

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

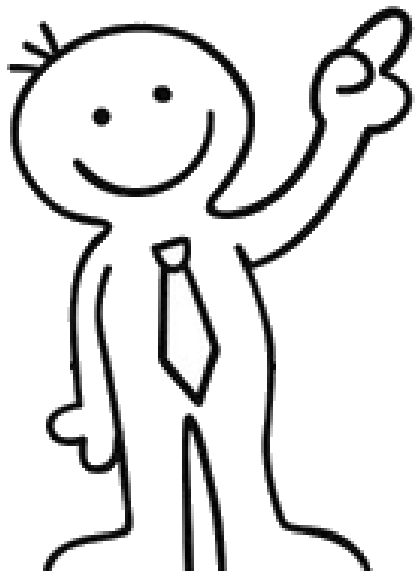
**Carlos Mera**  
**John Willian Branch**

Departamento de Ciencias de la Computación y de la Decisión  
Investigador del Grupo de I+D en Inteligencia Artificial – GIDIA

[camerab@unal.edu.co](mailto:camerab@unal.edu.co)

# Contenido

1. Definición
2. Motivación – ¿por qué reconocer patrones?
3. Conceptos relacionados Reconocimiento de Patrones
4. Métodos de Aprendizaje de Máquina:
  - a. Métodos Supervisados
  - b. Métodos No Supervisado
5. Regresión Lineal
6. Regresión Logística
7. Clasificador kNN
8. Naïve Bayes
9. LDA y QDA
10. Máquinas de Vectores de Soporte
11. Árboles de Decisión
12. Redes Neuronales



¿QUÉ ES EL RECONOCIMIENTO DE PATRONES?

# Definición



El **RECONOCIMIENTO DE PATRONES** es la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos o abstractos, con el propósito de extraer información que permita establecer propiedades de entre conjuntos de dichos objetos [\[wikipedia\]](#)

El **RECONOCIMIENTO DE PATRONES** es un área de investigación en la que:

- Se encuentran ...
- Se reconocen ...
- Se descubren ...

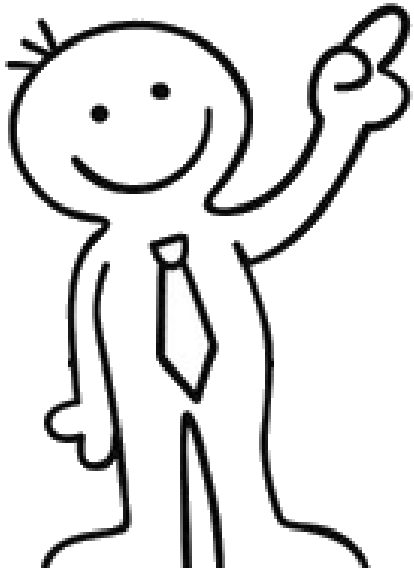
**PATRONES EN LOS DATOS**

## TÉRMINOS RELACIONADOS:

- Aprendizaje de máquina
- Minería de datos
- Clasificación
- Clustering, ...

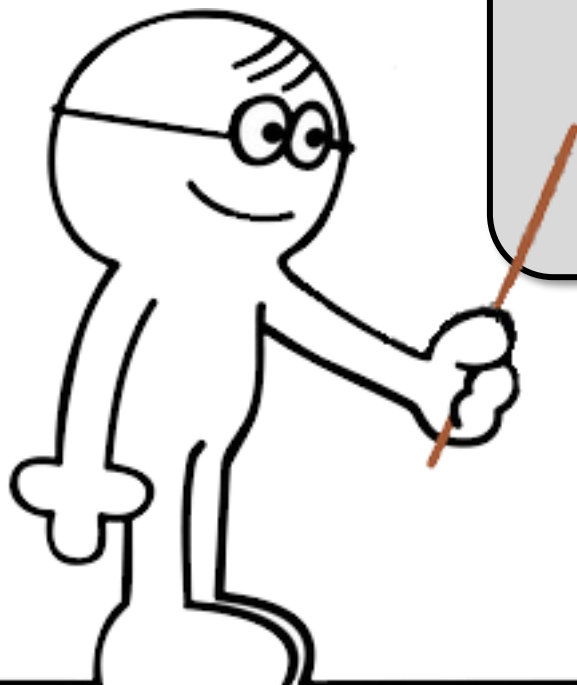


# Motivación



¿POR QUÉ RECONOCER PATRONES?

<https://www.ted.com/talks/fei-fei-li-how-we-re-teaching-computers-to-understand-pictures?language=es>



# MÉTODOS DE APRENDIZAJE DE MÁQUINA

# Conceptos



## EL APRENDIZAJE DE MÁQUINA:

Es una forma de la IA que permite a un sistema aprender de los datos en lugar programar de manera explícita un conjunto de reglas [IBM].

