

## Exercice 6 Algorithmes de Clustering

### Distance

<b>Distance euclidienne:</b>	$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$
<b>Distance de Manhattan :</b>	$d(A, B) = \sum_{i=1}^n  A_i - B_i $
<b>Distance de Hamming :</b>	$d(A, B) = \# \{ i : A_i \neq B_i \}$

### Algorithme

<b>Algorithme : k-means</b>
<b>Entrée :</b> D : Dataset ; k : le nombre de cluster à former ; <b>Sortie :</b> D: Dataset étiqueté ; <b>Début</b> Choisir aléatoirement k instances comme centroïdes. <b>Répéter</b> Calculer la distance entre chaque instances D[i] et les k centroïdes ; Affecter chaque instances D[i] au groupe dont il est le plus proche de son centre ; Calculer le nouveau centre de chaque cluster et modifier le centroïde ; <b>Jusqu'à</b> D[i] identique à D[i-1] ; <b>Retourner</b> D ; <b>Fin.</b>

### Questions :

- 1- Écrire une fonction java permettant de calculer la distance de Manhattan entre deux instances du dataset.
- 2- Écrire une fonction java permettant de calculer le centroïde d'un ensemble d'instances.
- 3- Écrire une fonction java permettant de trouver le cluster dont une instance donnée est la plus proche.
- 4- Implémenter en java l'algorithme k-means.
- 5- Déduire les clusters formés (sans considérer la classe comme un attribut).