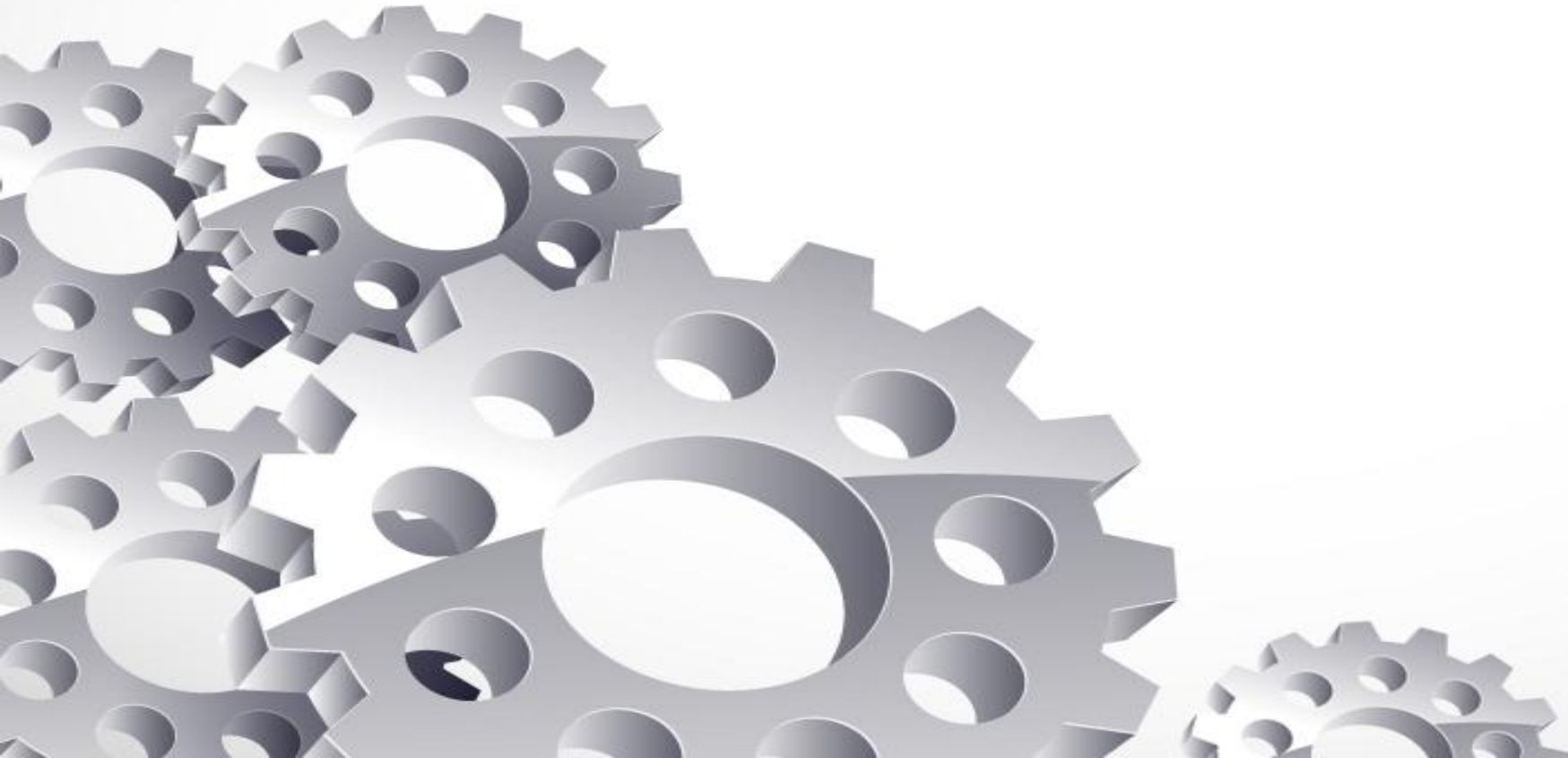


Machine Learning

Isabel Maniega





Regresión

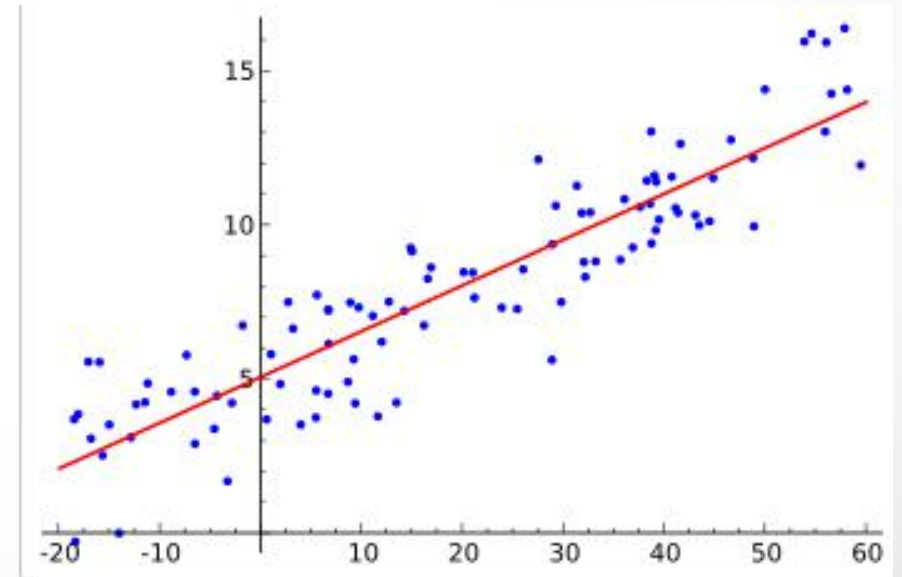
Regresión Lineal Simple

- Basado en la ecuación de la recta:

Variable dependiente $y = b + mx$ *Variable Independiente*

Constante b *Coefficiente* m

Ir al ejemplo práctico Regresión Lineal Simple

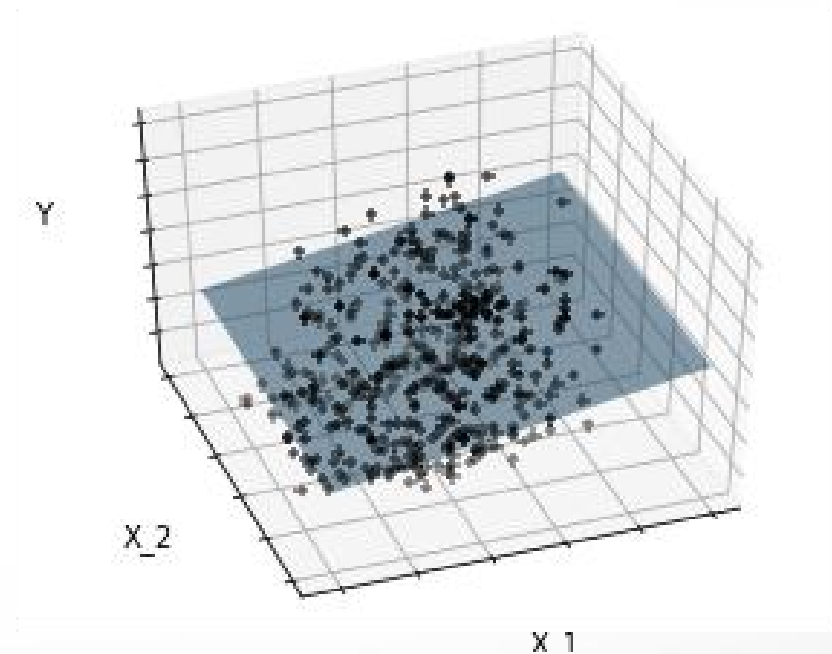


Regresión Lineal Multiple

- Basado en la ecuación:

$$y = b_0 + m_1 x_1 + m_2 x_2 + \dots + m_n x_n$$

Ir al ejemplo práctico Regresión Lineal Multiple

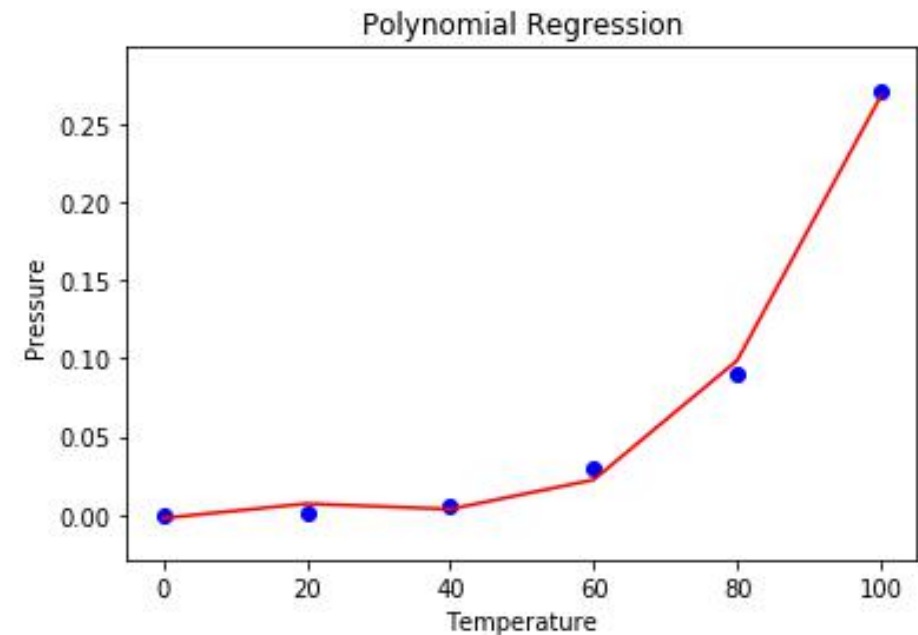


Regresión Lineal Polinomial

- Basado en la ecuación:

$$y = b_0 + m_1 x_1 + m_2 x_1^2 + \dots + m_n x_1^n$$

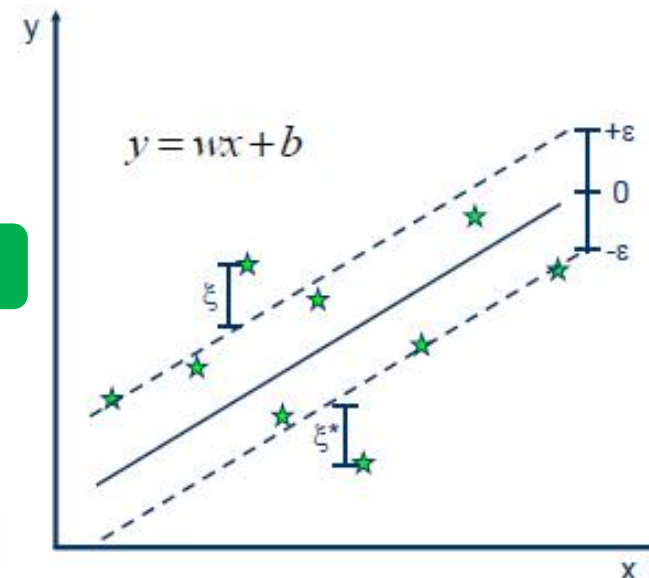
Ir al ejemplo práctico Regresión Lineal Polinomial



Support Vector Regression (SVR)