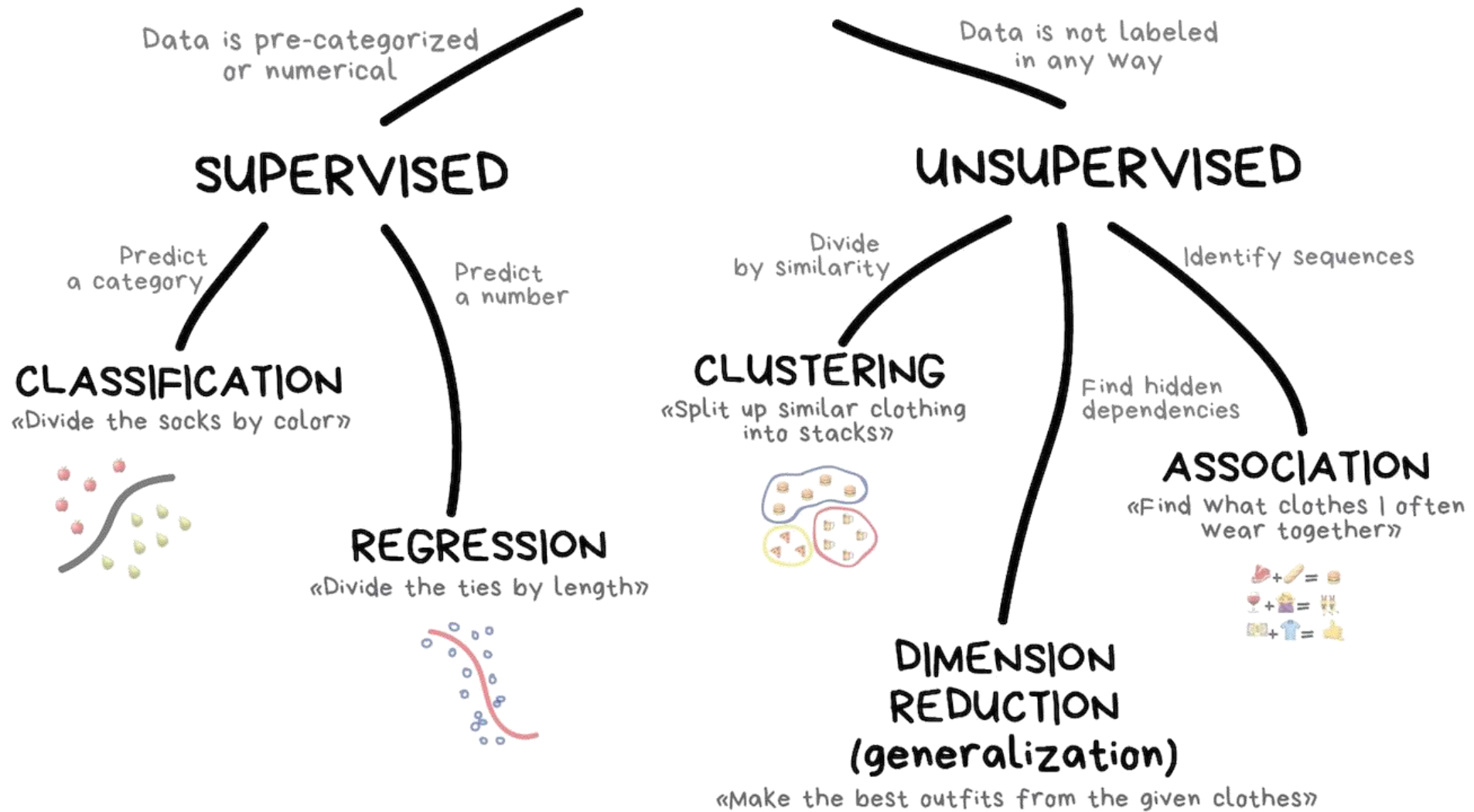


## Objetivo:

**Crear un modelo** de los datos optimizando un criterio (o medida) de rendimiento.

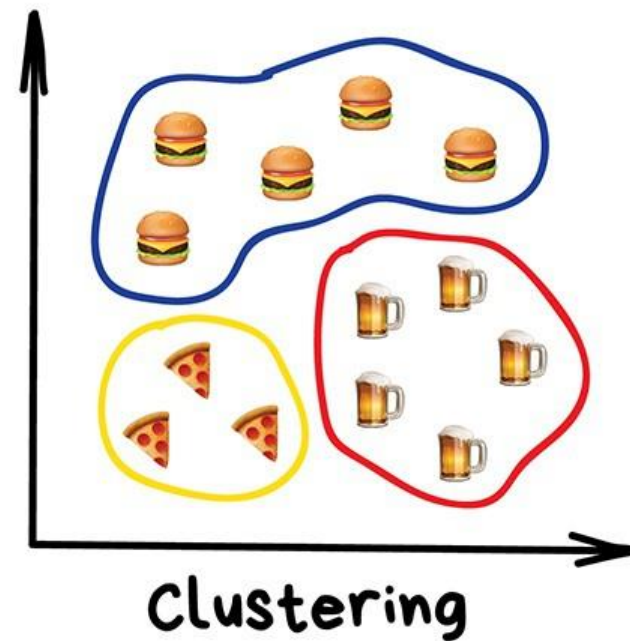
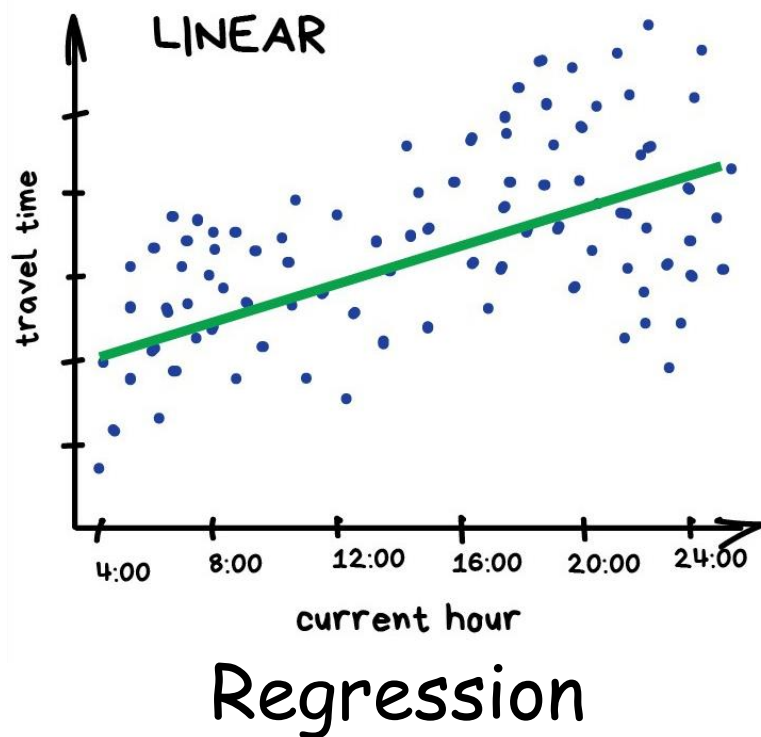
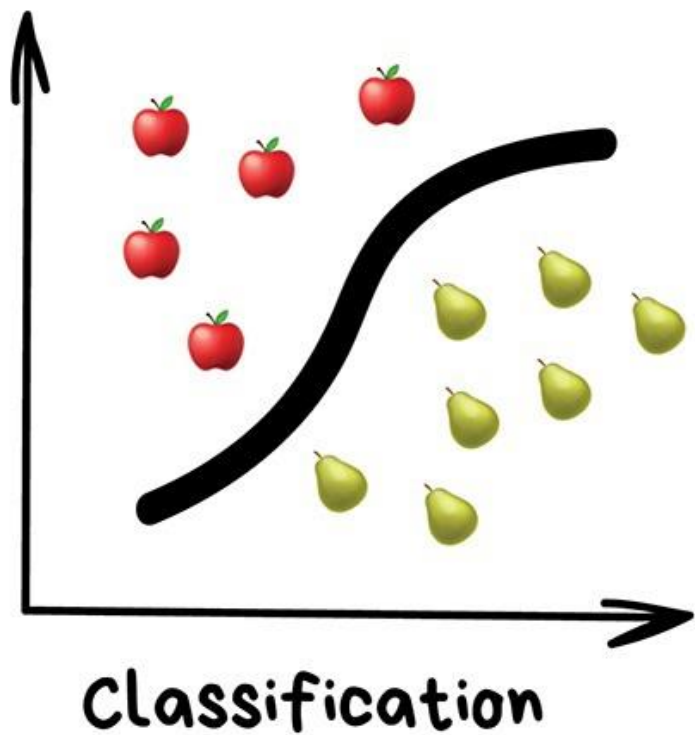


# CLASSICAL MACHINE LEARNING



Tomada de: [https://vas3k.com/blog/machine\\_learning/](https://vas3k.com/blog/machine_learning/)

# Aprendizaje Supervisado vs No Supervisado



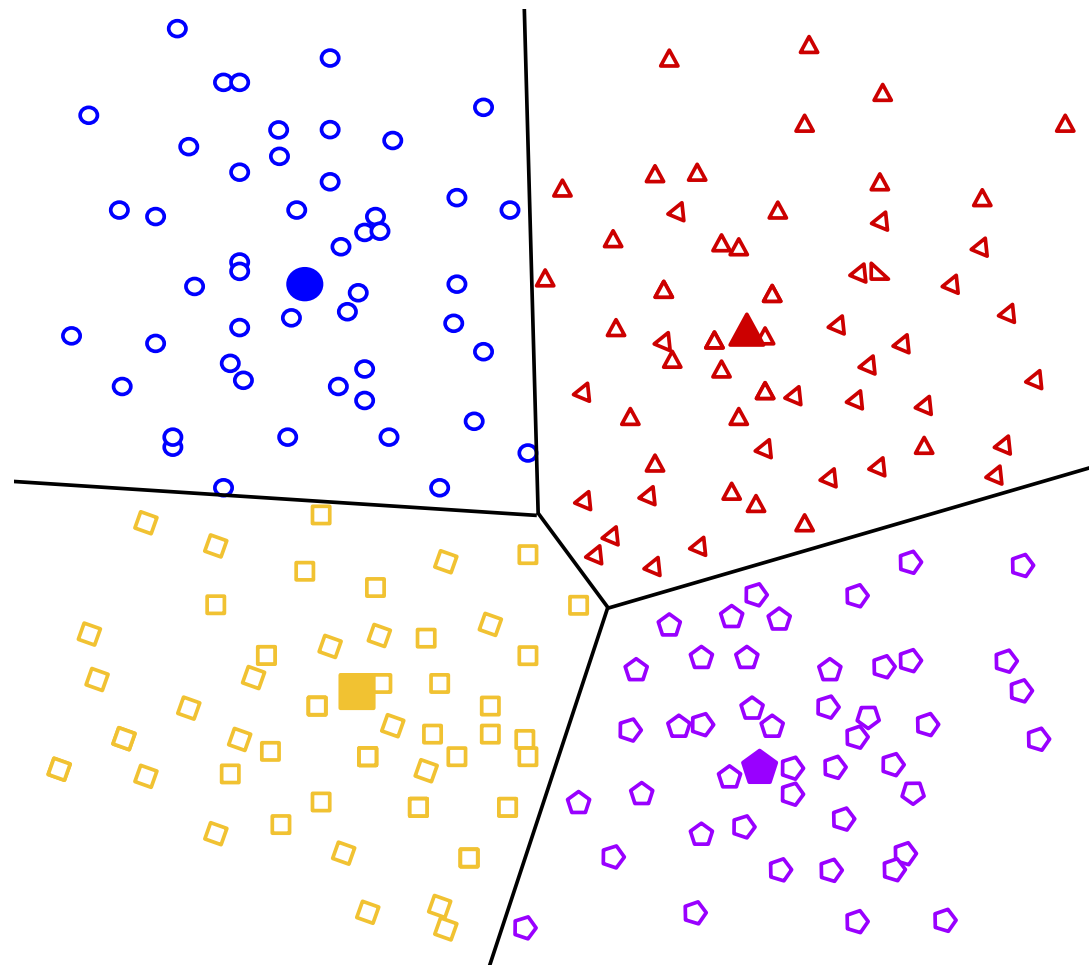
# Aprendizaje Supervisado VS Aprendizaje NO Supervisado

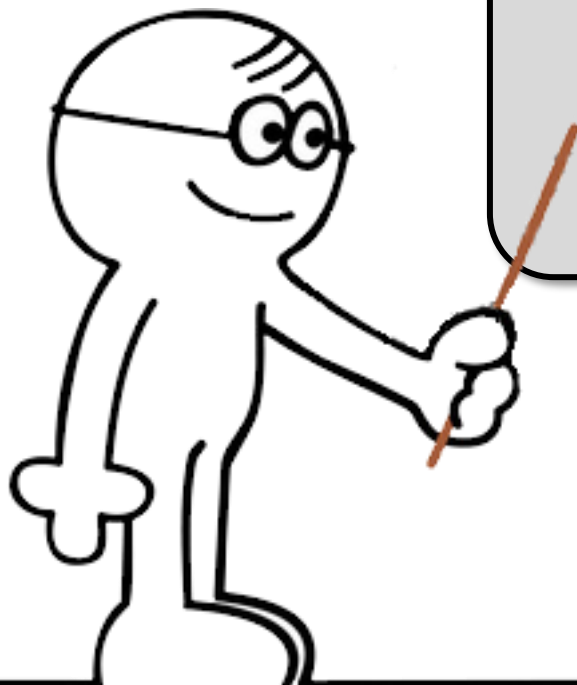
**Aprendizaje Supervisado:** requiere de un conjunto de datos conocidos a partir del cual se crea el **modelo** para predecir el valor de una variable de salida. El aprendizaje supervisado se puede usar en dos tareas:

- **Clasificación:** en este caso la variable de salida es una etiqueta que determina la clase a la que pertenecen los datos de entrada, es decir, la variable de salida es una **variable discreta**.
- **Regresión:** en este caso los algoritmos de aprendizaje buscan predecir el valor de una **variable continua** a partir de los datos de entrada. Un ejemplo de una tarea de regresión es el de estimar la longitud de un salmón en función de su edad y su peso.

En el **Aprendizaje no supervisado** se cuenta con un conjunto de datos de entrenamiento, pero no hay una variable específica de salida (se desconocen las clases). En este sentido, el objetivo de los problemas del aprendizaje no supervisado es, por ejemplo, el de agrupar los datos de entrada con base en algún criterio de similitud o disimilitud o determinar la distribución estadística de los datos, conocida como estimación de la densidad.

## Aprendizaje Supervisado VS Aprendizaje No Supervisado





¿QUÉ SE ENTIENDE POR APRENDIZAJE?

# Aprendizaje de Máquina

✂ SI PARTIMOS DE QUE SE TIENE ...



Un Conjunto de datos  $D$



Una medida de desempeño  $M$



Una tarea (de clasificación o predicción)  $T$

## Aprendizaje

Se dice que un sistema **aprende** de  $D$  un modelo para realizar la tarea  $T$  si después de aprender, el rendimiento del sistema en realizar la tarea  $T$  mejora según lo medido por  $M$ .

En otras palabras, el modelo aprendido ayuda al sistema a realizar mejor la tarea  $T$ , en comparación con la ausencia del aprendizaje.

# Aprendizaje de Máquina

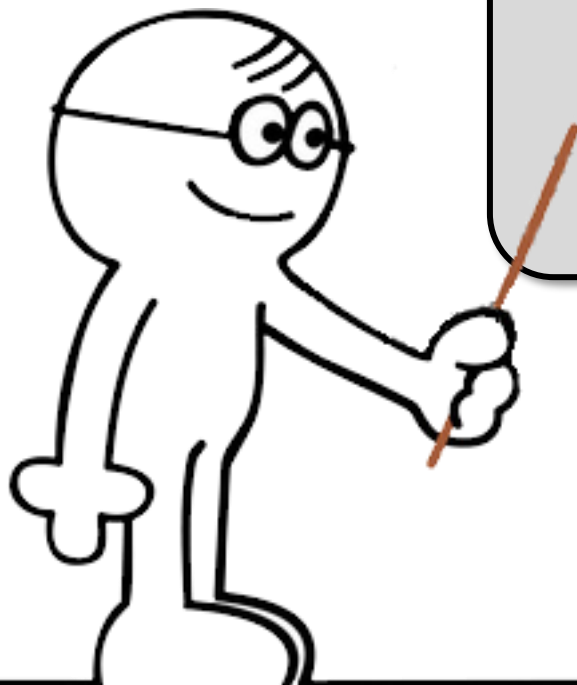
🚩 HAY UN SUPUESTO BASTANTE IMPORTANTE ...



