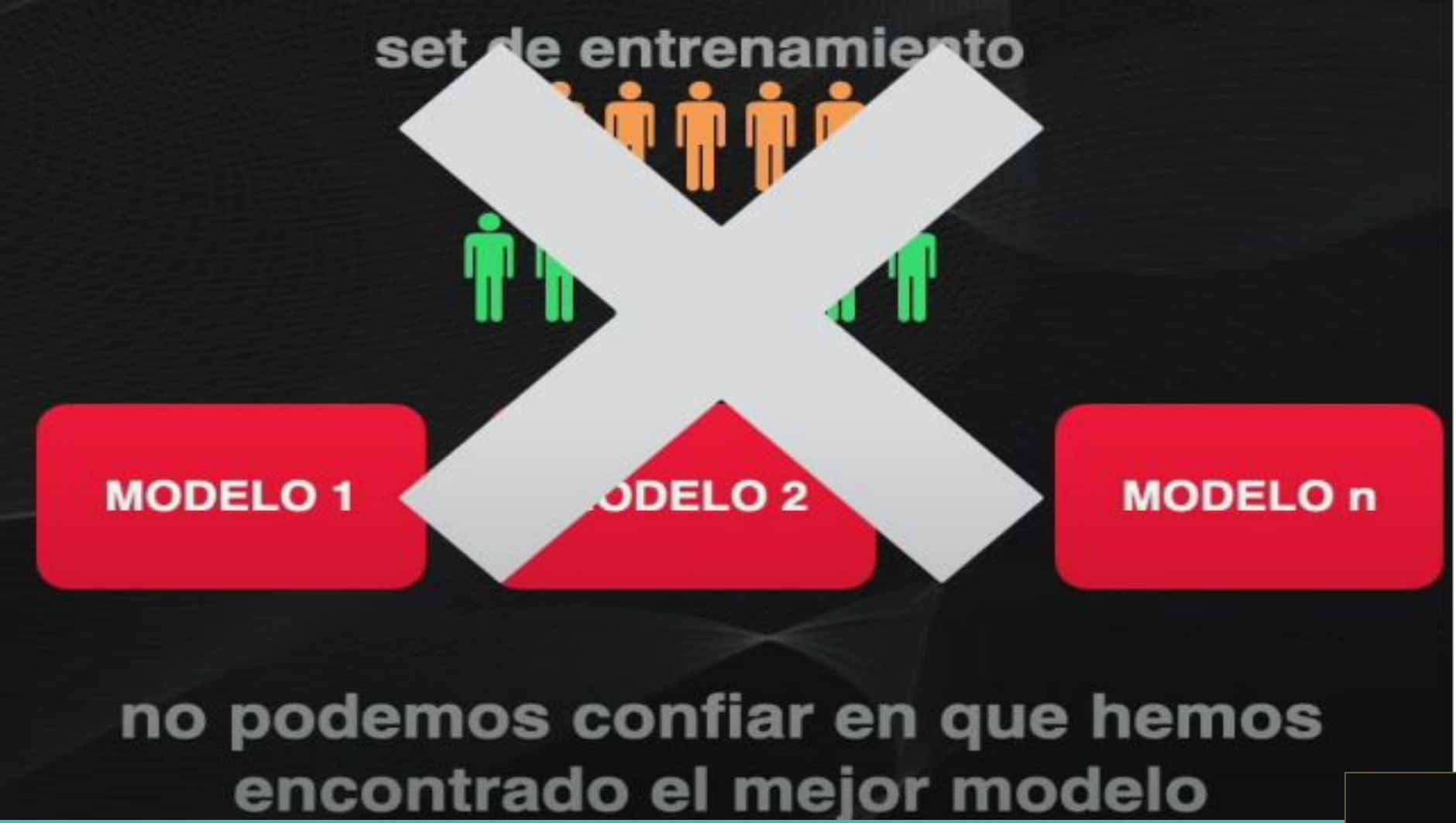


no hipertenso

# ¿cómo afinar los hiperparámetros?




sobre-ajuste (*overfitting*)


set de entrenamiento
MODELO 1
ODELO 2
MODELO n
no podemos confiar en que hemos
encontrado el mejor modelo