- Creamos una recta que pase entre los puntos, entonces se establece un margen de tolerancia (epsilon) cerca del vector con el fin de minimizar el error tomando en cuenta que parte de ese error es tolerado

Ir al ejemplo práctico Support Vector Regression



- Minimize:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

- Constraints:

$$y_i - wx_i - b \leq \epsilon + \xi_i$$

$$wx_i + b - y_i \leq \epsilon + \xi_i^*$$

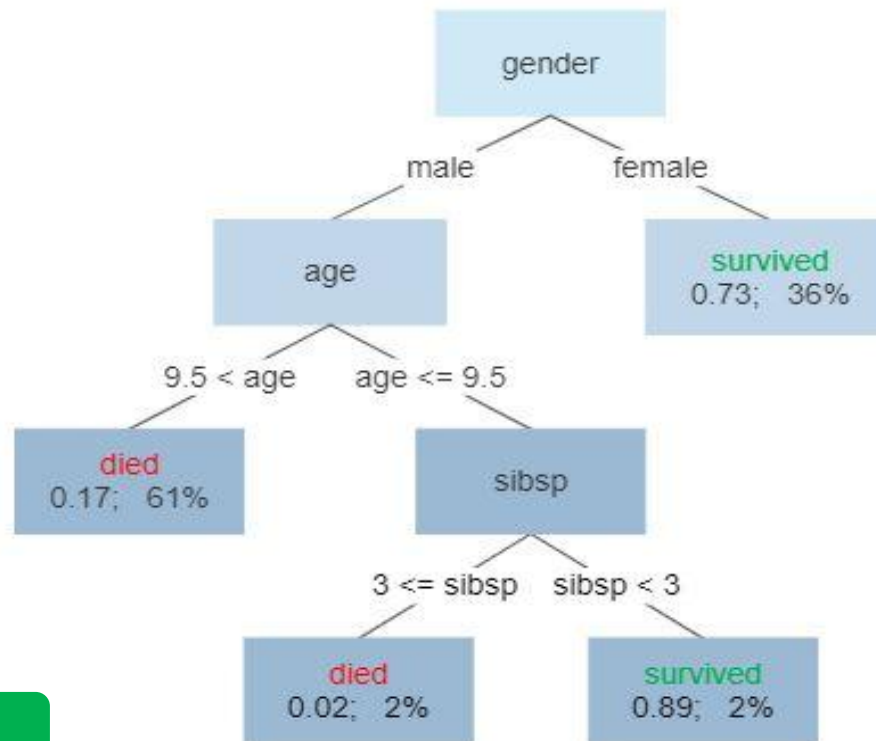
$$\xi_i, \xi_i^* \geq 0$$

Decission Tree Regression

- Es en un proceso de dividir los datos de entrada, este es un procedimiento numérico en el que se alinean todos los valores y se prueban diferentes puntos de división utilizando distintos métodos. error es tolerado

Ir al ejemplo práctico Decission Tree Regression

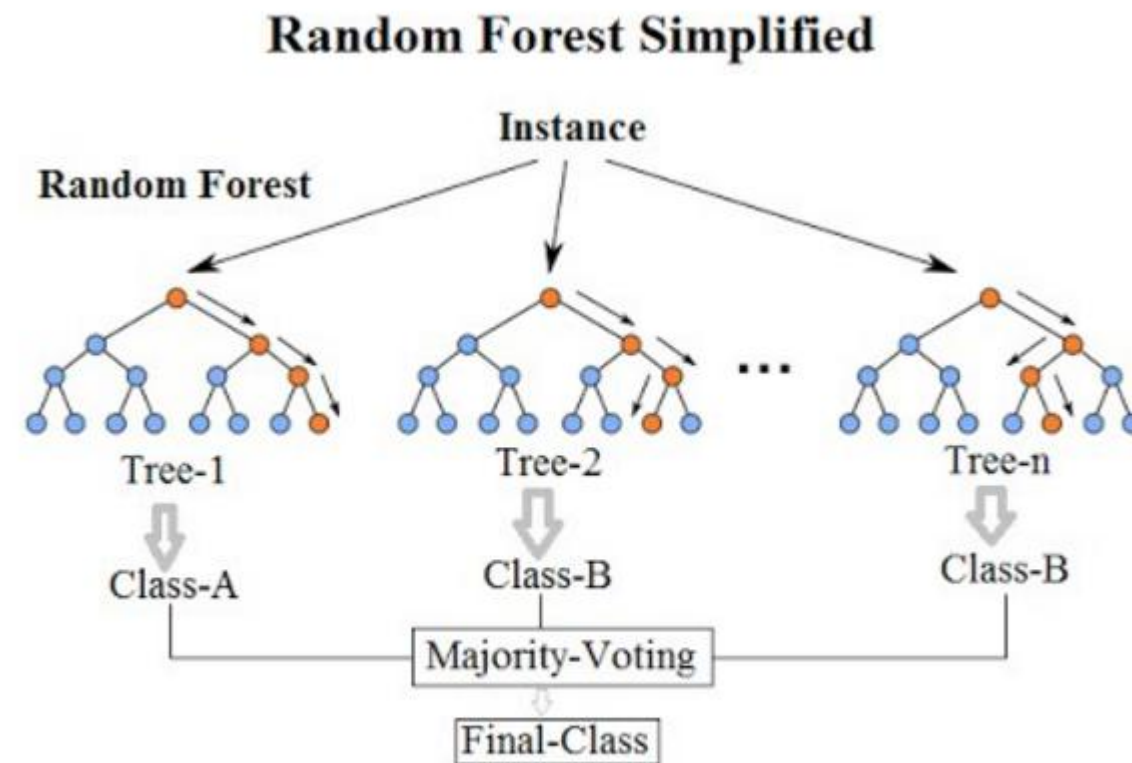
Survival of passengers on the Titanic



https://en.wikipedia.org/wiki/Decision_tree_learning.