## SUPUESTO:

La distribución estadística de los datos de entrenamiento debe ser idéntica a la distribución de los datos con los que se usará el sistema

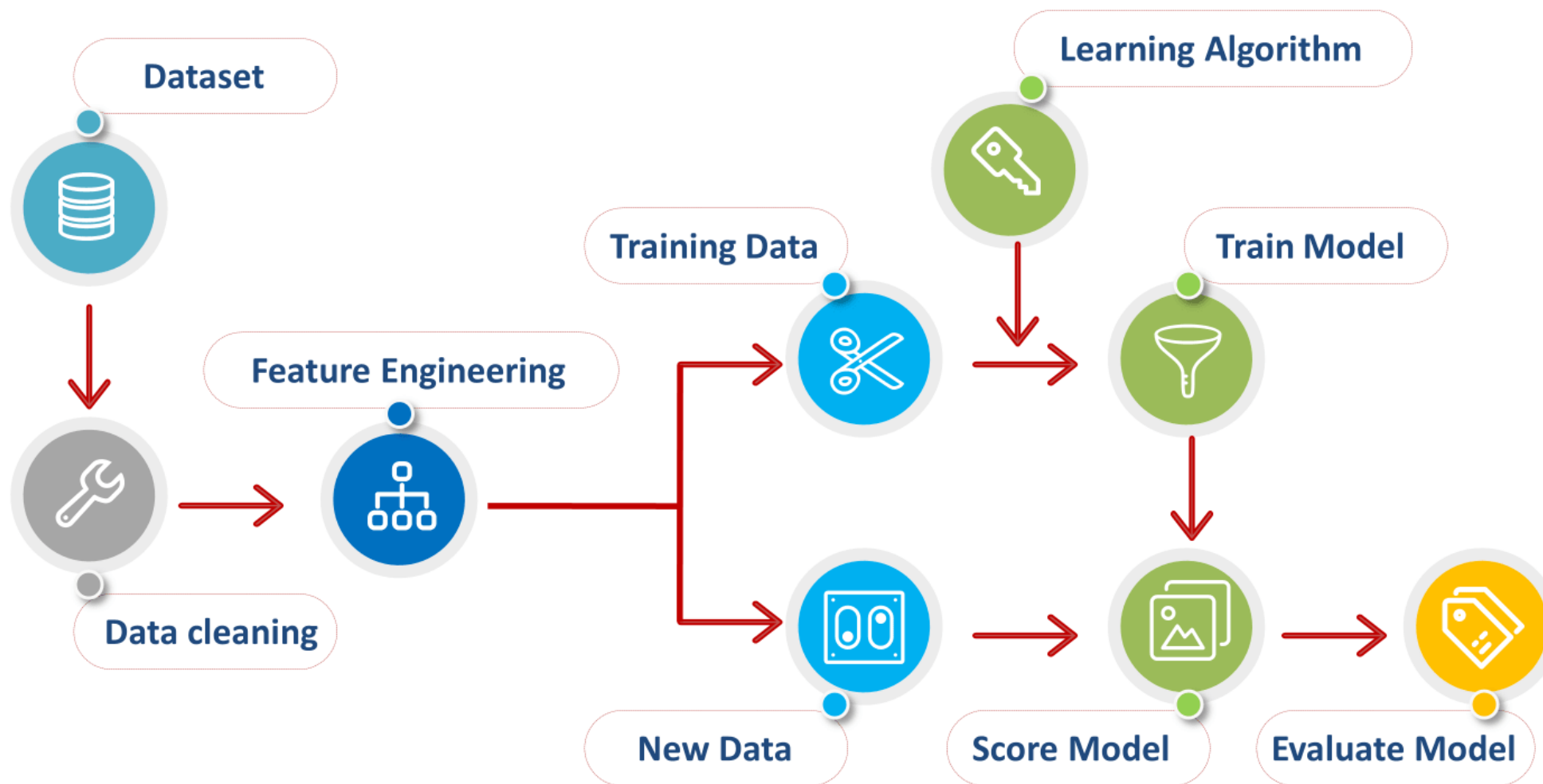
- ✓ En la práctica, esta suposición a menudo se viola en cierto grado.
- ✓ Desviaciones en los datos claramente darán lugar a una clasificación (o predicción) deficiente.
- ✓ Para lograr una buena precisión en los datos de prueba, los ejemplos de entrenamiento deben ser suficientemente representativos de los datos de prueba.



## FLUJO DE TRABAJO DE LOS MÉTODOS DE APRENDIZAJE AUTOMÁTICO



# ■ Aprendizaje de Máquina



Tomada de: <http://www.cs.us.es/~fsancho/images/2019-12/ml-process.png>

# Aprendizaje de Máquina

## PARA EL APRENDIZAJE ...

- Partimos de que tiene un conjunto de datos conocidos y etiquetados, a partir del cual se crea un **modelo de clasificación o de regresión**.



**Conjunto de Datos (D):** es un conjunto de instancias (objetos, individuos, ejemplos, casos o registros) etiquetadas  $(x_i, y_i)$  que están definidas por:

- $k$ -atributos (o características):  $x_i = x_i^1, x_i^2, \dots, x_i^k$
- Una etiqueta  $y_i$  que define la clase predefinida de  $x_i$



**Atributo:** es una característica (campo o propiedad) que describe una instancia  $x_i$ . Un atributo puede tener valores continuos o discretos



**Clase :** es un conjunto de individuos que tienen algunas características importantes en común, que los hacen similares

# Construcción de los conjuntos de datos

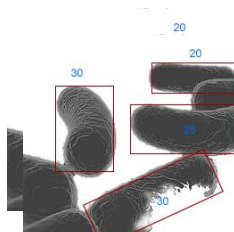
En imágenes: problema reconocimiento de bacterias



Conjunto de imágenes



Se segmentan los objetos



Se extraen características

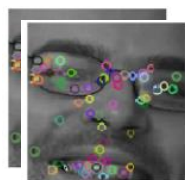
Longitud	Perímetro	Área	Intensidad	Etiqueta
120	567	2456	127	Bac
350	980	3870	98	No-Bac
125	546	2145	125	Bac

Matriz de características y etiquetas

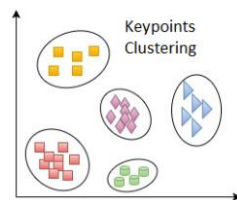
En imágenes: problema de reconocimiento de caras



Conjunto de imágenes



Extracción de Key Points



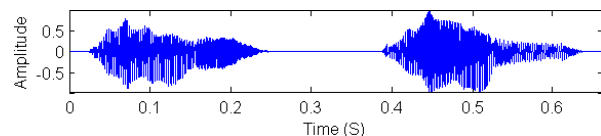
Uso de un diccionario de palabras

R1	R2	R3	R4	R5	Etiqueta
0	3	2	5	6	Cara
7	6	1	0	2	No-Cara
0	2	3	4	8	Cara

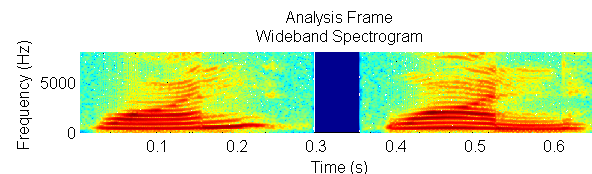
Matriz de características y etiquetas

# Construcción de los conjuntos de datos

En sonido: problema de detección de transacciones fraudulentas en llamadas



Conjunto de grabaciones



Pre-Procesamiento de la señal

Longitud de onda	Frecuencia dominante	Pico de frecuencia	Máxima energía	Etiqueta
120	67	0.8	445	Fraude
350	35	0.9	398	No-Fraude
125	56	0.8	289	No-Fraude

Matriz de características y etiquetas

Datos estructurados: problema de clasificar un cliente con base en sus datos bancarios

Estrato	Ingresos	Deudas	Tipo_C	Buen Cliente?
2	3500000	5520000	"A"	Si
4	3800000	3408000	"AAA"	Si
6	9800000	19560000	"A"	No

Datos bancarios para determinar si se hace o no un prestamos a un cliente

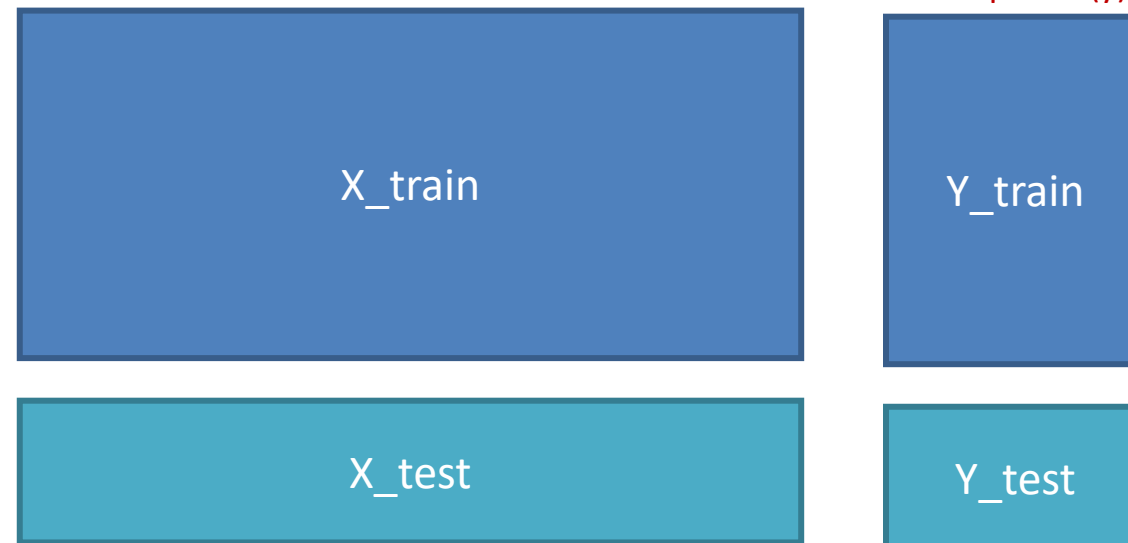
# Partición del Conjunto de Datos

Cuando se tienen “suficientes” datos estos se dividen en dos partes:

- **ENTRENAMIENTO:** son los datos que se usan para enseñar al modelo a reconocer los patrones. Estos se suelen sub-dividir en dos partes, una para entrenar (training) y otra para medir el error del entrenamiento (test)

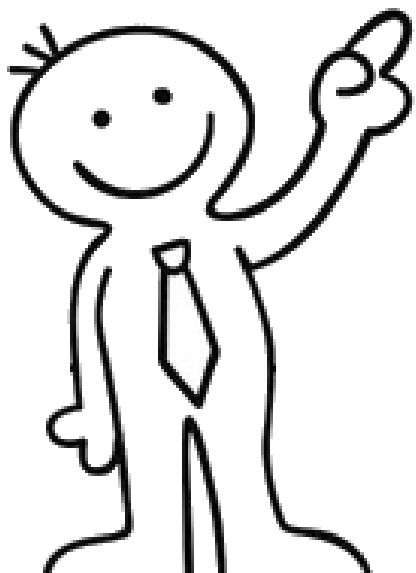
Matriz de características (X)

Vector de  
Etiquetas (y)



- **VALIDACIÓN:** es un conjunto de datos independiente que nunca se ha mostrado al modelo y determinar la precisión del mismo







¿CÓMO PARTICIONAMOS EL CONJUNTO DE  
DATOS?