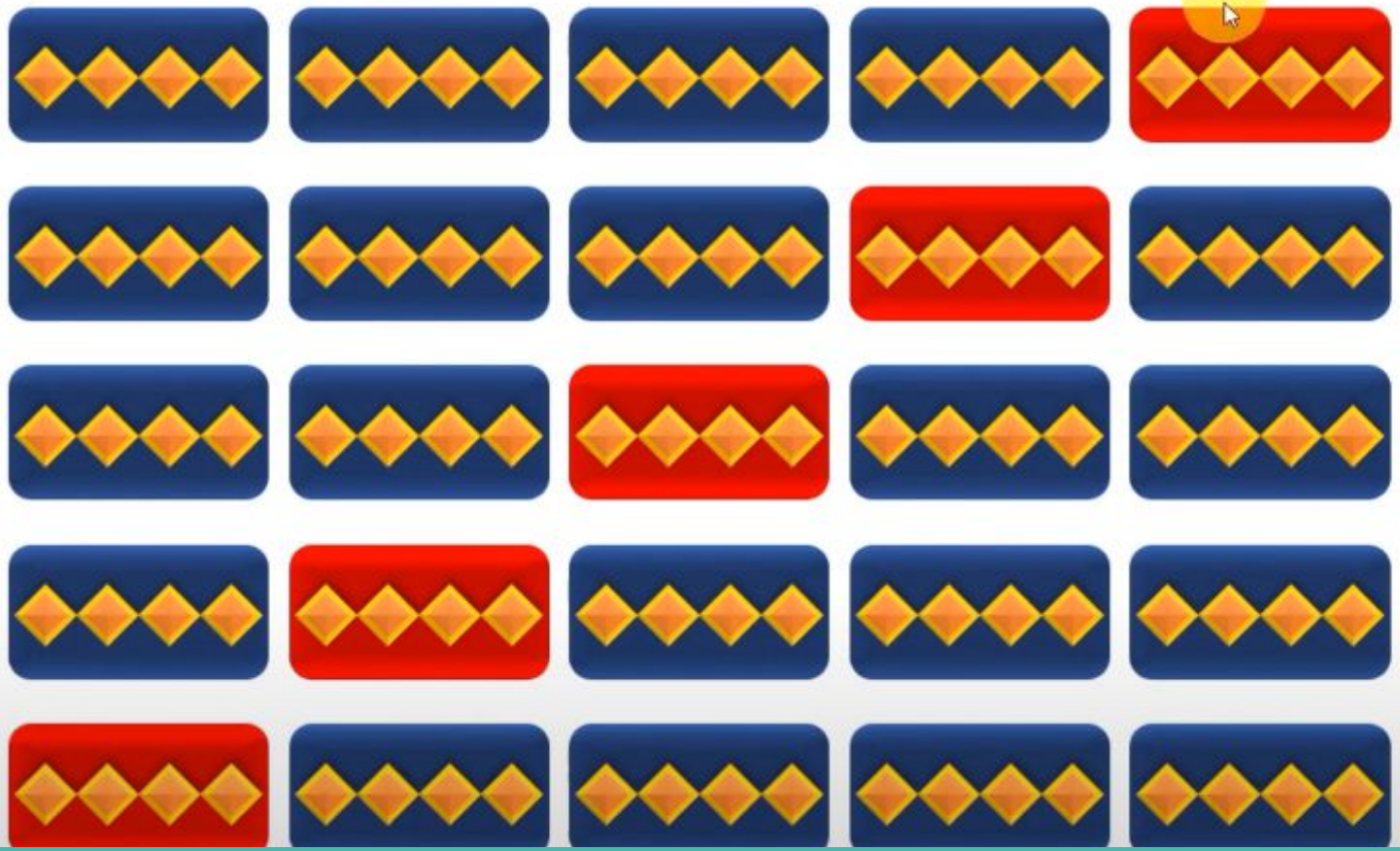
Trabajamos con la librería SKLEARN

Parámetros: Variable numérica interna del modelo que no se puede controlar al programar el algoritmo de entrenamiento y que por el contrario se calcula de forma automática a partir de los datos.

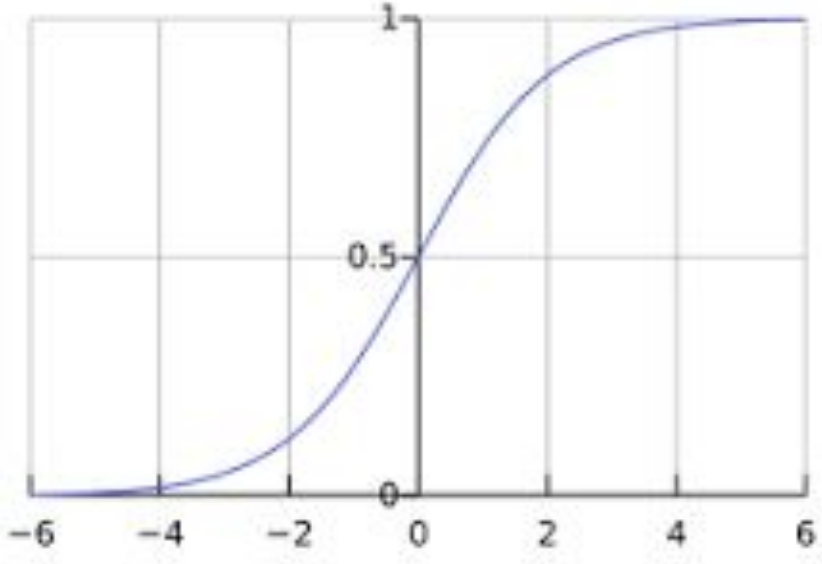
Hiper-parámetros:  Son variables externas al modelo y que se definen manualmente al momento de programar el algoritmo de entrenamiento.

Esta validación es utilizada más que todo en el deeplearning, cuando tenemos muchos datos (miles), cuando hay menos datos se recomienda trabajar con validación cruzada