Random Forest Regression

- Es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos



Ir al ejemplo práctico Random Forest Regression



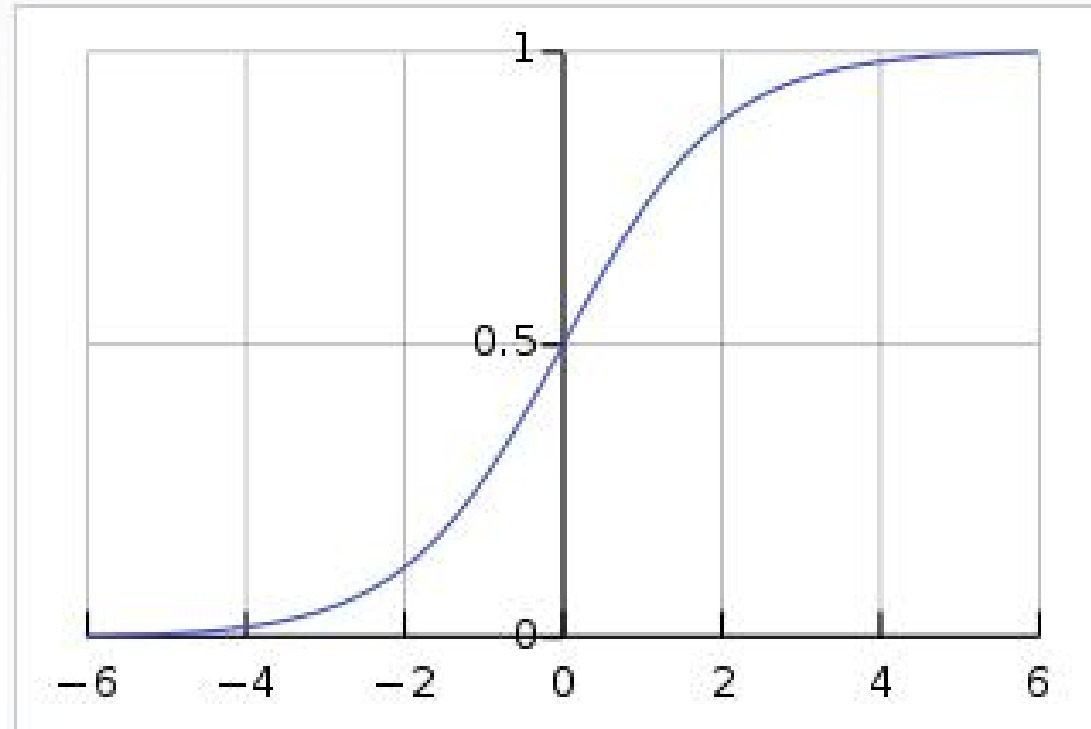
Clasificación

Logistic Regression



- Es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras.
- Trata de correlacionar la probabilidad de una variable cualitativa binaria (asumiremos que puede tomar los valores reales "0" y "1") con una variable escalar x . La idea es que la regresión logística aproxime la probabilidad de obtener "0" (no ocurre cierto suceso) o "1" (ocurre el suceso) con el valor de la variable explicativa x . En esas condiciones, la probabilidad aproximada del suceso se aproximará mediante una función logística

Logistic Regression



Función logística con $\beta_0 + \beta_1 x + e$ en el eje horizontal y $\pi(x)$ en el eje vertical.

Ir al ejemplo práctico Logistic Regression

https://es.wikipedia.org/wiki/Regresi%C3%B3n_log%C3%ADstica

K-Nearest Neighbors

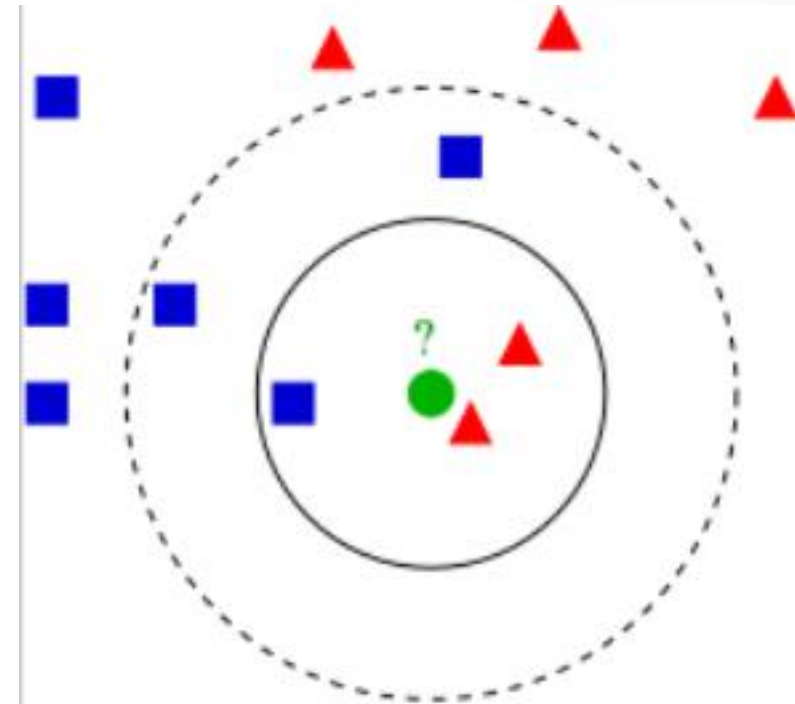


- Este es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C_j a partir de la información proporcionada por el conjunto de prototipos.

