**CORRELATION CLUSTERING**

El **Clustering** o la clusterización es un proceso importante dentro del **Machine Learning**. Este proceso desarrolla una acción fundamental que le permite a los algoritmos de **aprendizaje automatizado**entrenar y conocer de forma adecuada los datos con los que desarrollan sus actividades.

**USO:**

El **Clustering**es una tarea que tiene como finalidad principal lograr el agrupamiento de conjuntos de objetos no etiquetados, para lograr construir subconjuntos de datos conocidos como Clusters. Cada clúster dentro de un grafo está formado por una colección de objetos o datos que a términos de análisis resultan similares entre sí, pero que poseen elementos diferenciales con respecto a otros objetos pertenecientes al conjunto de datos y que pueden conformar un clúster independiente.

Este tipo de proceso es aplicado en modelos de [Machine Learning](https://www.grapheverywhere.com/tipos-machine-learning/) de tipo no supervisado. Gracias a su implementación el sistema puede analizar los datos, realizar la tarea y encontrar los posibles errores dentro de su funcionamiento. El Clustering, en este caso sirve para segmentar datos en grupos de dimensiones similares en base a características para facilitar este proceso.

**METODOLOGÍA:**

Existen adicionalmente algunos otros métodos de **clusterización de datos** que debemos conocer, pues, pueden ser funcionales para nuestros proyectos. Descubramos cuáles son a continuación:

* **Algoritmo de k-medias:**

Es tal vez el método clásico para aplicar y entender el proceso de agrupamiento. Se establece un número de grupos previamente determinado. En este caso el algoritmo buscará los mejores centroides para realizar el agrupamiento, de manera que los miembros de cada grupo estén lo más cerca posible de sus centroides. El algoritmo funciona de forma iterativa, actualizando el centro de los clústeres de manera de ir reduciendo las distancias entre los miembros de cada clúster y su centro.

* **Clustering jerárquico:**

Uno de los métodos más utilizados, debido a la visualización práctica en forma de dendrograma que se obtiene. El clustering jerárquico puede realizarse tanto en forma divisiva o aglomerativa. Este método nos permite analizar alternativas para distintos números de grupos. Para entender un poco acerca de su funcionamiento si bien su procedimiento es bastante simple,  fijémonos por ejemplo en el caso aglomerativo:

1. Se parte de tantos grupos como individuos haya.
2. De acuerdo a la medida de similitud previamente seleccionada, unimos los dos grupos con mayor similitud para formar uno solo.
3. Continuamos de la misma forma hasta formar un solo grupo.

Dependiendo de los objetivos del proyecto o el problema a resolver podremos elegir quedarnos con algún agrupamiento especifico de los producidos durante el proceso, lo que nos permitirá ajustar el número de grupos finales.

**CARACTERÍSTICAS:**

* Es un tipo de aprendizaje no supervisado
* Agrupa los objetos de un dataset según su similaridad.
* La noción de clúster no puede ser definido con precisión

**APLICACIÓN:**

Las técnicas de agrupamiento encuentran aplicación en diversos ámbitos.

* En [biología](https://es.wikipedia.org/wiki/Biolog%C3%ADa) para clasificar animales y plantas.
* En [medicina](https://es.wikipedia.org/wiki/Medicina) para identificar enfermedades.
* En [marketing](https://es.wikipedia.org/wiki/Marketing) para identificar personas con hábitos de compras similares.
* En [teoría de la señal](https://es.wikipedia.org/w/index.php?title=Teor%C3%ADa_de_la_se%C3%B1al&action=edit&redlink=1) pueden servir para eliminar ruidos.
* En [biometría](https://es.wikipedia.org/wiki/Biometr%C3%ADa) para identificación del locutor o de caras.

**BIBLIOGRAFÍA:**

[**https://www.grapheverywhere.com/que-es-el-clustering/**](https://www.grapheverywhere.com/que-es-el-clustering/)

**https://es.wikipedia.org/wiki/Algoritmo\_de\_agrupamiento**