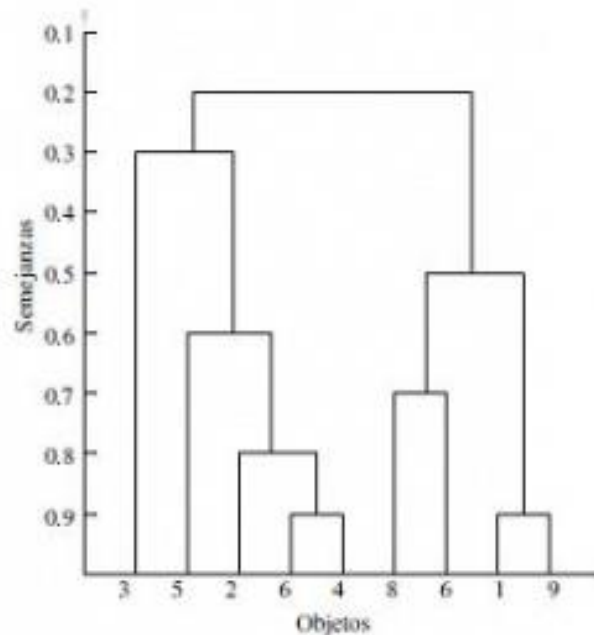


## Lección 8: tipos de algoritmos de agrupamiento y reglas de asociación

Dentro del aprendizaje no supervisado se encuentran diversos tipos de algoritmos que se enmarcan en la premisa de esta clasificación como se mencionaran a continuación:

- **Algoritmos jerárquicos**

Los algoritmos de agrupamiento jerárquicos son utilizados comúnmente para el agrupamiento de patrones de los cuales no se conoce la organización interna que poseen. Es decir, no existe conocimiento sobre la etiqueta de clase a la cual son pertenecientes. Estos algoritmos trabajan a través de la unión o división en cada paso o iteración, el par de grupos más semejante.



Fuente: Ecured

Los algoritmos jerárquicos generan una secuencia anidada de particiones del conjunto de objetos, lo que quiere decir que, los grupos de organización se forma jerárquica y cada uno de los grupos o clúster, puede tomarse como la unión entre otros grupos, logrando distintos niveles de jerarquía de grupos. La organización jerárquica es generalmente representada por un árbol llamado **dendrograma**, el cual proporciona una taxonomía de la información procesada.

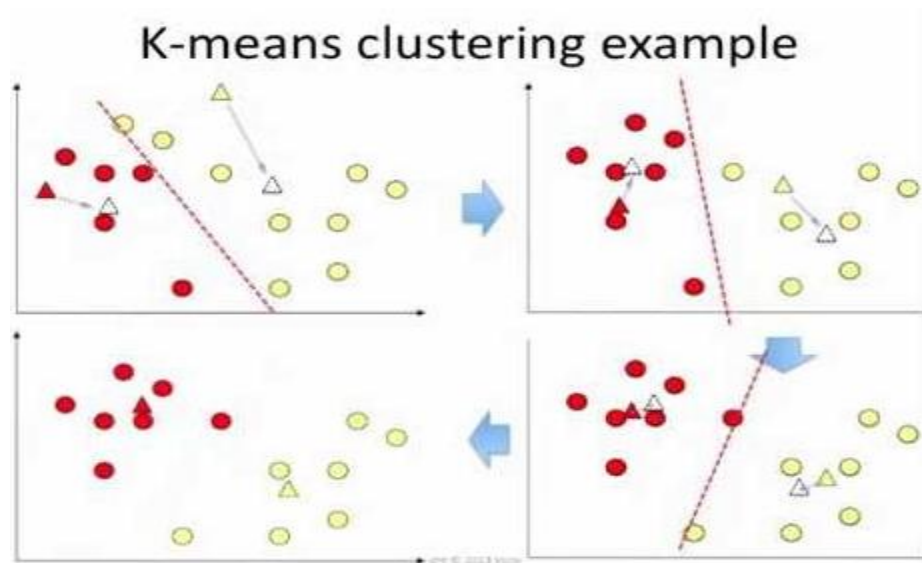
Debido a la forma en que se crea la jerarquía, los algoritmos se clasifican en:

1. Divisivos: parten a raíz de la consideración del conjunto de objetos como un grupo, y en cada una de las iteraciones fraccionan un grupo en dos, hasta que permanezcan tantos grupos como objetos individuales existentes, o hasta que cumplan con el criterio de parada.
2. Aglomerativos: son basados en distancias. Comienzan con la consideración de cada objeto como grupos unitarios y en cada iteración se acoplan los dos grupos más contiguos hasta que se obtenga un único grupo, o hasta que se cumpla un criterio de parada.

Estos algoritmos se diversifican entre sí, por la forma en cómo se calculan los trayectos entre grupos.

- **K-medias**

El algoritmo es aplicable en aquellos casos en que se tenga una inmersión de ejemplos en un espacio geométrico. Este algoritmo pretende encontrar una división de ejemplos en K agrupaciones, de manera que cada uno de los ejemplos concierna a una de ellas, específicamente a aquella cuyo valor geométrico este más colindante.