# Partición del Conjunto de Datos

## MÉTODOS DE MUESTREO

-  Recuerde que entre mayor sea la muestra esta tendrá a ser mas representativa y, por tanto, el error del modelo podrá ser menor.
-  Se debe evitar tener ordenado el conjunto de datos



### Muestreo Aleatorio

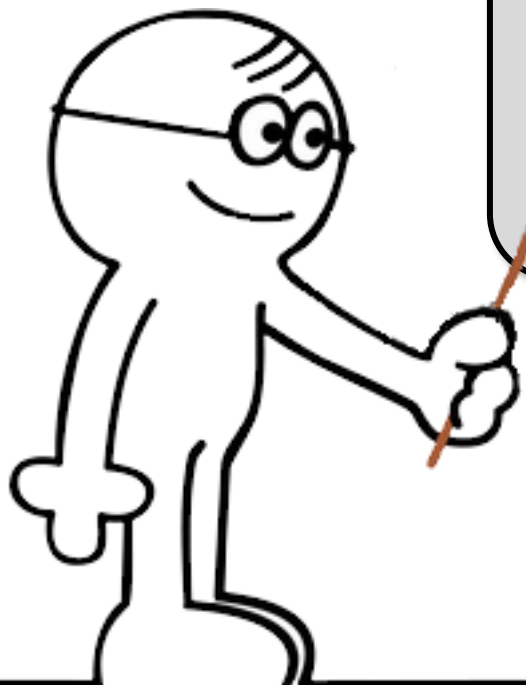
Todos los elementos de la población tienen la misma probabilidad de ser seleccionados:

- Muestreo con reemplazo
- Muestreo sin reemplazo



### Muestreo Estratificado

Con base en el número de clases, manteniendo la distribución de las mismas en la población



¿QUÉ ALGORITMO DE APRENDIZAJE UTILIZAR?



# Algoritmos de Aprendizaje

## Aprendizaje Supervisado - Clasificación

### Aprendizaje Tradicional:

- Regresión Logística
- Clasificador Ingenuo de Bayes
- Clasificador Lineal (LDA)
- Clasificador Cuadrático (QDA)
- Clasificador de Vecinos más Cercanos
- Máquinas de Vectores de Soporte (SVM)
- Árboles de decisión
- Redes Neuronales

### Aprendizaje No-Tradisional:

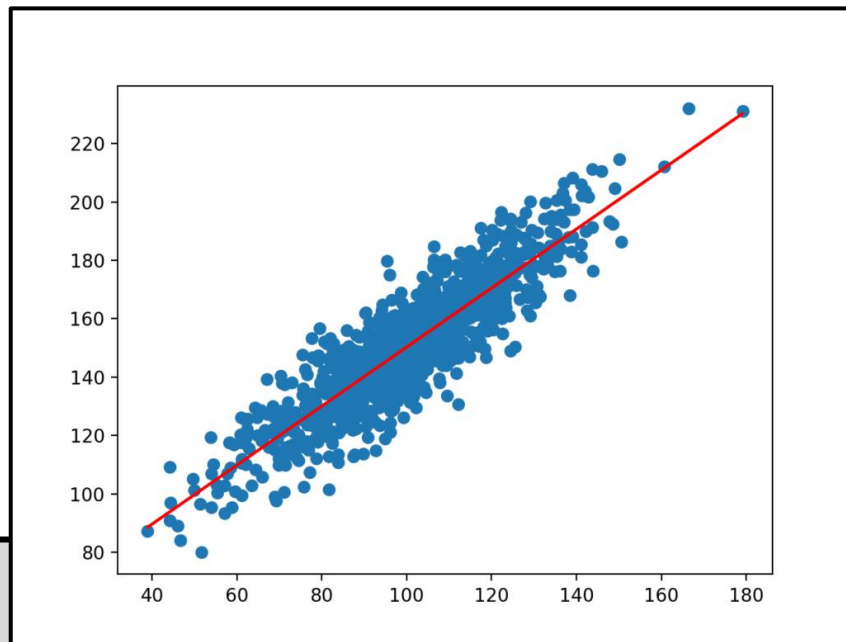
- Lógica difusa
- Métodos basados en ensambles
- Métodos de aprendizaje profundo
- Métodos de aprendizaje incremental

## Aprendizaje Supervisado - Regresión

- Regresión Lineal
- SVM para regresión
- Redes neuronales para regresión
- RANSAC

## Aprendizaje NO Supervisado

- K-Means
- Mean shift
- Clustering Espectral
- DBSCAN
- GMM - Modelos de Mezclas de Gaussianas



## REGRESIÓN LINEAL

# Regresión Lineal

## PUNTO DE PARTIDA

- Se cree que hay una relación de dependencia entre una variable cuantitativa  $y$ , y un grupo de variables independientes  $x_i$ . Es decir, se considera que hay una correlación entre ellas de tal manera que la variable  $y$  varía sistemáticamente con respecto a los valores de las variables independientes.
- Por ejemplo:**
  - ✓ ¿Hay relación entre la temperatura de una ciudad y el número de helados que se venden en una heladería?
  - ✓ ¿Puede identificar otras correlaciones?

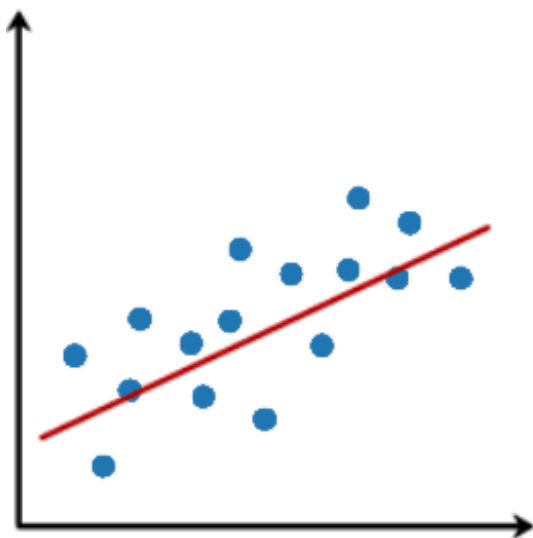


**OBSERVACIÓN:** Si sabemos que la variable  $x$  está correlacionada con  $y$ , quiere decir que podemos **predecir** la variable  $y$  a partir de la variable  $x$ .

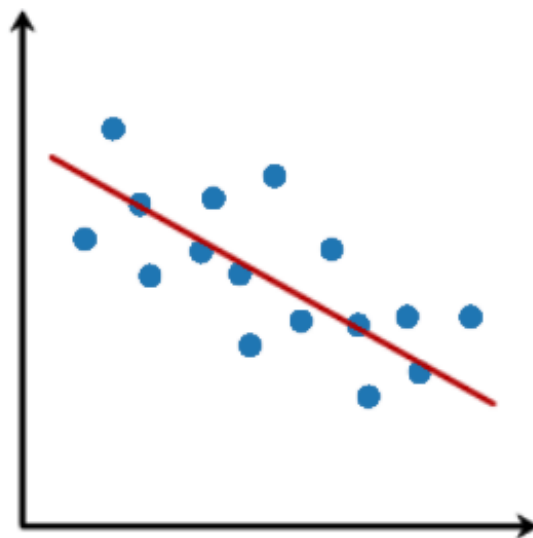
# Regresión Lineal

## TIPOS DE CORRELACIÓN:

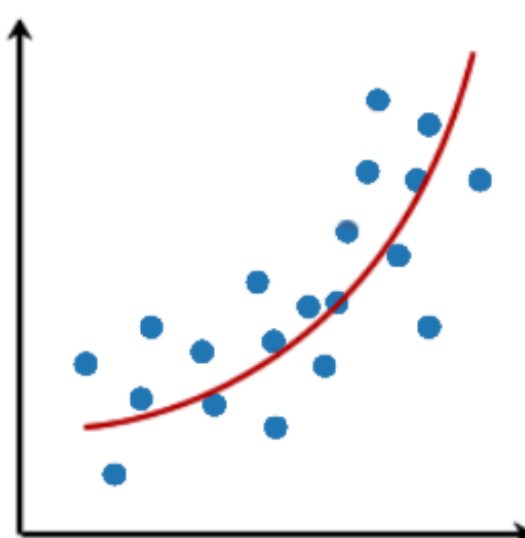
Hay cuatro tipos básicos de correlación: positiva, negativa, no-lineal y nula (sin correlación):



