

DBDM – Clustering

TD

1 Clustering

Let us consider the following points $\mathcal{P} = \{A : (1, 1), B : (8, 3), C : (3, 3), D : (4, 9), E : (2, 7), F : (4, 6), G : (6, 5), H : (5, 1), I : (9, 1), J : (5, 8), K : (5, 5)\}$.

1. k -means.
 - (a) Compute a bi-partition ($k = 2$) with the k -means algorithm. Use G and K as the first centroids. To ease the distance computations, we consider the Manhattan distance.
 - (b) Compute a 3-partition ($k = 3$) with the k -means algorithm. Use A , B and I as the first centroids.
2. Hierarchical clustering
 - (a) Build the distance (Manhattan) matrix
 - (b) Perform a hierarchical clustering until obtaining 3 classes. The dissimilarity between clusters is the maximal distance among the pairs composed of objects from the two clusters (Complete Link Method).
 - (c) Perform a hierarchical clustering until obtaining 3 classes. The dissimilarity between clusters is the minimal distance among the pairs composed of objects from the two clusters (Single Link Method).