Fuente: CS

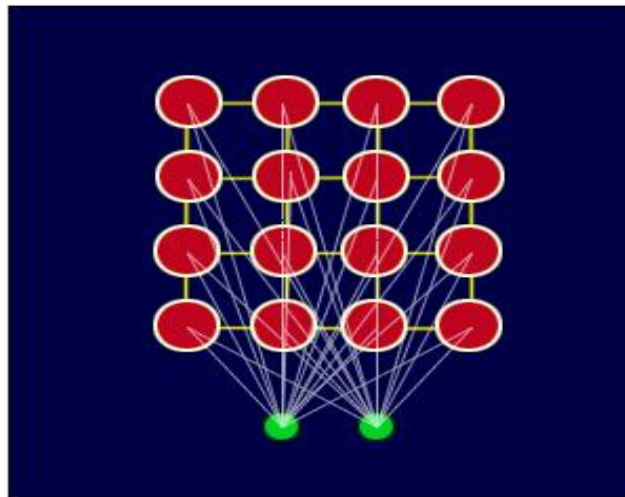
El mejor valor de K para que la clasificación sea separada de la manera más óptima posible con sus ejemplos, no se conoce a priori y depende básicamente de los datos con los que se trabaje. El algoritmo pretende minimizar la varianza total del sistema.

El algoritmo por seguir es:

1. Seleccionar al azar K puntos como centros de grupos
2. Asignar los ejemplos al centro más cercano
3. Calcular el centro geométrico de los ejemplos asociados a cada grupo
4. Repetir el paso 2 hasta que no exista resignación de centros

El algoritmo es eficiente y usualmente se requieren de pocos pasos para que el proceso pueda ser estabilizado. Pero, por otra parte, es necesario conocer el número de gremios a priori, y el sistema es sensible a la posición inicial de los K centros haciendo que no consigan un mínimo global, sino que se ubique en un mínimo local.

- **Mapas auto-organizados**



Fuente: CS

Los mapas auto-organizados, o SOM, brindan una forma de representar datos multidimensionales en espacios de dimensión interior, generalmente en 2D. El sistema para reducir la dimensionalidad de vectores es una estrategia de comprensión de datos que se conoce como cuantización vectorial. La técnica almacena la información de manera que las relaciones topológicas del conjunto de entrenamiento sean mantenidas.

Uno de los aspectos más interesantes entorno a los SOM, es el hecho de que aprenden a clasificar sin supervisión, por lo que no es necesario un objetivo que aproximar, sino que crea la distribución en base a la semejanza entre vectores. Al no existir objetivo de aproximación, en las zonas en las cuales la red tiene nodos con pesos que concuerdan con vectores de entrenamiento, el resto de nodos de su medio tienden a aproximarse también a ese mismo vector. De esa manera, procura aproximarse a una distribución de pesos constante.

- **Análisis de componentes principales**

La metodología de componentes principales tiene como objetivo la transformación de un conjunto de variables denominadas originales, en un conjunto nuevo de variables designadas componentes principales, que se identifican por estar correlacionadas entre sí, y, asimismo, pueden ordenarse con base a la información que llevan asociada. En este método, cuando mayor sea la varianza, mayor es la cantidad de información que lleva incorporada dicho componente, dicho esto, se elige como primer componente aquel que posea mayor varianza, y el último componente será el de menor varianza.

Este método es considerado como un sistema para la reducción de datos, tratamiento de problemas como la rotación de factores, contrastes, entre otros. Se realiza dentro del análisis factorial que conlleva más formalización. En este aspecto, el método de componentes principales se ubica dentro de la estadística descriptiva.

- **Análisis de componentes independientes**

El análisis de componentes principales es un método clásico de análisis de datos cuya principal aplicación se localiza en el campo de la extracción de características y la comprensión de datos. El objetivo principal de este método es proporcionar un sistema que permita hallar una representación lineal de los datos no gaussianos, de manera que los componentes sean estadísticamente independientes. Un modelo de esta tipología permite obtener la organización fundamental de los datos en muchas aplicaciones, dentro de las que se encuentran la separación de señales. Este análisis es bastante útil en aplicaciones de reconocimiento facial y se objetos.

La tarea real de los usuarios de una tecnología tan avanzada como el machine learning es transformar o traducir los problemas de la sociedad en un idioma comprensible para las maquinas, es decir, en números. Una vez realizada esta acción se inicia el proceso de entrenamiento de la máquina de ML que lograra la solución de estos problemas. A pesar de ello, aún existen dos limitaciones claras para el aprendizaje automatizado, los problemas que conciernen a la creatividad y las emociones siguen siendo un obstáculo.

### **Referencias:**

- Bibing. (s.f.). Bibing. Recuperado el 7 de octubre de 2018, de <http://bibing.us.es/proyectos/abreproy/11088/fichero/Proyecto+Fin+de+Carrera%252F5.pdf>
- CS. (2017). CS. Recuperado el 7 de octubre de 2018, de <http://www.cs.us.es/~fsancho/?e=77>
- De la fuente, S. (2011). Estadística. Recuperado el 7 de octubre de 2018, de [http://www.estadistica.net/Master-Econometria/Componentes\\_Principales.pdf](http://www.estadistica.net/Master-Econometria/Componentes_Principales.pdf)
- Ecured. (s.f.). Ecured. Recuperado el 7 de octubre de 2018, de [https://www.ecured.cu/Algoritmos\\_jer%C3%A1rquicos](https://www.ecured.cu/Algoritmos_jer%C3%A1rquicos)