

PROPUESTA DE TRABAJO DE FIN DE MÁSTER  
MÁSTER EN MATEMÁTICAS AVANZADAS

Director(es): Álvaro Martínez Pérez

Tutor UCM: (sólo en caso de que no haya ningún director de la UCM)

Alumno(a): Damián Fromilhague García

Curso: Master Matemáticas Avanzadas

Título: Algoritmos de Clustering Jerárquico

Resumen: Un método de agrupamiento (o *clustering*) es un algoritmo que toma un espacio finito  $X$  dotado de una distancia (típicamente un espacio métrico finito) y da como salida una partición de  $X$ . Kleinberg propuso una aproximación axiomática al agrupamiento definiendo unas pocas propiedades básicas que una función de agrupamiento debía satisfacer. Entonces, probó que ninguna función de agrupamiento podía satisfacer esas condiciones simultáneamente. Esto no implica la imposibilidad de definir un algoritmo de agrupamiento consistente. El resultado de imposibilidad de Kleinberg solo se cumple cuando el único *input* en el algoritmo son el espacio y el conjunto de distancias. Se puede evitar incluyendo como parte del *input*, por ejemplo, el número de conjuntos (o *clusters*) que se desea obtener.

G. Carlsson y F. Mémoli, estudian el problema análogo para algoritmos de agrupamiento que producen como salida descomposiciones jerárquicas (*hierarchical clustering* o HC) del espacio en lugar de particiones del mismo. Los métodos de agrupamiento jerárquico también toman como *input* un espacio métrico finito pero la salida es una familia jerárquica de particiones de  $X$ . Dar una descomposición jerárquica es equivalente a dotar al espacio de una ultramétrica o, lo que es lo mismo, generar un árbol en el que los puntos del espacio se corresponden con los finales del árbol y el proceso de ramificación representa la familia jerárquica de particiones. Estos autores se aproximan al problema procurando una base teórica al estudio del *clustering* jerárquico (HC). En el espíritu del trabajo de Kleinberg, definen una serie de condiciones razonables que un HC debería satisfacer y prueban que el único algoritmo que satisface tres condiciones elementales es el agrupamiento jerárquico por encadenamiento singular o *single linkage hierarchical clustering*, SL HC.

En este trabajo vamos a repasar algunas de las estrategias para definir algoritmos de *clustering* jerárquico y analizar algunas de sus principales propiedades.

## Bibliografía:

- Carlsson, G. y Mémoli, F., Characterization, Stability and Convergence of Hierarchical Clustering Methods, *Journal of Machine Learning Research*, 11, 1425\_1470 (2010).
- Carlsson, G. y Mémoli, F., Multiparameter Hierarchical Clustering Methods, in *Classification as a Tool for Research*, ed. by C.W.H. Locarek- Junge. Proc. 11th Conference of the International Federation of Classification Societies (IFCS-09) (Springer, Heidelberg), 63\_70 (2010).
- Kleinberg, J. M. An Impossibility Theorem for Clustering, in *Advances in Neural Information Processing Systems*, 15 (NIPS 2002), eds. S. Becker, S. Thrun, and K. Obermayer, Cambridge MA: MIT Press, 446\_453 (2002).
- Martínez-Pérez, A density sensitive hierarchical clustering method, *Journal of Classification*. 35, 481\_510 (2018)
- Martínez-Pérez, Gromov-Hausdor\_ Stability of Linkage-based hierarchical clustering methods, *Advances in Mathematics*. 279, pp. 234 - 262. Elsevier, (2015).