K-Nearest Neighbors

- Ejemplo del algoritmo Knn. El ejemplo que se desea clasificar es el círculo verde. Para $k = 3$ este es clasificado con la clase triángulo, ya que hay solo un cuadrado y 2 triángulos, dentro del círculo que los contiene. Si $k = 5$ este es clasificado con la clase cuadrado, ya que hay 2 triángulos y 3 cuadrados, dentro del círculo externo.

Ir al ejemplo práctico K-Nearest Neighbors

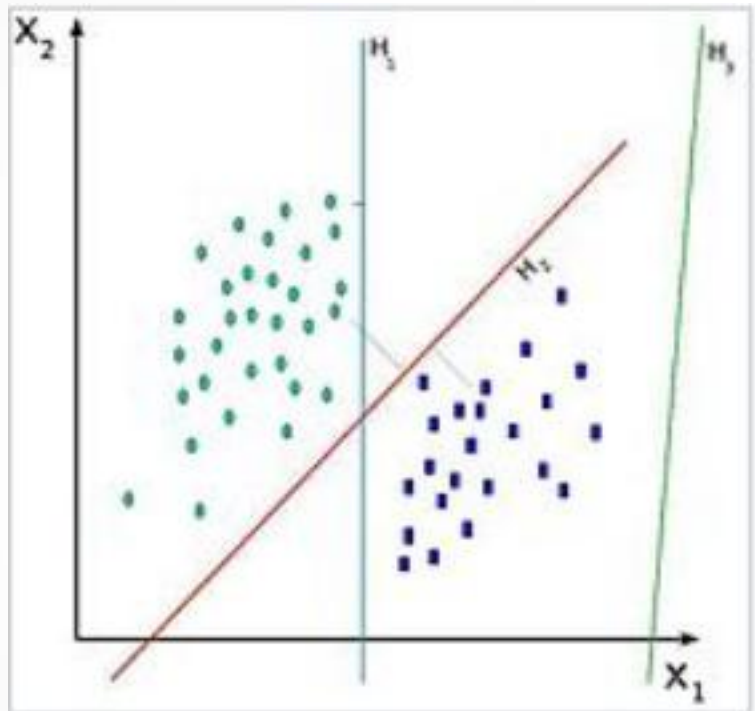


Support vector Machine (SVM)

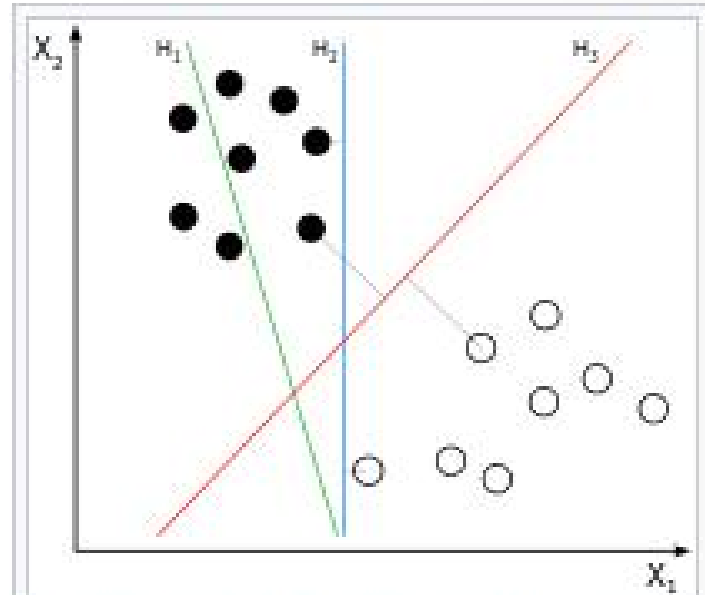


- Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases, más cercanos al que se llama vector soporte. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o la otra clase.

Support Vector Machine (SVM)



Hay infinitos hiperplanos posibles



H_1 no separa las clases. H_2 las separa, pero solo con un margen pequeño. H_3 las separa con el margen máximo.

Ir al ejemplo práctico Support Vector Machine

https://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte

Naives Bayes



- Basado en el Teorema de Bayes con una suposición de independencia entre los predictores.
- Un clasificador de Naive Bayes asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable

[Ir al ejemplo práctico Naives Bayes](#)