# Evaluación con Validación Cruzada



# Regresión logística

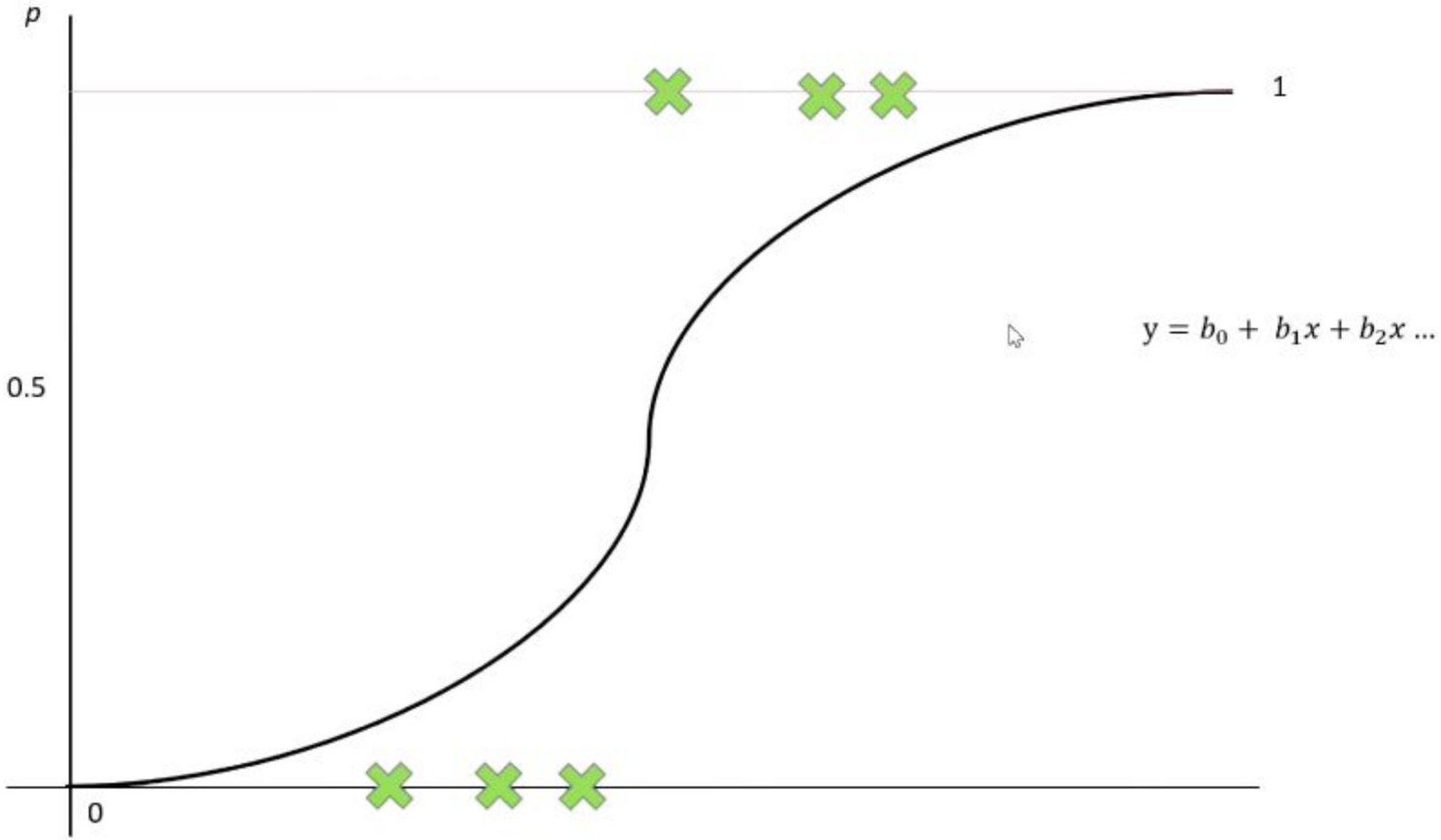



## Descripción

En estadística, la regresión logística es un tipo de análisis de clasificación utilizado para predecir el resultado de una variable categórica en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo en función de otros factores