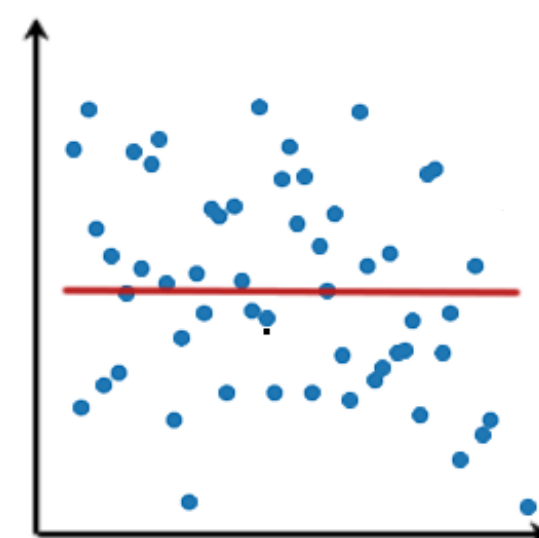
Correlación Lineal Positiva



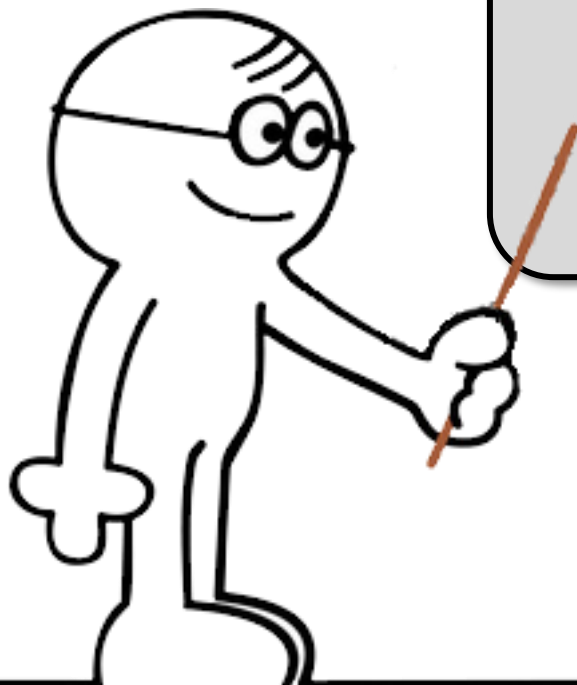
Correlación Lineal Negativa



Correlación No Lineal

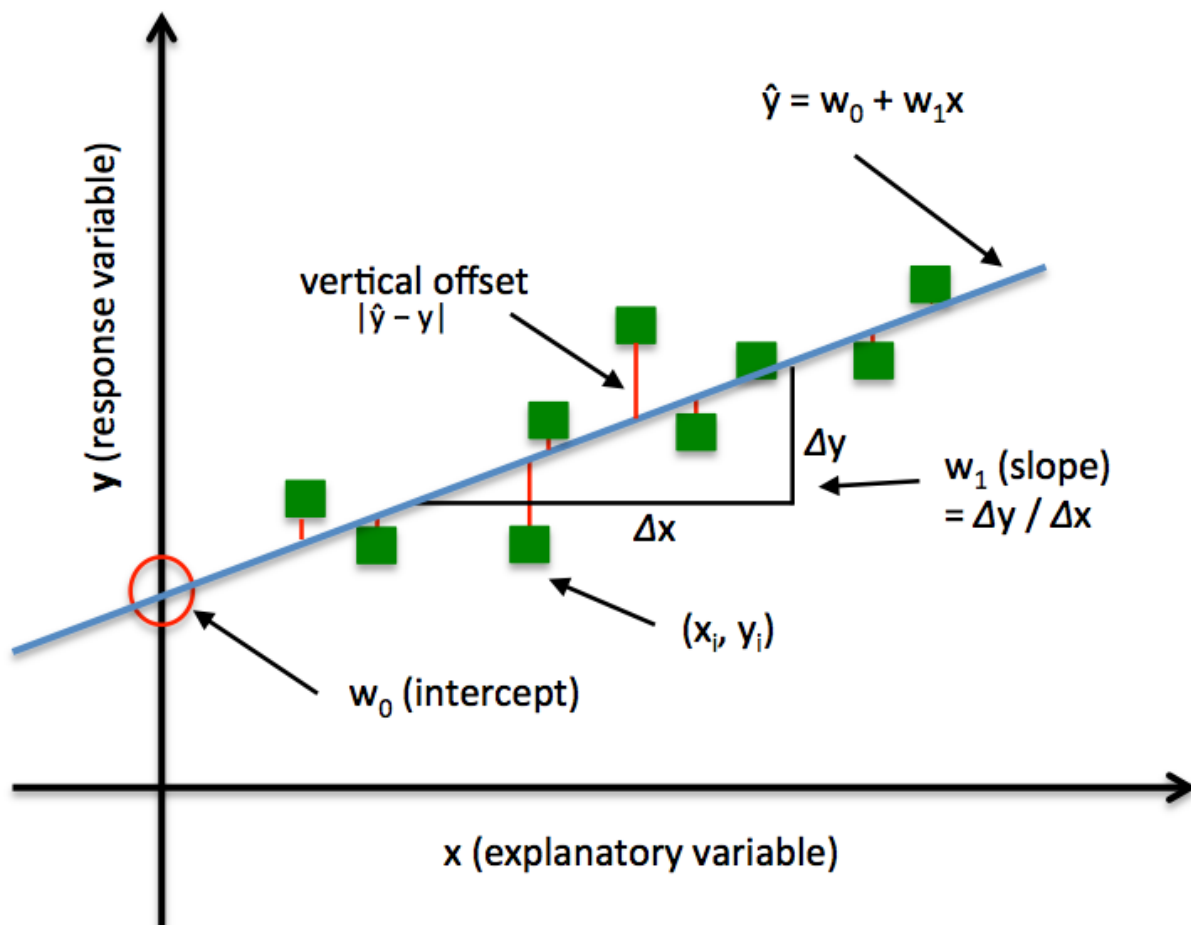


Sin Correlación



¿CUÁL ES EL PRINCIPIO DE LA REGRESIÓN LINEAL?

# Regresión Lineal



## ■ Generalidades:

- En la **Regresión Lineal** el modelo a ajustar es una línea recta:

$$\hat{y} = w_0 + w_1x$$

- Los métodos de regresión buscan modelar la relación entre 2 variables
- El modelo se ajusta usando una **medida de error** sobre las predicciones que éste hace

Puede haber múltiples líneas rectas dependiendo de los valores de intercepción y pendiente. Básicamente, lo que hace el algoritmo de regresión lineal es ajustar varias líneas y retornar la línea que produce el menor error.

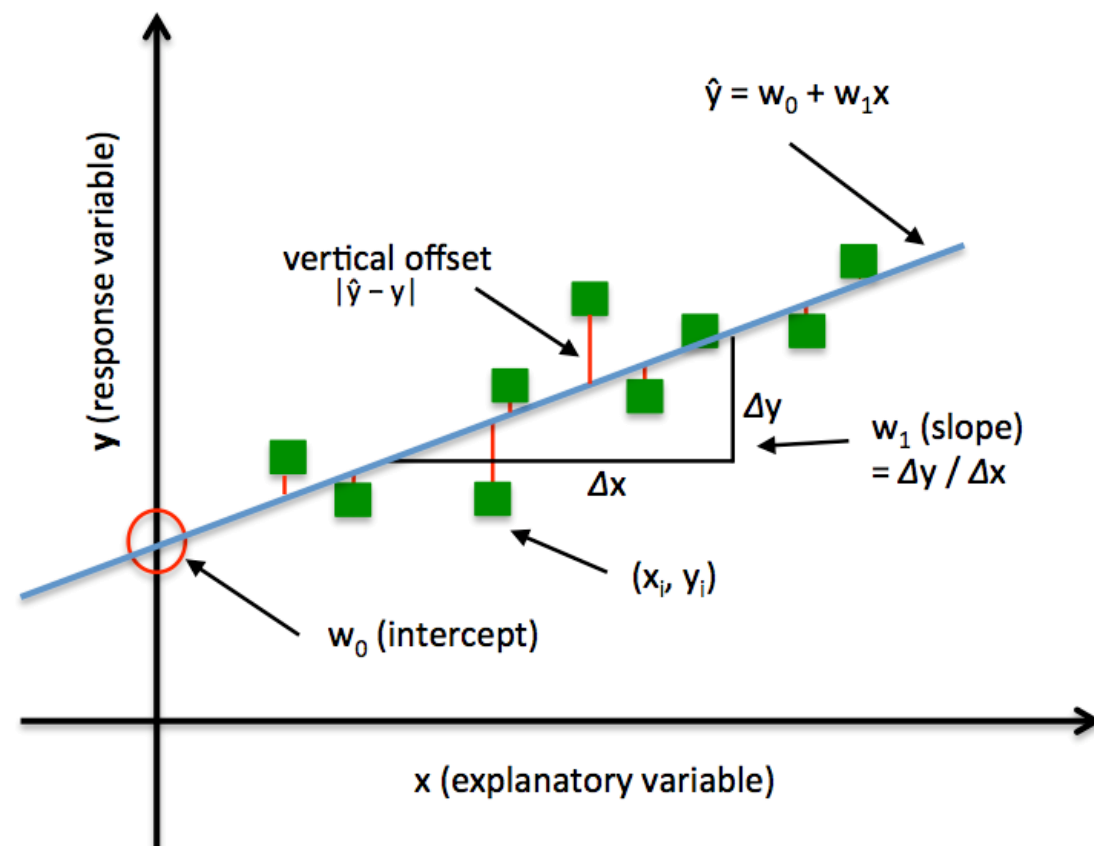
# Regresión Lineal

## GENERALIDADES

En la **REGRESIÓN LINEAL POR MÍNIMOS CUADRADOS**, el objetivo es encontrar la línea (o hiperplano) que minimiza la suma de los errores al cuadrado (SSE) o el error cuadrático medio (MSE) entre la variable objetivo ( $y$ ) y la salida del modelo sobre todas las muestras  $x_i$ . Entonces, buscamos aquella línea tal que minimice la función de costo:

$$J(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Predicción  
del Modelo  
↓  
↑  
Valor  
Observado



# Regresión Lineal

## REGRESIÓN LINEAL MÚLTIPLE:

Cuando la **Regresión Lineal** se usa para predecir una variable  $y$  a partir de **más de una variable** ( $x^i$ ), el modelo debe ajustar un hiperplano a los datos dados:

$$\hat{y} = w_0 + w_1x^1 + w_2x^2 + \dots + w_dx^d$$

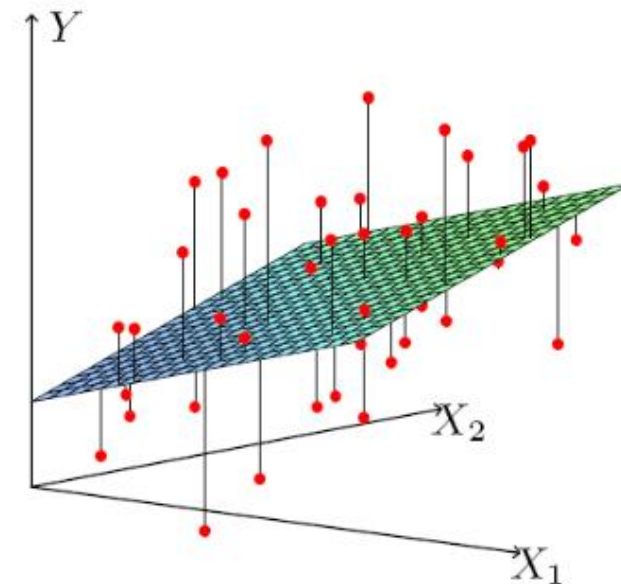
$$\hat{y} = h_{\mathbf{W}}(\mathbf{x}) = \sum_{k=0}^d w_k x^k = \mathbf{W}^T \cdot \mathbf{x}, \text{ con } x^0 = 1$$

Donde,

- $h_{\mathbf{W}}$  es la función de hiperplano
- $\mathbf{W}$  es el vector de parámetros del modelo (a estimar)
- $\mathbf{x} = \{x^1, x^2, \dots, x^d\}$  es el vector de características d-dimensional

La función de costo a minimizar sigue siendo la suma de los errores al cuadrado:

$$J(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^n (\mathbf{W}^T \cdot \mathbf{x}_i - y_i)^2$$





# Regresión Lineal

## SOLUCIÓN POR EL MÉTODO TEÓRICO CON ALGEBRA LINEAL:

Para encontrar el valor de  $W$  que minimiza la función de error, hay una solución de forma cerrada, en otras palabras, una ecuación matemática que da el resultado directamente.