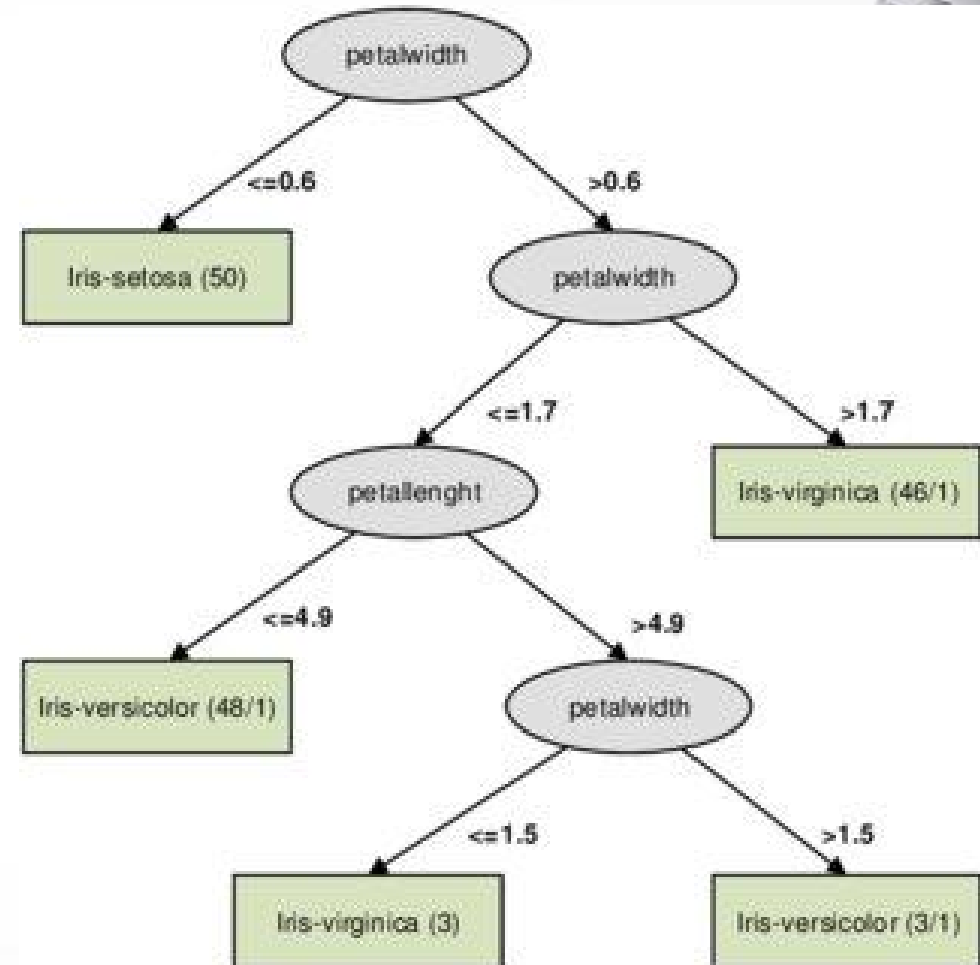
Decission Tree Classification



- Los modelos de árbol, donde la variable de destino puede tomar un conjunto finito de valores se denominan árboles de clasificación. En estas estructuras de árbol, las hojas representan etiquetas de clase y las ramas representan las conjunciones de características que conducen a esas etiquetas de clase. Los árboles de decisión, donde la variable de destino puede tomar valores continuos (por lo general números reales) se llaman árboles de regresión

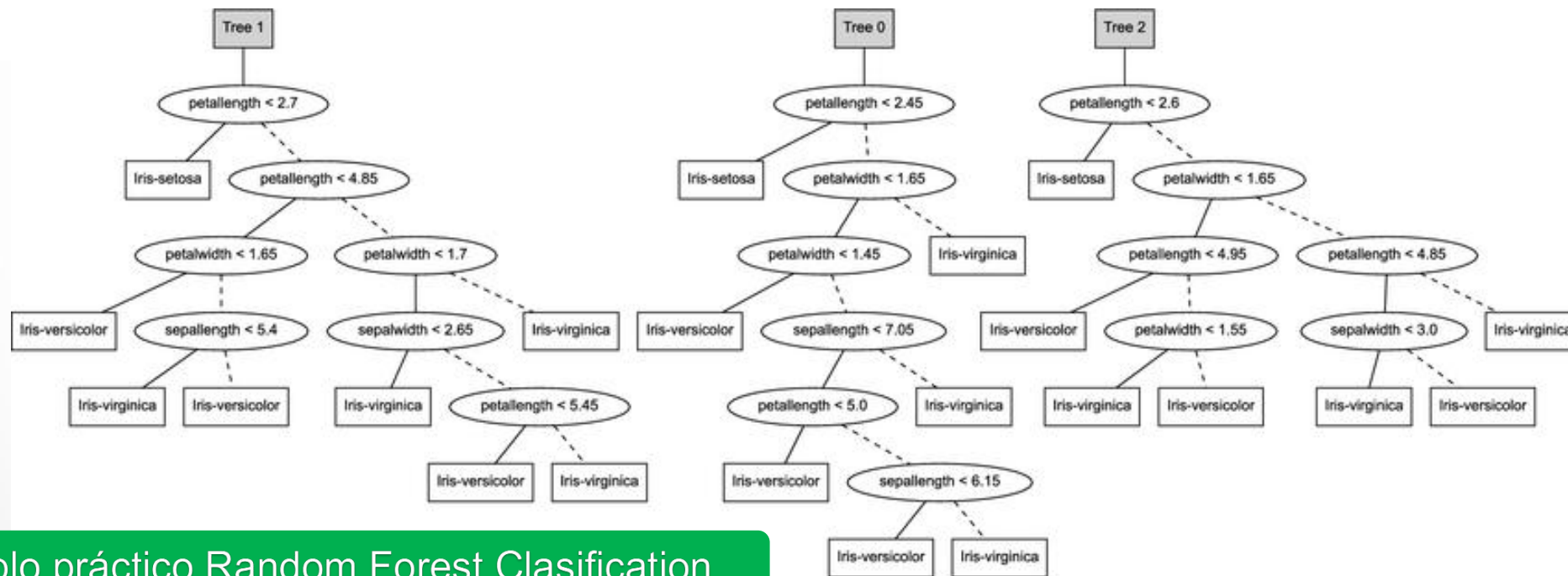
Decission Tree Classification

Ir al ejemplo práctico Decission Tree Clasification



Random Forest Classification

- Es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.



Ir al ejemplo práctico Random Forest Clasificación

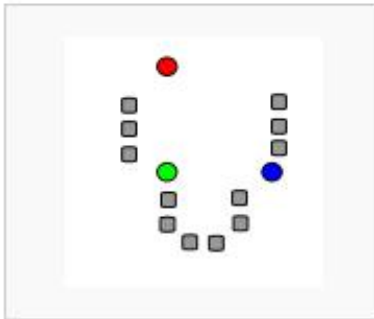


Clustering

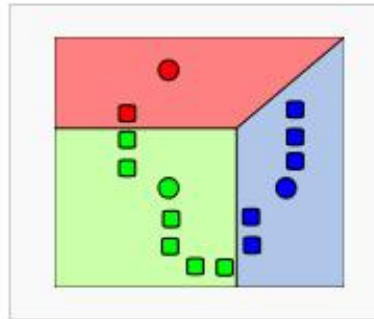
K-Means Clustering

- Es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.