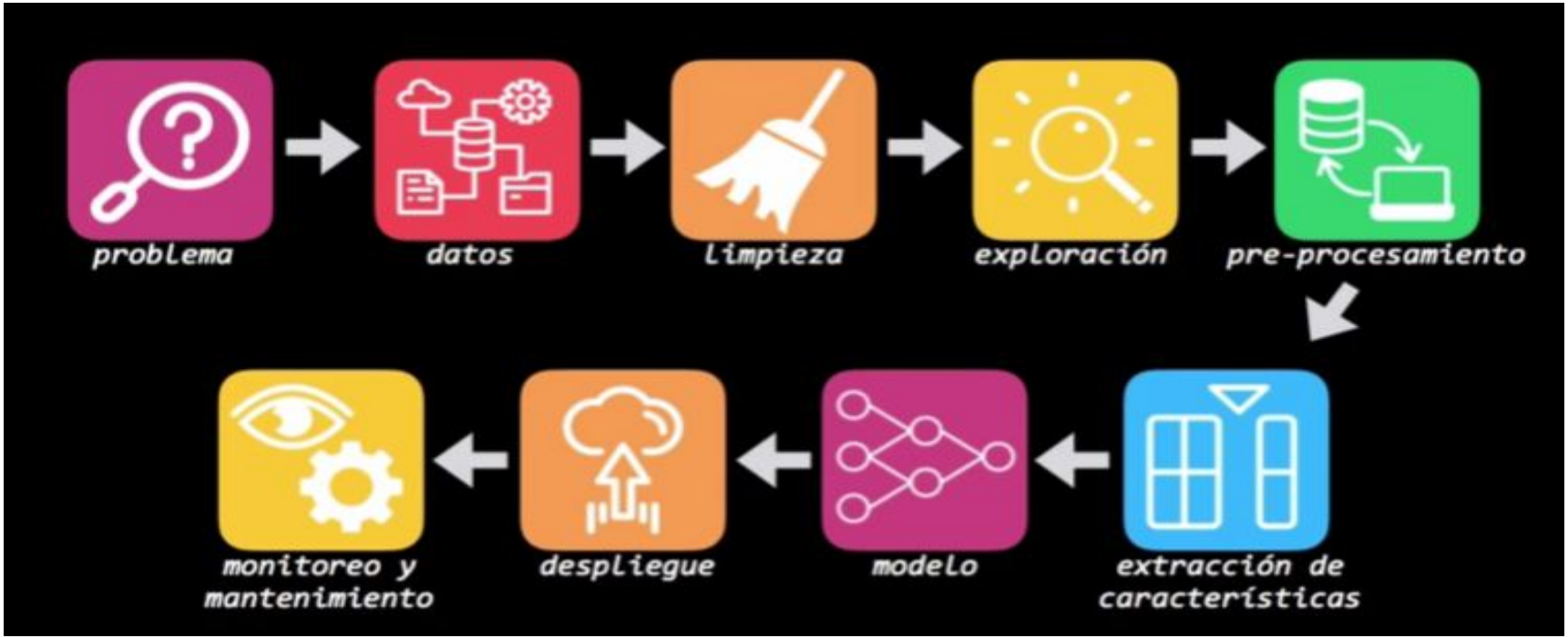
# Qué es la regresión logística

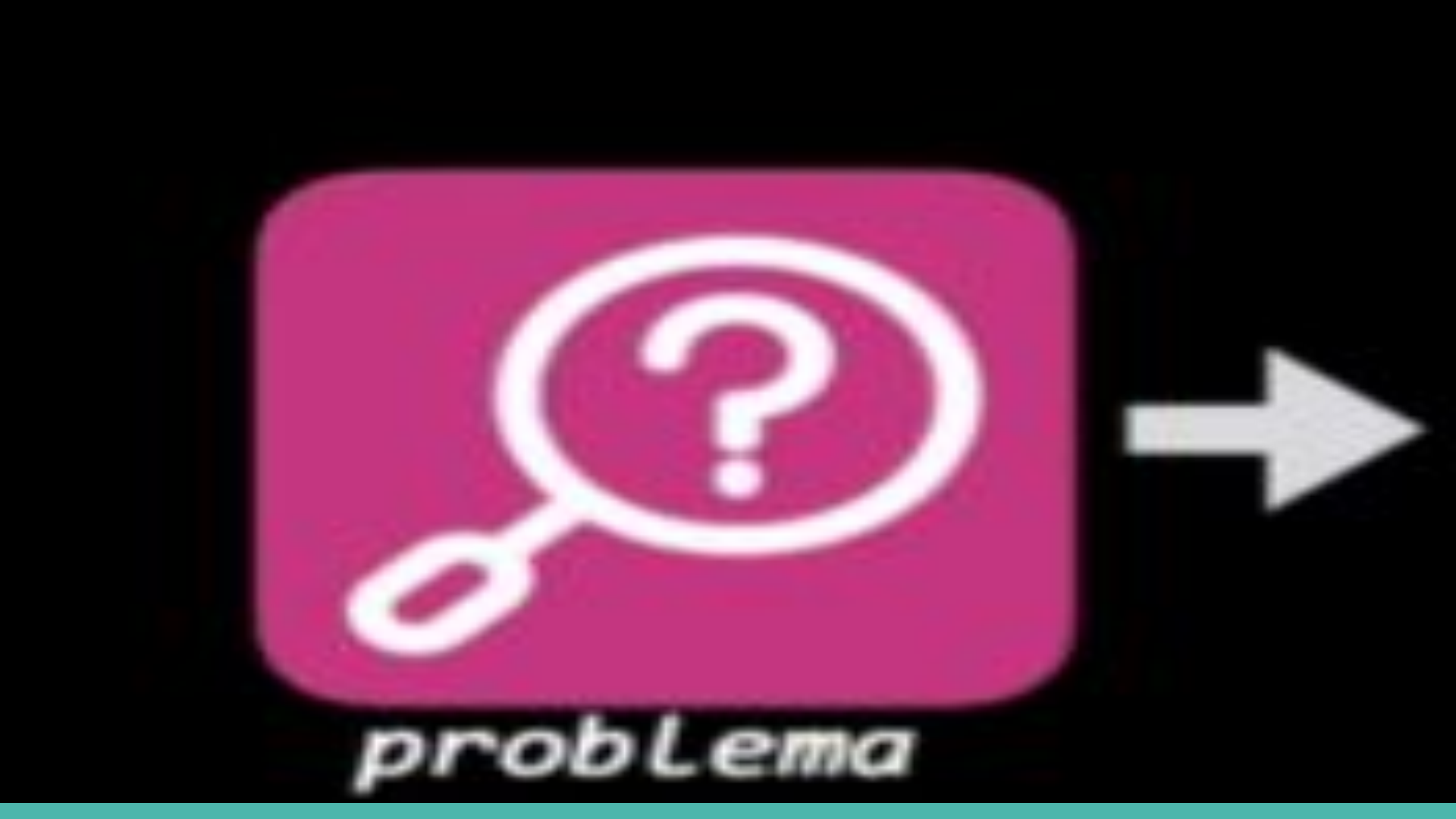
La regresión logística estima la probabilidad de que ocurra un evento, como votar o no votar, en función de un conjunto de datos determinado de variables independientes.
Este tipo de modelo estadístico (también conocido como *modelo logit*) se utiliza a menudo para la clasificación y el análisis predictivo. Dado que el resultado es una probabilidad, la variable dependiente está limitada entre 0 y 1. En la regresión logística, se aplica una transformación logit a las probabilidades, es decir, la probabilidad de éxito dividida por la probabilidad de fracaso.


p
1
0.5
0
$y = b_0 + b_1x + b_2x \ldots$

# Ciclo de vida de un proyecto de machine learning

El ciclo de vida del machine learning consta de pasos que proporcionan estructura al proyecto de machine learning y dividen eficazmente los recursos de la empresa.




problema

Una empresa del sector inmobiliario desea determinar si resulta viable invertir en la compra de propiedades en diferentes zonas de una ciudad. Es decir que dada una propiedad la idea es determinar su precio dadas algunas características de esta propiedad (ubicación, número de habitaciones,la densidad de la población, entre otras, etc).

