

Aprendizaje No Supervisado

Dra. Marcela Riccillo



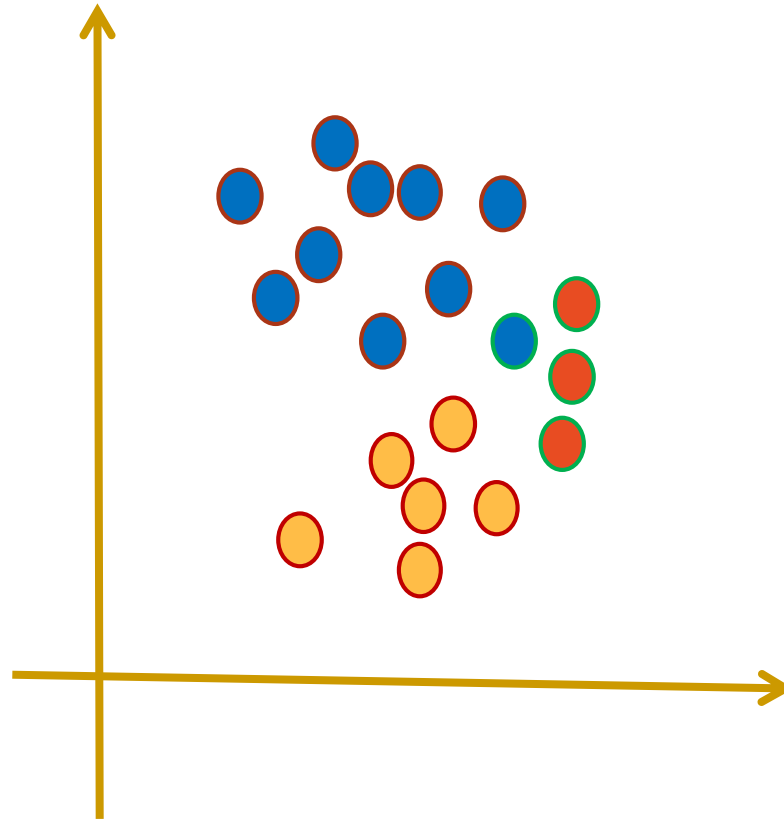
En el Aprendizaje Supervisado existe una variable a predecir

V1	V2	V3	V4	Sí/NO
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X

Variable a
Predecir

En el Aprendizaje No Supervisado no hay variable a predecir

La idea es que el algoritmo arme grupos con los registros



Aprendizaje No Supervisado

Jerárquico

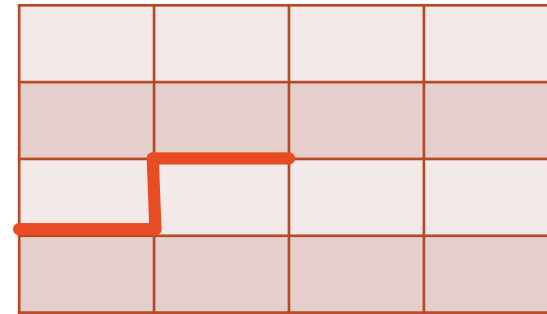
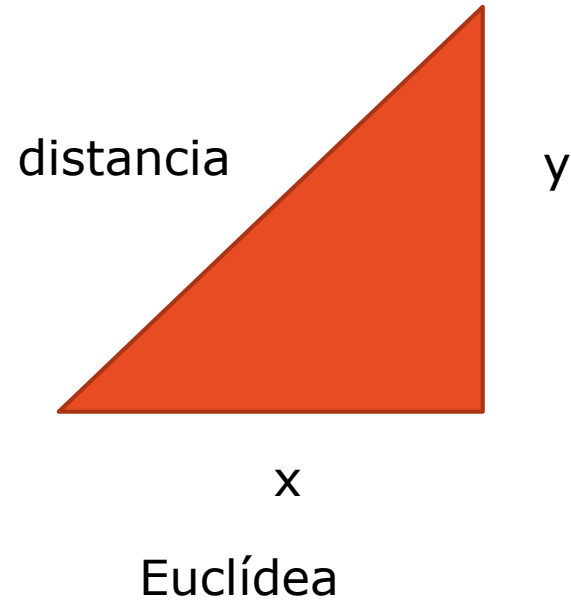
Dendograma

No Jerárquico

k-Means

Distancias

Ejemplos



Manhattan

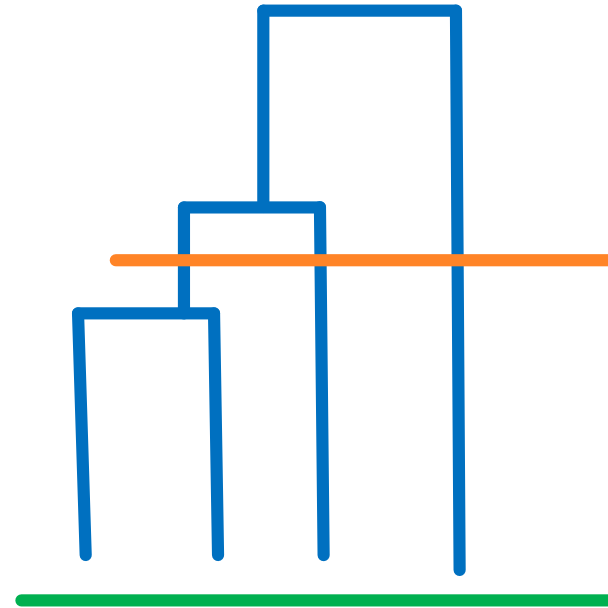
Dendograma

El método de Dendograma corresponde a un Aprendizaje No Supervisado Jerárquico

Primero se calcula la matriz de distancia entre cada uno de los puntos

Luego se van haciendo grupos intermedios con los puntos según metodologías como "vecino más cercano", "vecino más lejano", "vecino promedio"

Y luego se recalcula la matriz de distancias con los grupos que se van formando



k- Means

Se determinan la cantidad k de grupos a obtener

El algoritmo inicializa k centroides al azar

Luego recalcula los grupos según los puntos más cercanos a cada centroide

Luego calcula los nuevos centroides de cada grupo

Vuelve a rearmar los grupos y calcular los nuevos centroides hasta que los mismos ya no cambian

Práctica (1/5)

¿Cómo se calcula la distancia Euclídea entre dos puntos?

Práctica - Dendograma (2/5)

Dada una base:

Mostrar un resumen de las características de los datos

¿Qué tipos de variables se pueden identificar?

¿Cuántas variables tiene la base? ¿Cuántos registros?

Obtener la matriz de distancias

Obtener el dendograma

Dibujar el dendograma

¿Cuántos grupos quedaron? ¿Algún registro quedó separado del resto?

Práctica - k-means (3/5)

Hallar una agrupación de k-means de los datos de n grupos

Hallar los valores de los centroides

¿Cuántos registros quedaron en cada grupo? ¿Se detecta algún outlier?

Realizar un gráfico de dispersión de 2 variables coloreado según los grupos formados

Agregar en el gráfico los centroides

Práctica - k-means (4/5)

Analizar y caracterizar los grupos resultantes

Realizar el mismo análisis para diferentes cantidades de grupos

Práctica - k-means (5/5)

Considerar una base con una variable a predecir

Sacar la variable a predecir y realizar un agrupamiento k-means

Comparar los resultados con la clasificación original