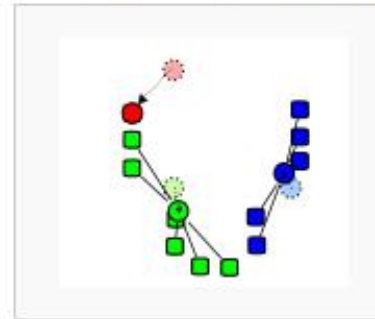
Demostración del algoritmo estándar



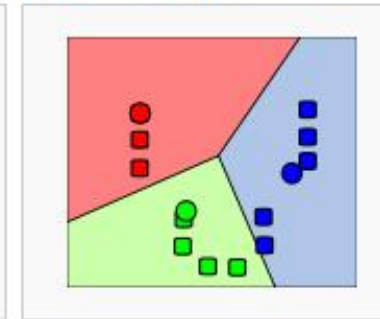
1) k centroides iniciales (en este caso $k=3$) son generados aleatoriamente dentro de un conjunto de datos (mostrados en color).



2) k grupos son generados asociándole el punto con la media más cercana. La partición aquí representa el [diagrama de Voronoi](#) generado por los centroides.

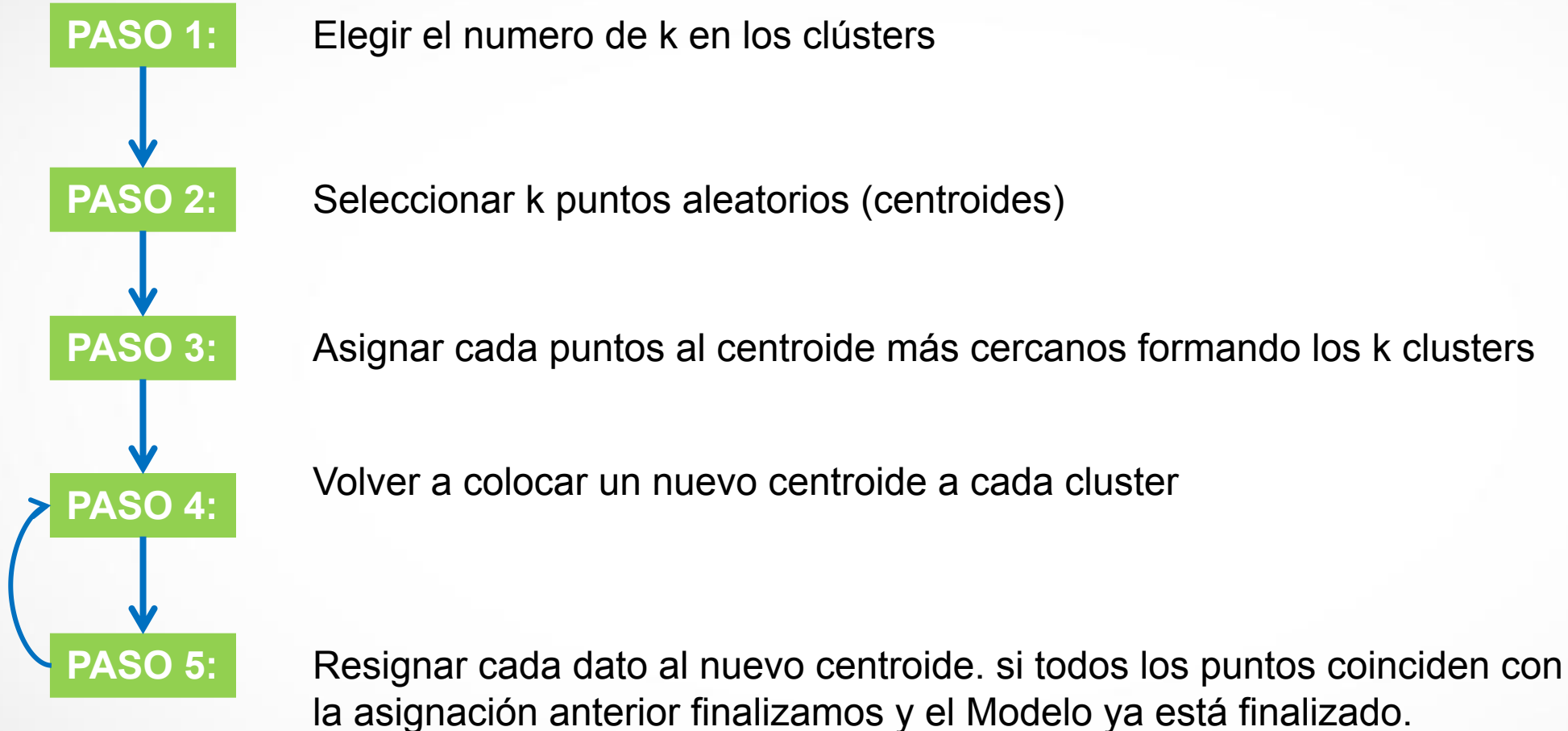


3) EL [centroide](#) de cada uno de los k grupos se recalcula.



4) Pasos 2 y 3 se repiten hasta que se logre la convergencia.