**Fase 1. Tomar el problema descrito en un lenguaje general**

**y llevarlo a un lenguaje específico del ML.**

**“ Desarrollar un modelo que permita predecir el precio de una propiedad dadas características como ubicación, número de habitaciones etc.”**

# Recolección de los datos

Identificar las características o variables de entrada y salida.

Librerías a utilizar:
pandas



# Limpieza de datos

Optimizar los datos.

Librerías a utilizar:
pandas



# Exploración de datos

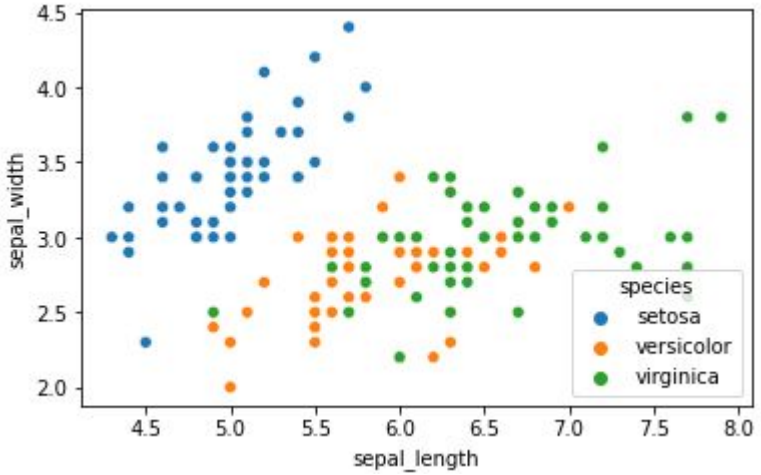
Análisis de correlación
Graficar
Matriz de correlación



# Algoritmos de aprendizaje no supervisado

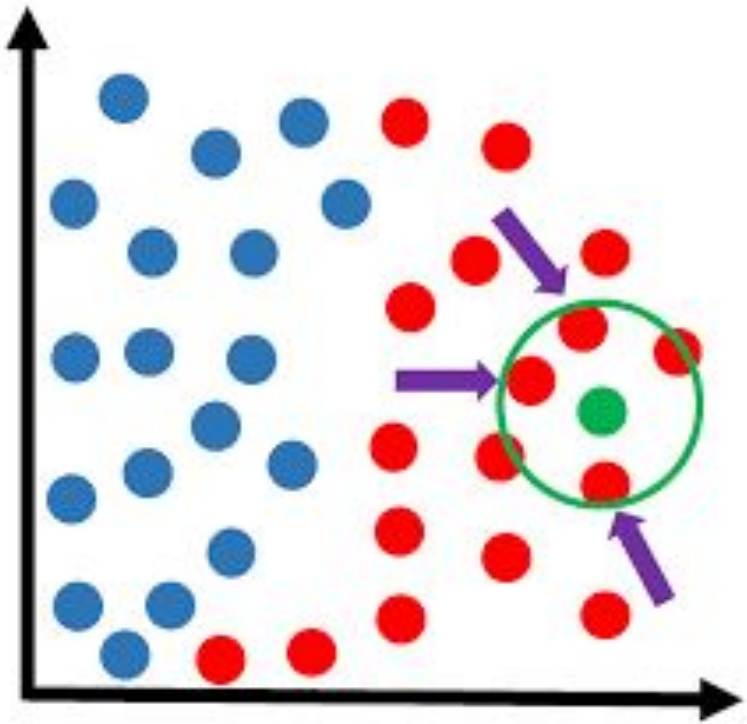
Los **algoritmos de aprendizaje no supervisado** son una herramienta importante en el campo del aprendizaje automático, y se utilizan para hacer predicciones o tomar decisiones basadas en datos. A diferencia de los algoritmos supervisados, que se entrenan utilizando datos etiquetados, los algoritmos no supervisados no tienen datos de entrenamiento etiquetados. En lugar de eso, utilizan la estructura inherente de los datos para aprender y hacer predicciones.

