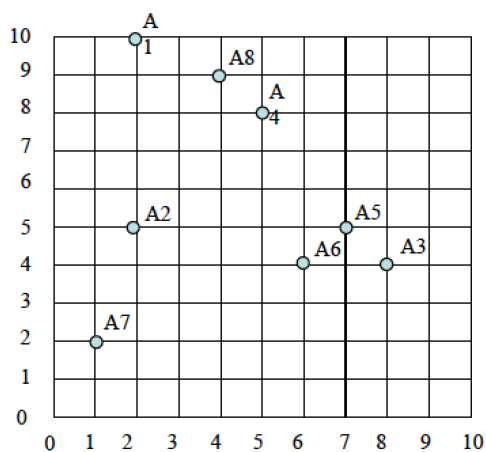


## TD 2 – Apprentissage non supervisé

### Exercice 1

On considère un jeu de points dans  $\mathbb{R}^2$  suivant que l'on veut partitionner avec une méthode de classification non supervisée:

$$A_1 = (2, 10), A_2 = (2, 5), A_3 = (8, 4), A_4 = (5, 8) \\ A_5 = (7, 5), A_6 = (6, 4), A_7 = (1, 2), A_8 = (4, 9)$$



### Kppv

1. Calculer la matrice des distances euclidiennes au carré entre les points  $(A_i)_{i=1..8}$ .
2. Réaliser la méthode des K-ppv avec  $k = 1$  et un seuil de 16.

### kmeans

1. Réaliser une itération de la classification par K-means en prenant des centres initiaux les points  $A_1$ ,  $A_4$  et  $A_7$ .
2. Obtient-on la même partition que pour les Kppv ?

### Classification hiérarchique

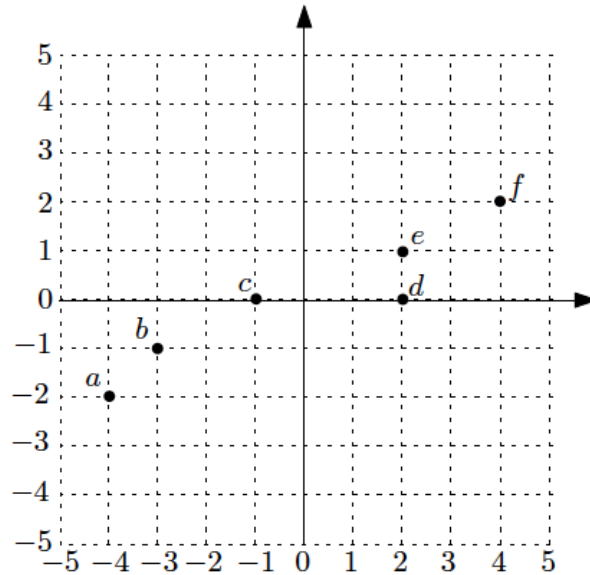
1. Réaliser une classification hiérarchique par lien simple et lien complet.
2. Représenter les dendrogrammes respectifs.
3. Pour quelle(s) mesure(s) de dissimilarité obtient-on la même partition que pour les Kppv et Kmeans?

*Rappel: Soient  $G$  et  $H$  deux groupes de données et  $d$  la distance euclidienne classique,*

$$\text{Distance du lien simple/single linkage} \quad \left| \quad \text{Distance du lien complet/complete linkage} \right. \\ d_{SL}(G, H) = \min_{i \in G, i' \in H} d_{ii'} \quad \left| \quad d_{CL}(G, H) = \max_{i \in G, i' \in H} d_{ii'}$$

## Exercice 1 : ACP et Classification

On considère le jeu de données suivant :



### Questions

1. Calculer la matrice de variance-covariance  $\Sigma$ .
2. Calculer le premier axe principal et les composantes principales correspondantes.
3. Appliquer, sur les composantes principales 1D, l'algorithme des  $k$ -plus proches voisins pour  $k = 1$  en supposant que le seuil est égal à  $\frac{20}{7}$ .