Dicha ecuación es la siguiente:

$$W = (X^T X)^{-1} X^T y$$



Para más información se sugiere consultar:

- <https://www.oreilly.com/library/view/hands-on-machine-learning/9781491962282/ch04.html>
- <https://www.ritchieng.com/one-variable-linear-regression/>

# Regresión Lineal

## SOLUCIÓN POR EL MÉTODO ESTADÍSTICO:

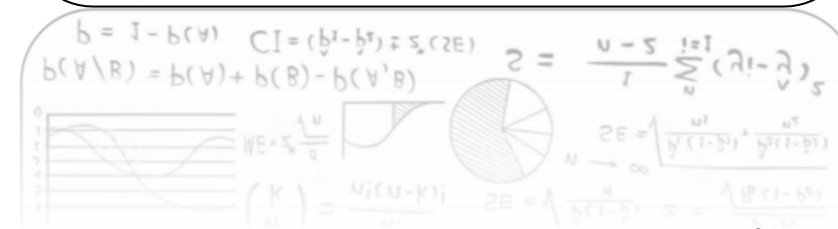
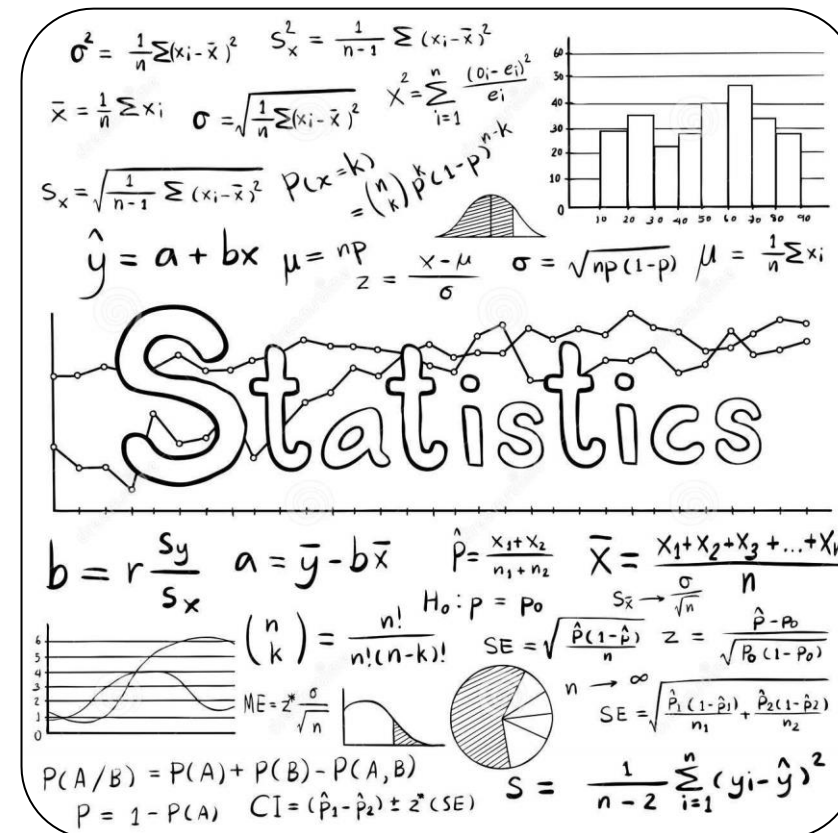
Para encontrar el valor de  $W$ , tenemos que:

$$w_0 = \bar{y} - w_1 \bar{x}$$

$$w_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

Para más información se sugiere consultar:

- <https://www.oreilly.com/library/view/hands-on-machine-learning/9781491962282/ch04.html>
- <https://www.ritchieng.com/one-variable-linear-regression/>



# Regresión Lineal

## SOLUCIÓN POR EL MÉTODO DE OPTIMIZACIÓN:

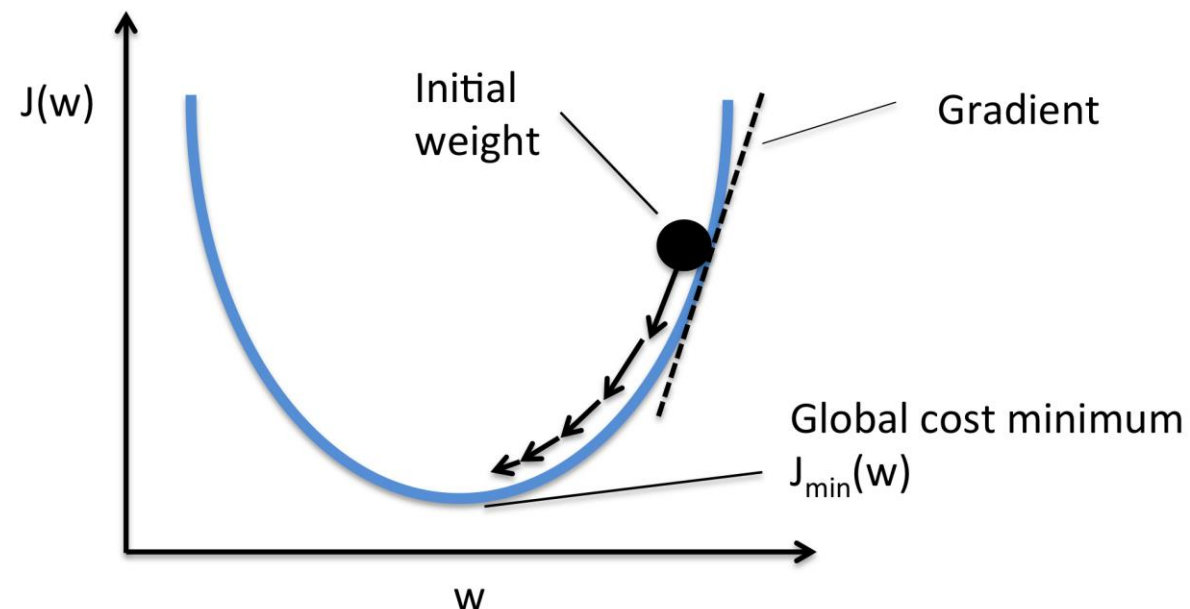
En este caso el valor del parámetro  $W$  se busca utilizando métodos de optimización, por ejemplo usando Gradiente Descendente.

En este caso los pesos se actualizan gradualmente después con base en la función de costo, denominada  $J(\cdot)$ , que es la suma de los errores al cuadrado.

En este caso, la magnitud y la dirección de la actualización de  $W$  se calcula dando un paso en la dirección opuesta del gradiente de la función de costo:

$$\Delta w_j = -\eta \frac{\partial J}{\partial w_j}$$

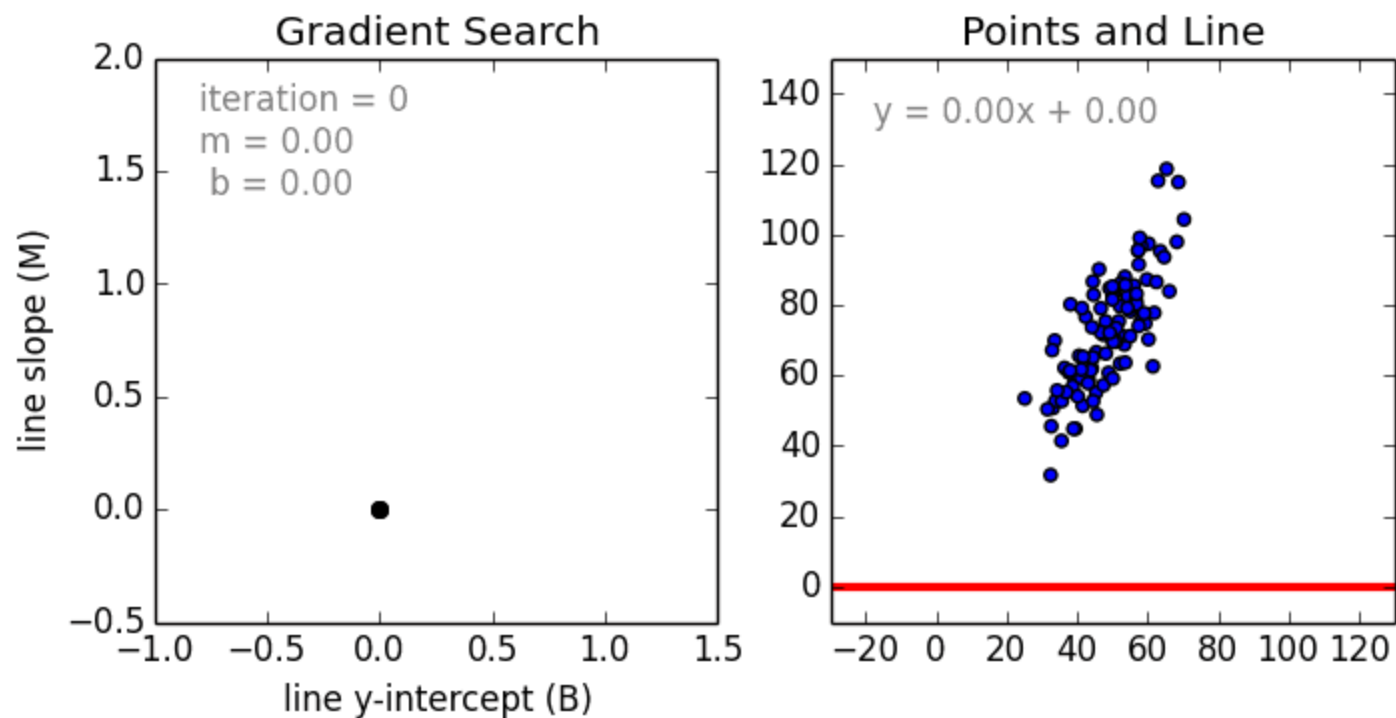
Siendo  $\eta$  la razón (o tasa) de aprendizaje.



# Regresión Lineal

## SOLUCIÓN POR EL MÉTODO DE OPTIMIZACIÓN:

En este caso el valor del parámetro **W** se busca utilizando métodos de optimización, por ejemplo usando Gradiente Descendente.





VAMOS A CODIFICAR!



¿QUÉ ES LA REGRESIÓN LOGÍSTICA?

# Regresión Logística

## 🦋 GENERALIDADES

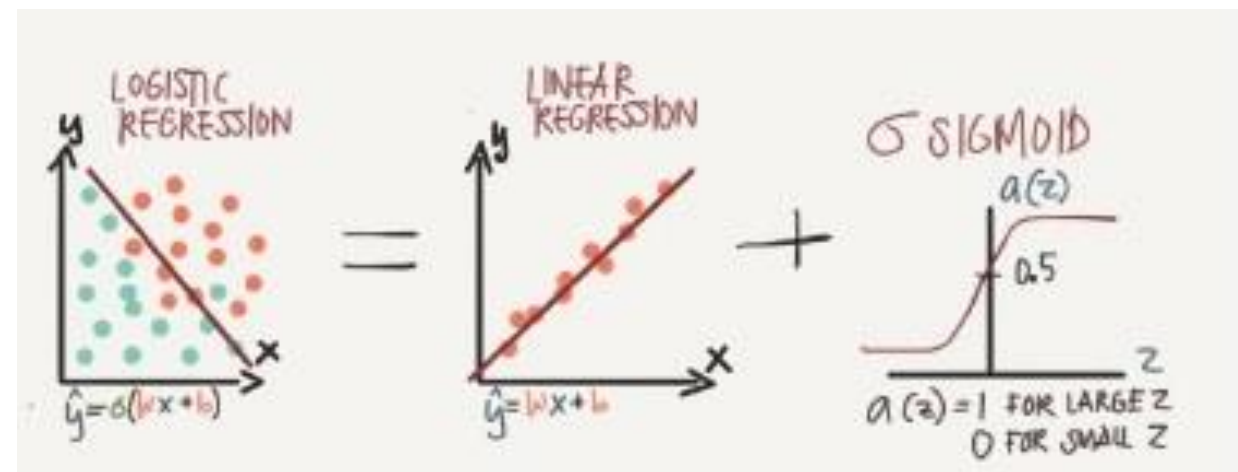
La regresión logística es un modelo de clasificación que se utiliza para **predecir la probabilidad** de que se presente el evento (o valor) 1 **de una variable dependiente categórica** ( $y$ ) en función de las variables independientes  $x$ . En otras palabras, el modelo busca predecir

$$p(y = 1|x)$$

Donde, la variable  $y$  representa la ocurrencia o no de un suceso, es decir, esta se codificada como 1 cuando ocurre el suceso (positivo, éxito, etc.) o 0 cuando no ocurre (negativo, falla, etc.).