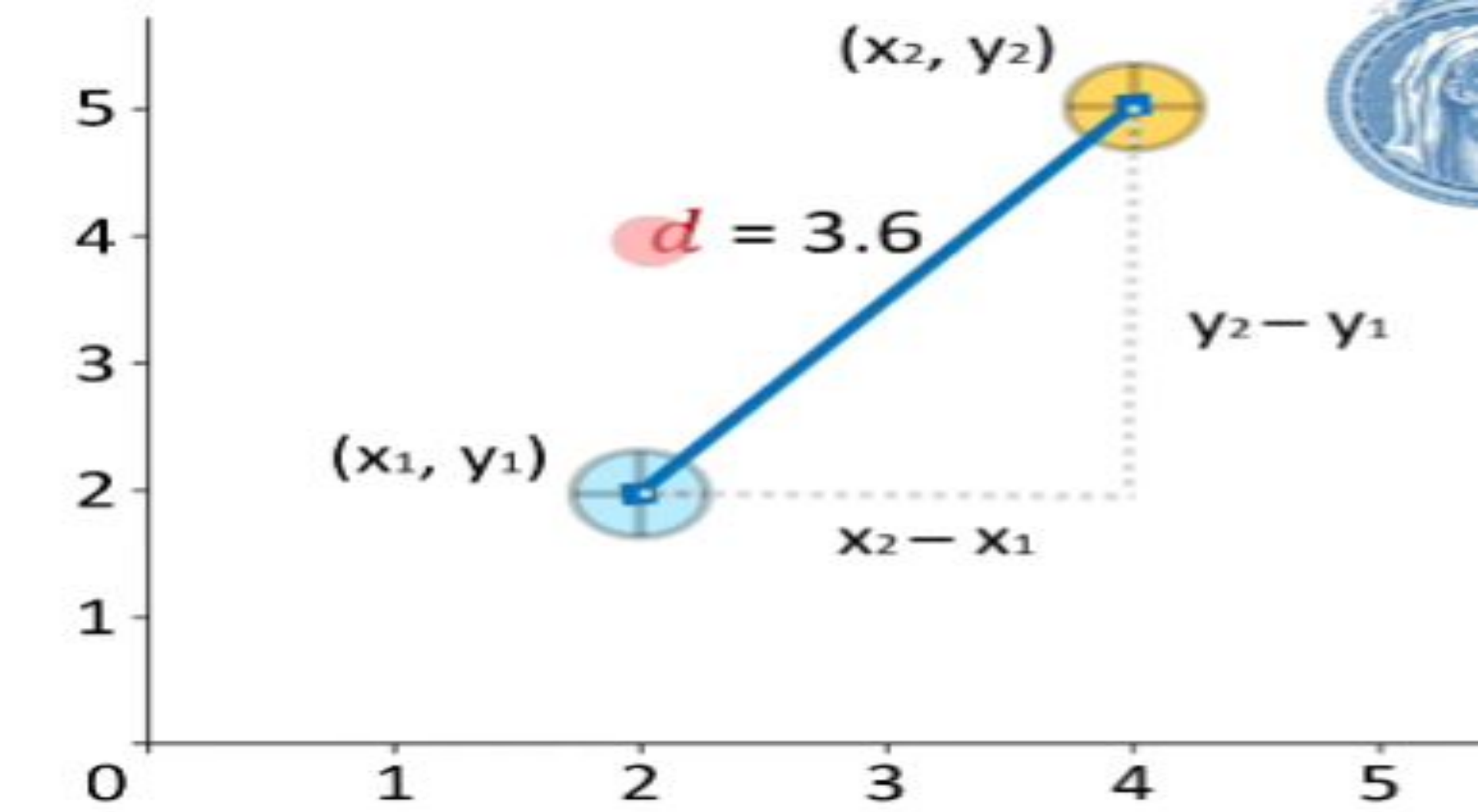
Un ejemplo común de un algoritmo de aprendizaje no supervisado es el **clustering** (agrupamiento), que se utiliza para dividir un conjunto de datos en grupos o clústeres basados en la similitud entre los datos:



# ¿Qué es KNN?

El algoritmo k vecinos más cercanos (KNN) es un clasificador de aprendizaje supervisado no paramétrico, que emplea la proximidad para realizar clasificaciones o predicciones sobre la agrupación de un punto de datos individual. Es uno de los clasificadores de clasificación y regresión más populares y sencillos que se emplean actualmente en el machine learning.