K-Means Clustering



[Ir al ejemplo práctico K-Means Clustering](#)

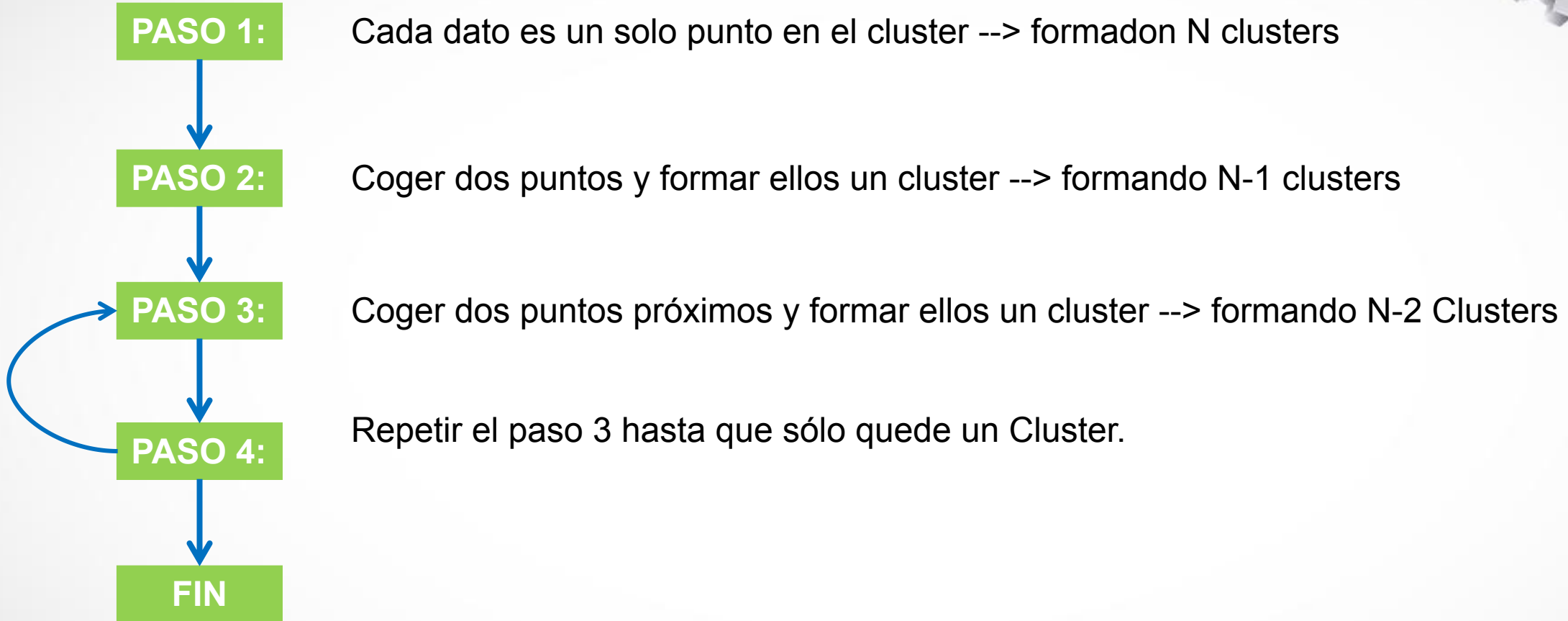
Hierarchical Clustering



Es un método de análisis de grupos puntuales, el cual busca construir una jerarquía de grupos. Estrategias para agrupamiento jerárquico generalmente caen en dos tipos:

- **Aglomerativas:** Este es un acercamiento ascendente: cada observación comienza en su propio grupo, y los pares de grupos son mezclados mientras uno sube en la jerarquía.
- **Divisivas:** Este es un acercamiento descendente: todas las observaciones comienzan en un grupo, y se realizan divisiones mientras uno baja en la jerarquía.

Hierarchical Clustering



[Ir al ejemplo práctico Hierarchical Clustering](#)

A decorative graphic in the bottom-left corner of the slide. It features several interlocking gears of different sizes, rendered in a light gray, 3D style with soft shadows. The gears are arranged in a cluster, with some overlapping others, creating a mechanical or industrial aesthetic.

Association Rule Learning

Associate Rule Learning

- Las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos
- Ejemplo en los datos de ventas de un supermercado, indicaría que un consumidor que compra cebollas y verdura a la vez, es probable que compre también carne.



Apriori Algoritmo

Procede identificando los ítems individuales frecuentes en la base y extendiéndolos a conjuntos de mayor tamaño siempre y cuando esos conjuntos de datos aparezcan suficientemente seguidos en dicha base de datos.

PASO 1:

Elegir la frecuencia para cada uno de los ítems



PASO 2:

Coger todos los subconjuntos que forman cada uno de las distintas combinaciones de ítems y su frecuencia



PASO 3:

Coger todos los subconjuntos aumentando un ítem cada vez y ver su frecuencia



PASO 4:

Repetir el paso 3 hasta que sólo quede un Cluster con los ítems con mayor frecuencia.

Apriori Algoritmo



PASO 1:

Conjuntos:

{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}

PASO 2:

| Ítem | Confianza |
|------|-----------|
| {1} | 3 |
| {2} | 6 |
| {3} | 4 |
| {4} | 5 |

PASO 3:

| Ítem | Confianza |
|-------|-----------|
| {1,2} | 3 |
| {1,3} | 1 |
| {1,4} | 2 |
| {2,3} | 3 |
| {2,4} | 4 |
| {3,4} | 3 |

PASO 4:

| Ítem | Confianza |
|---------|-----------|
| {2,3,4} | 2 |

[Ir al ejemplo práctico Apriori Algoritmo](#)

Market Basket Analysis

Análisis de la cesta de mercado, que se usa para observar el comportamiento de compra de los consumidores en los supermercados.

Se basa en el el algoritmo Apriori:

PASO 1:

Elegir la frecuencia para cada uno de los ítems

PASO 2:

Coger todos los subconjuntos que forman cada uno de las distintas combinaciones de ítems y su frecuencia

PASO 3:

Coger todos los subconjuntos aumentando un ítem cada vez y ver su frecuencia

PASO 4:

Repetir el paso 3 hasta que sólo quede un Cluster con los ítems con mayor frecuencia.

[Ir al ejemplo práctico Market Basket Analysis](#)