## Algunos ejemplos de aplicación:

- E-mail: spam/no spam
- Transacciones en línea: fraude/no fraude
- Tumores: maligno/no maligno



# Regresión Logística

## ▪ Funcionamiento:

El modelo de regresión logística establece la siguiente relación:

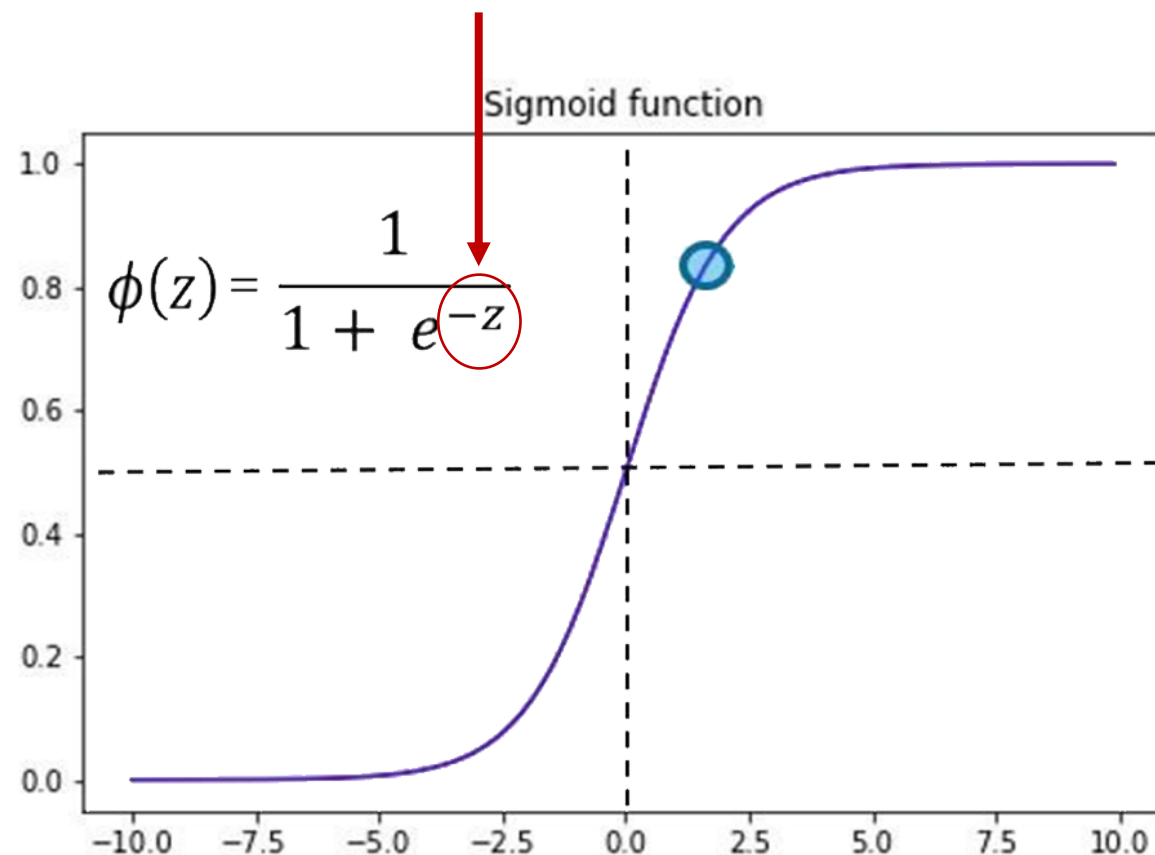
$$p(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(w_0 + w_1x^1 + w_2x^2 + \dots + w_dx^d)}}$$

Esta relación escrita de otra forma es:

$$p(y = 1|\mathbf{x}) = \phi(z) = \frac{1}{1 + e^{-z}}$$

donde  $z$  es la combinación lineal entre los pesos ( $w_i$ ) y las características ( $x^i$ ), es decir,  $z = \mathbf{W}^T \mathbf{x}$ .

Esta es la salida del modelo de regresión lineal





# Regresión Logística

## ■ Funcionamiento:

Usando la probabilidad de que ocurra el evento y la función logit se obtiene el modelo de regresión lineal:

$$\text{logit}(\phi(z)) = \log\left(\frac{\phi(z)}{1-\phi(z)}\right) = w_0 + w_1x^1 + w_2x^2 + \dots + w_dx^d$$

## ■ Ejemplo:

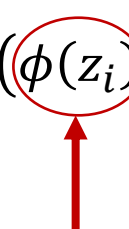
Se debe determinar si una persona tiene o no cáncer con base en el tamaño de un tumor.

- Variable de salida: tiene o no cáncer (salida binaria)
- Variable de entrada: el tamaño del tumor
- El modelo debe aprender la relación entre tener cáncer y el tamaño de un tumor como relación lineal que define la  $p(y = 1)$ . Es decir, el modelo aprende a estimar una probabilidad con base en las entradas.

# Regresión Logística

## ▪ Funcionamiento:

Para estimar los coeficientes  $\mathbf{W}$  se debe minimizar una función de costo. En nuestro caso la función de costo, interpretada como la suma de los errores, corresponde al logaritmo de la verosimilitud que es:

$$J(\mathbf{W}) = \sum_{i=1}^n -y_i \log(\phi(z_i)) - (1 - y_i) \log(1 - \phi(z_i))$$


Es la predicción hecha por el modelo

$$P(y = 1|\mathbf{x};\mathbf{W})$$

Este error se puede ver como la diferencia que hay entre la clase estimada y la etiqueta real de cada observación.

# Regresión Logística

## VENTAJAS:

- Es un modelo de clasificación eficiente y simple
- No es necesario disponer de grandes recursos computacionales
- Los resultados son altamente interpretables



## DESVENTAJAS:

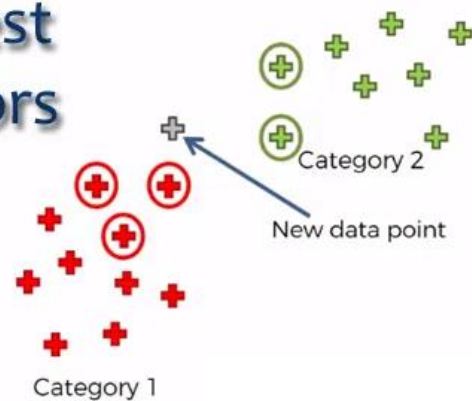
- Imposibilidad de resolver directamente problemas no lineales.
- La variable objetivo esta ha de ser linealmente separable.
- La regresión logística no es uno de los algoritmos más potentes que existen.





VAMOS A CODIFICAR!

## K-Nearest Neighbors






## CLASIFICADOR DE LOS K-VECINOS MÁS CERCANOS



# Clasificador kNN

## LA BASE DEL CLASIFICADOR

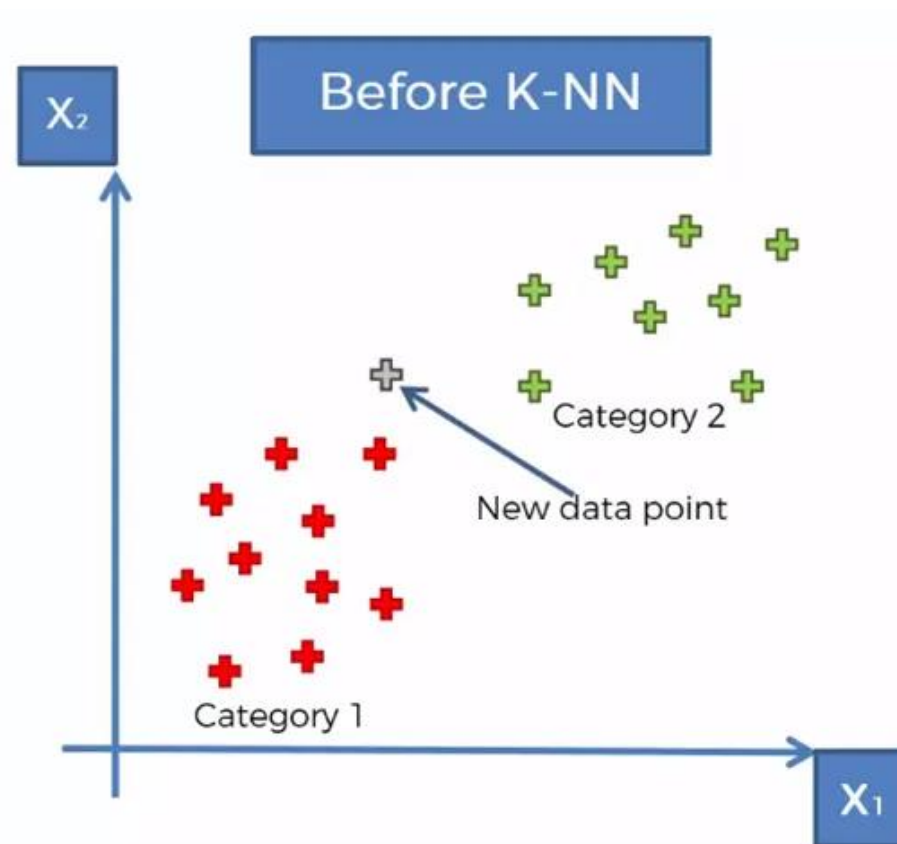
-  Este es un clasificador basado en las instancias en que el no se aprende o se crea un modelo generalizado de los datos, sino que se crean modelos locales a demanda:
  -  El clasificador simplemente almacena los datos de entrenamiento y el proceso de clasificación se realiza a partir de un voto mayoritario simple entre los  $k$ -vecinos del vecindario ( $V$ ) de un punto  $P$ . Es decir, se cuentan el número  $n$  de vecinos de  $P$  que pertenecen a cada clase y se escoge la clase con más votos.
  -  Esta regla basa su operación en el supuesto de que la clase a la que pertenece un punto  $P$  es la que tienen la mayor probabilidad de aparición en el vecindario de  $P$ .