(x2, y2)
d = 3.6
y2 − y1
(x1, y1)
x2 − x1
5
4
3
2
1
0
1
2
3
4
5

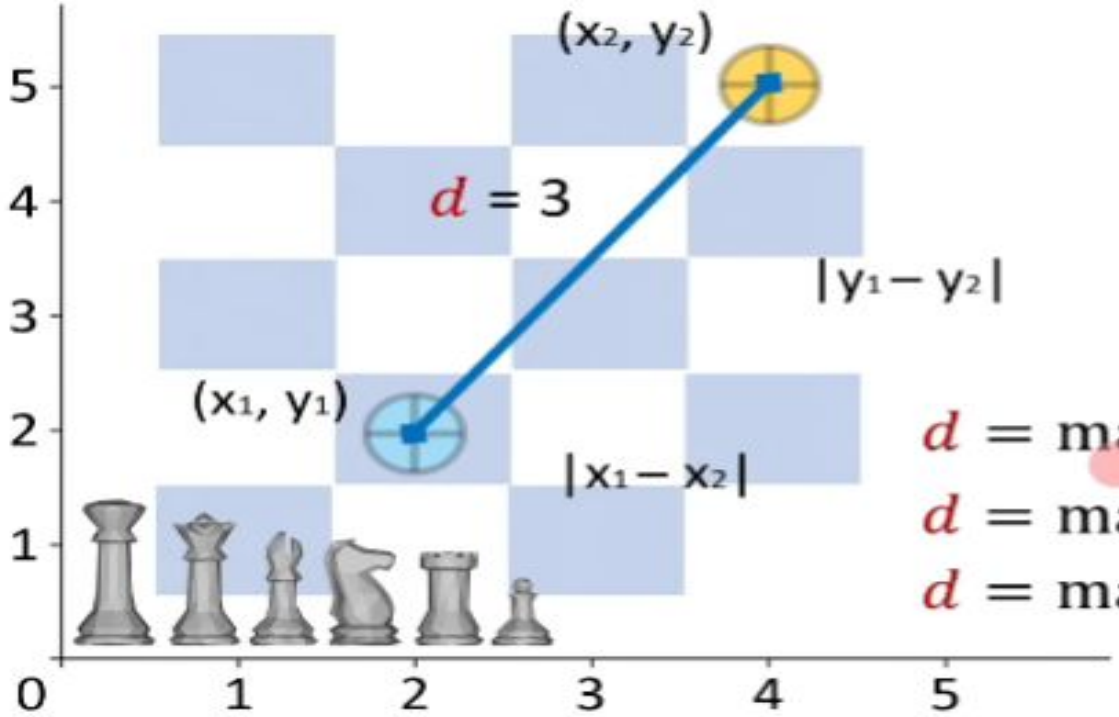
# Distancia Euclidiana

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$d = \sqrt{(4 - 2)^2 + (5 - 2)^2}$$

$$d = \sqrt{(2)^2 + (3)^2} = \sqrt{13}$$

# Distancia Chebyshev



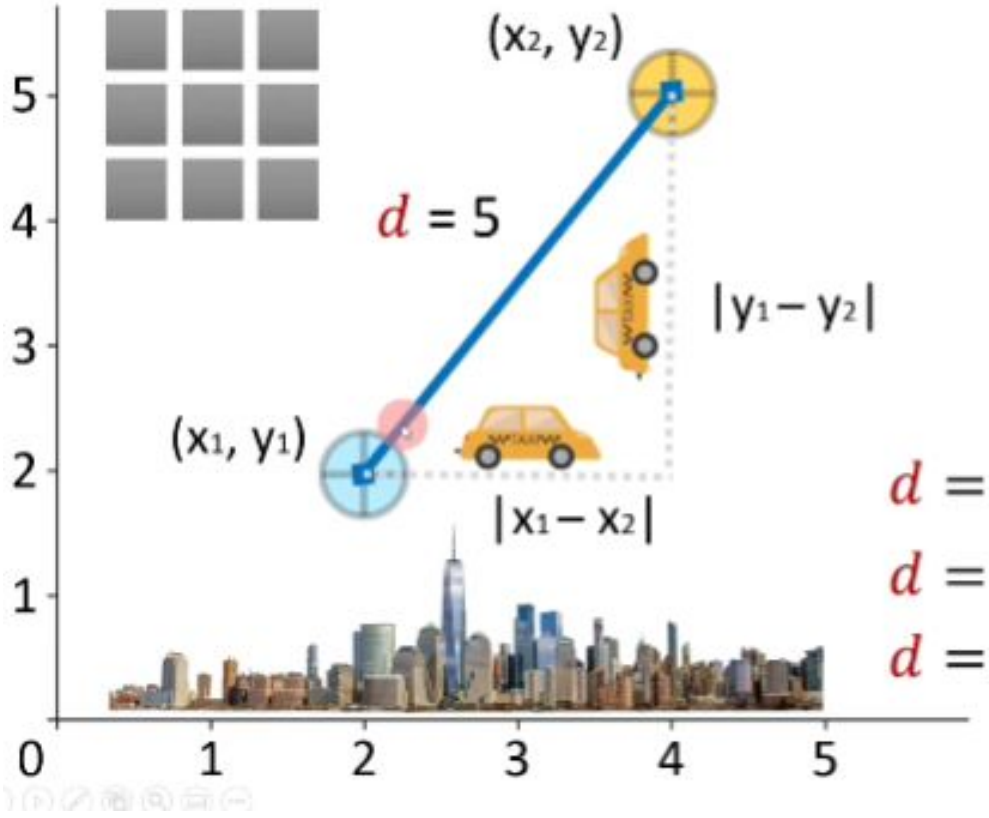

## Distancia Chebyshev

$$d = \max(|x_1 - x_2|, |y_1 - y_2|)$$

$$d = \max(|2 - 4|, |2 - 5|)$$

$$d = \max(2, 3) = 3$$

# Distancia Manhattan




## Distancia Manhattan

$$d = |x_1 - x_2| + |y_1 - y_2|$$

$$d = |2 - 4| + |2 - 5|$$

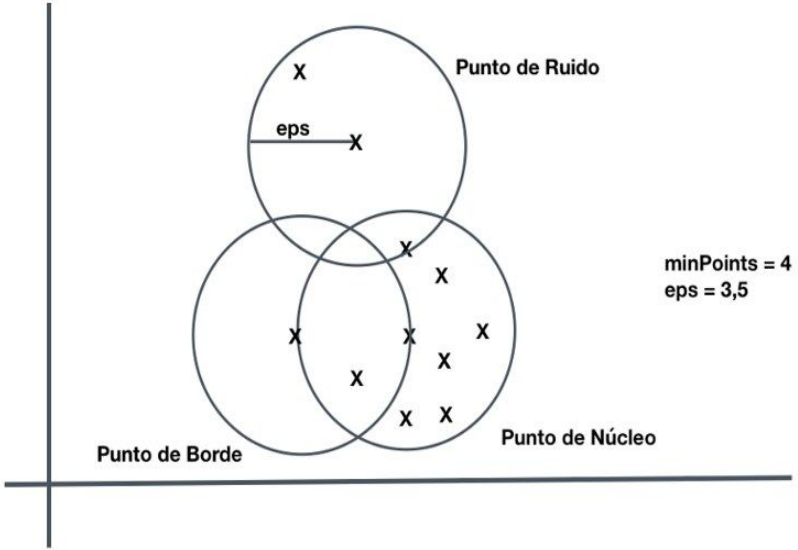
$$d = 2 + 3 = 5$$

# DBSCAN

El algoritmo DBSCAN requiere básicamente 2 parámetros:

**Épsilon (eps):** especifica lo cerca que deben estar los puntos entre sí para ser considerados parte de un clúster. Esto significa que, si la distancia entre dos puntos es menor o igual a este valor de épsilon, estos puntos se consideran vecinos.

**Puntos mínimos (minPts):** el número mínimo de puntos para formar una región densa.



# Aprendizaje No Supervisado



SEGMENTACIÓN DE CLIENTES PARA CAMPAÑAS DE MARKETING



DETECCIÓN DE GENES QUE SE COMPORTAN DE LA MISMA MANERA



ANÁLISIS DE RELACIONES EN LAS REDES SOCIALES



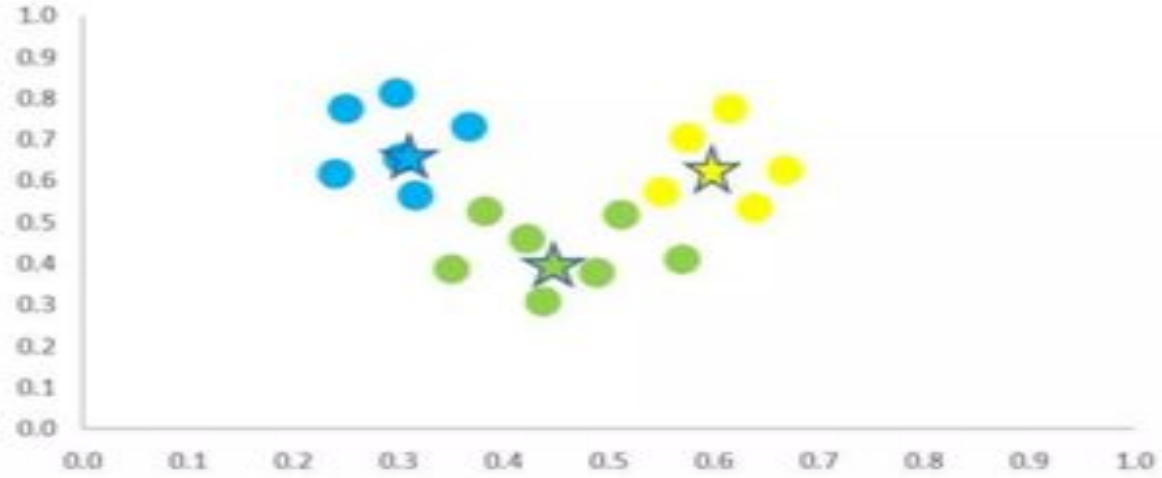
AGRUPACIÓN DE DOCUMENTOS



RECOMENDACIONES MUSICALES O DE CONTENIDOS

# Clustering K-Means

*Método no supervisado que sirve para hacer agrupaciones de objetos o individuos en base a la similitud de sus características*




*Se busca formar clusters compactos y bien definidos*

# Ventajas

*Es fácil de implementar*

*Es útil cuando queremos segmentar grandes volúmenes de datos (Cuando se trata de big data se utilizan variantes como el Mini-Batch K-Means)*

*Computacionalmente es menos costoso que el clustering Jerárquico*

*Ha sido ampliamente utilizado en bases de datos reales con buenos resultados*

# Desventajas

*Necesitamos conocer de antemano la cantidad de clusters a formar*

*Debido a que crea los grupos en base a las distancias entre los individuos, solamente maneja variables numéricas.*

*La calidad de los clusters depende en gran medida de los valores con los que se inicializa el algoritmo*

*Puede ser que se obtenga un resultado que sea óptimo localmente pero no globalmente*

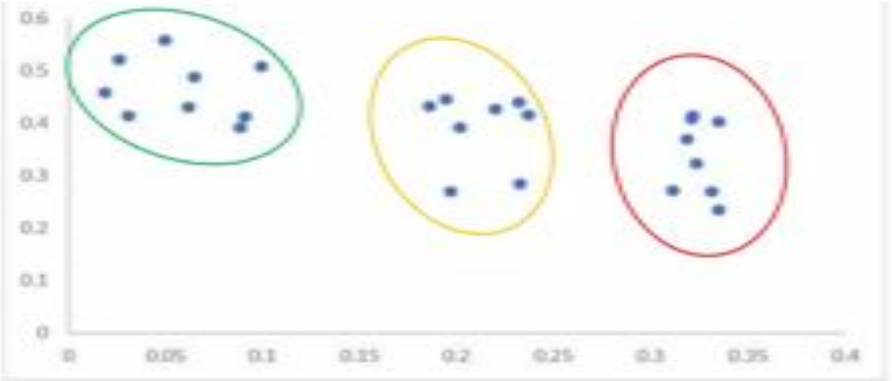
*El resultado se vé muy afectado por los outliers*

# Cómo afectan los Outliers?

Con Outliers




Sin Outliers




GRACIAS