# Clasificador kNN



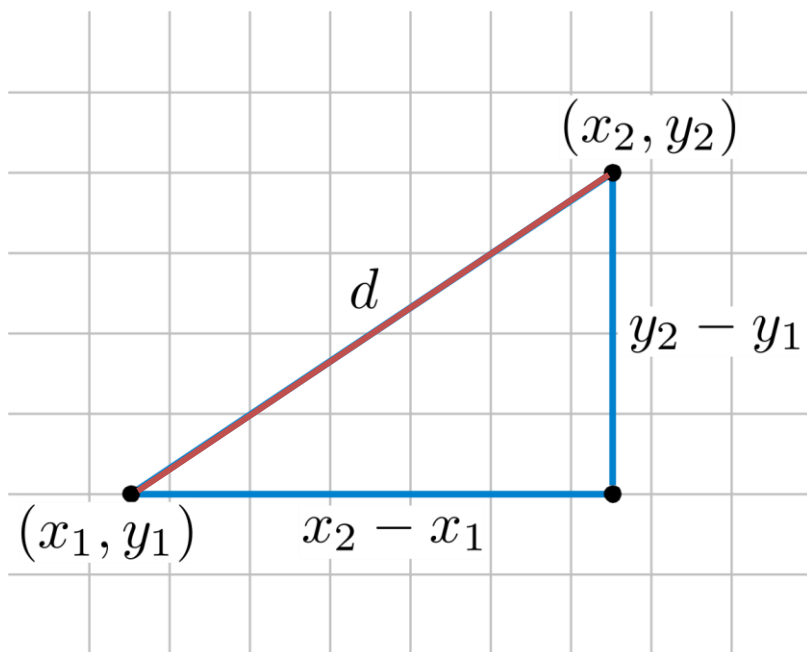
## ALGORITMO DE LOS K VECINOS MÁS CERCANOS

1. Se toma el punto P a evaluar (+) y se calcula la distancia de este a cada uno de los puntos en el conjunto de datos de entrenamiento.
2. Se seleccionan los k puntos más cercanos, es decir aquellos con menor distancia (según la función de distancia usada)
3. Se realiza una votación y se asigna a P la etiqueta de clase que predomine entre los k-vecinos



# Clasificador kNN

## ALGORITMO DE LOS K VECINOS MÁS CERCANOS – MÉTRICAS DE DISTANCIA



**DISTANCIA EUCLIDIANA:** distancia de la línea recta que conecta a los dos puntos

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

**DISTANCIA MANHATTAN:** distancia por bloques entre los puntos

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

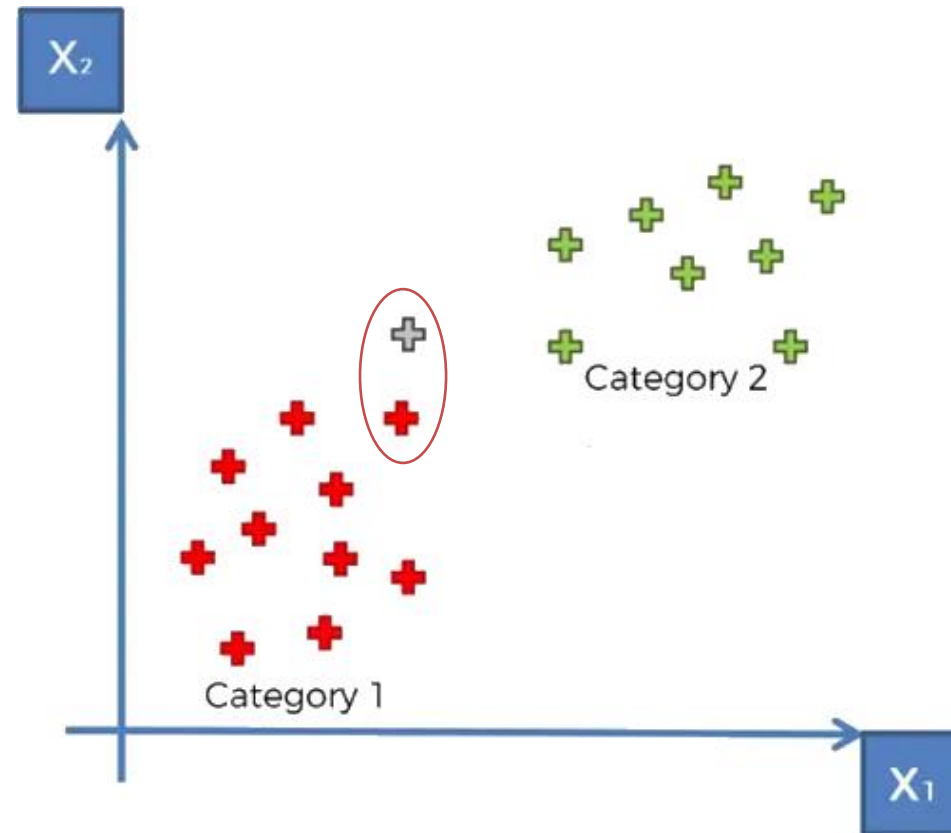
**DISTANCIA MINKOWSKI:** Es una generalización de las distancias

$$d = \left( \sum_{i=1}^n |x_2^i - x_1^i|^p \right)^{\frac{1}{p}}$$



# Clasificador kNN

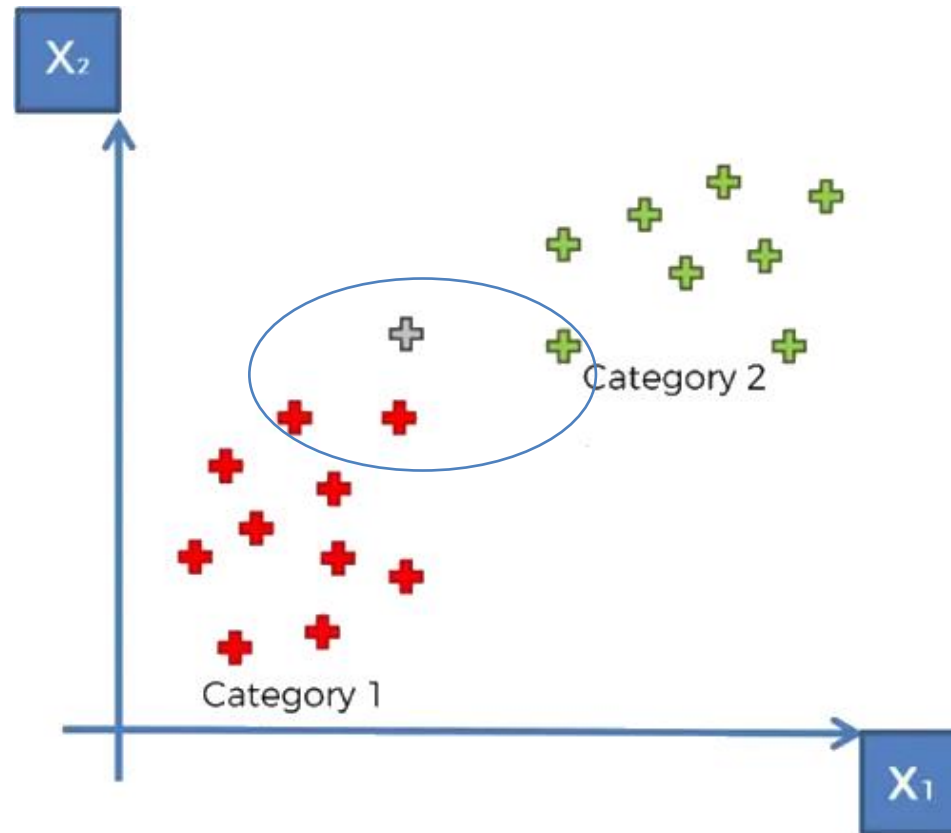
USANDO 1 VECINO



# Clasificador kNN

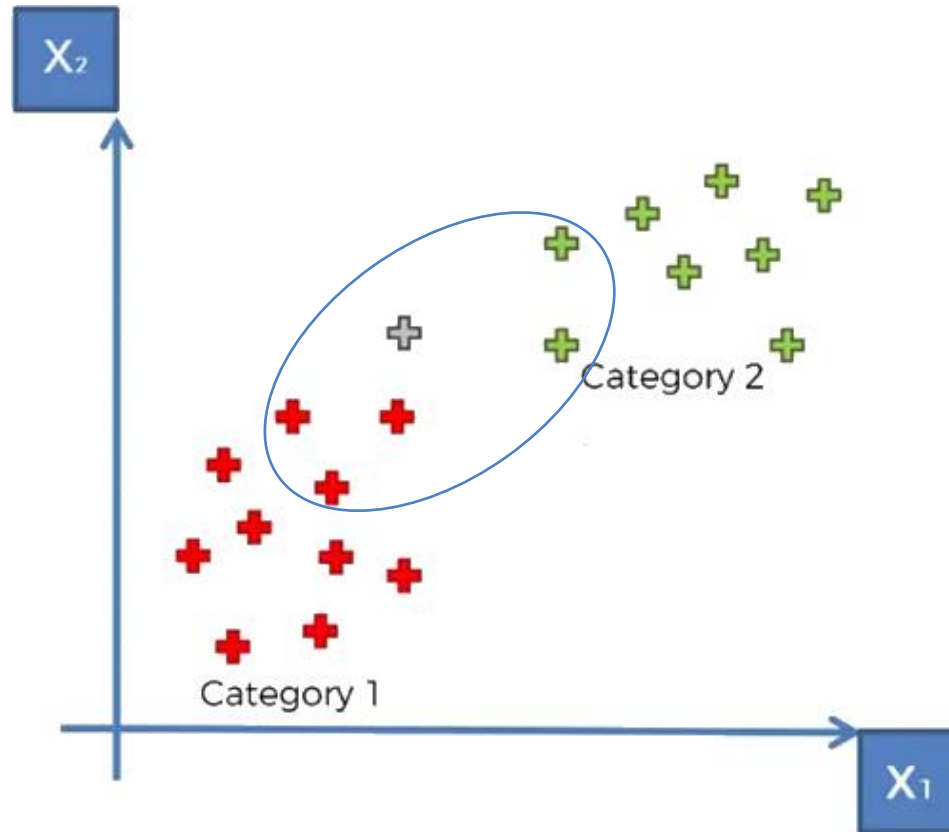


USANDO 3 VECINOS



# Clasificador kNN

USANDO 5 VECINOS



# Clasificador kNN

## VENTAJAS:

- Es clasificador bastante simple
- Puede lidiar con fronteras de decisión complejas y de formas arbitrarias
- Se ha demostrado que la precisión de este clasificador puede ser bastante fuerte y en muchos casos tan precisa como otros métodos elaborados
- En un clasificador **no paramétrico** por los que no hace suposiciones explícitas sobre la distribución subyacente de los datos
- Suele ser insensible a los valores atípicos



## DESVENTAJAS:

- Tiene requisito de memoria alta porque almacena todos los datos de entrenamiento
- Es computacionalmente costoso (lento) porque se deben calcular las distancias a todos los puntos del conjunto de entrenamiento con cada uso
- El “modelo” que genera el clasificador no es interpretable





VAMOS A CODIFICAR!