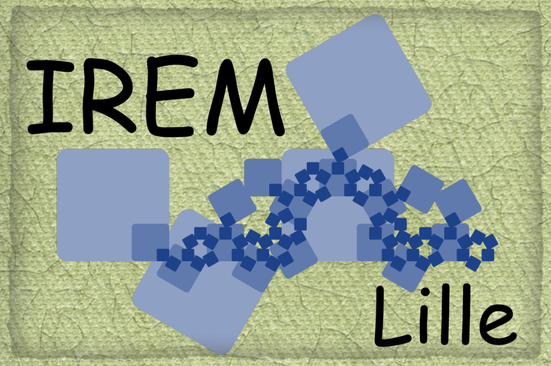


Groupe Info

IREM de Lille

Asli Grimaud

IA et Apprentissage Automatique

Mars - Avril 2025

asli.grimaud@ac-lille.fr

- Apprentissage supervisé
  - Algorithme des k-plus proches voisins
  - Algorithme ID3
- Apprentissage non supervisé
  - Classification hiérarchique ascendante
  - Algorithme des k-moyennes

Apprentissage non supervisé

Technique d'apprentissage automatique où l'objectif principal est de découvrir des structures ou des motifs cachés dans les données.

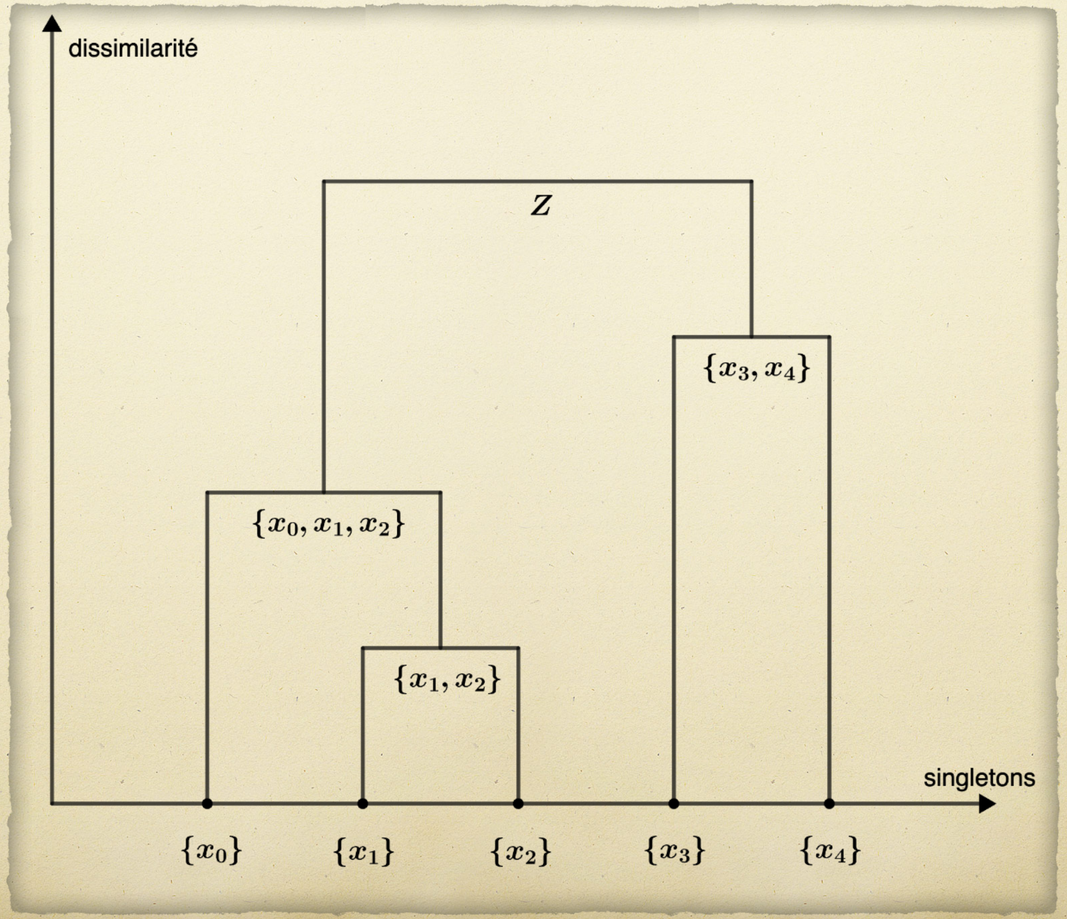
Ex : Regrouper des étoiles en clusters basés sur leurs caractéristiques spectrales et lumineuses sans connaissance préalable des catégories existantes.

Segmentation d'images en génétiques, identification des section d'ADN.

$$Z = \{x_i \mid i \in [\![0, n-1]\!], x_i \in \mathbb{R}^d\}$$

Classification hiérarchique ascendante

Idée : regrouper progressivement les données en attribuant initialement une classe distincte à chaque donnée de Z, puis fusionner ces classe pour en former des de plus en plus grande.



Classification hiérarchique ascendante

Calcul de distances entre deux individus x_i et x_j

choix : distance euclidienne

$$\delta(x_i, x_j) = \sqrt{\sum_{k=0}^{d-1} \left(x_{ik} - x_{jk}\right)^2}$$

Calcul de distances entre deux classes c_i et c_j

- lien simple : $\delta_{min}(c_i, c_j) = \min_{x_i \in c_i, x_j \in c_j} \delta(x_i, x_j)$
- lien complet : $\delta_{max}(c_i, c_j) = \max_{x_i \in c_i, x_j \in c_j} \delta(x_i, x_j)$
- lien moyen : $\delta_{av}(c_i, c_j) = \frac{\sum_{x_i \in c_i} \sum_{x_j \in c_j} \delta(x_i, x_j)}{|c_i||c_j|}$
- lien de Ward : $\delta_{Ward}(c_i, c_j) = \frac{w_i w_j}{w_i + w_j} \delta^2(g_i, g_j)$ où

  w_i et w_j sont les poids des ensembles, g_i et g_j sont leurs centre de gravités
- ...

choix à faire

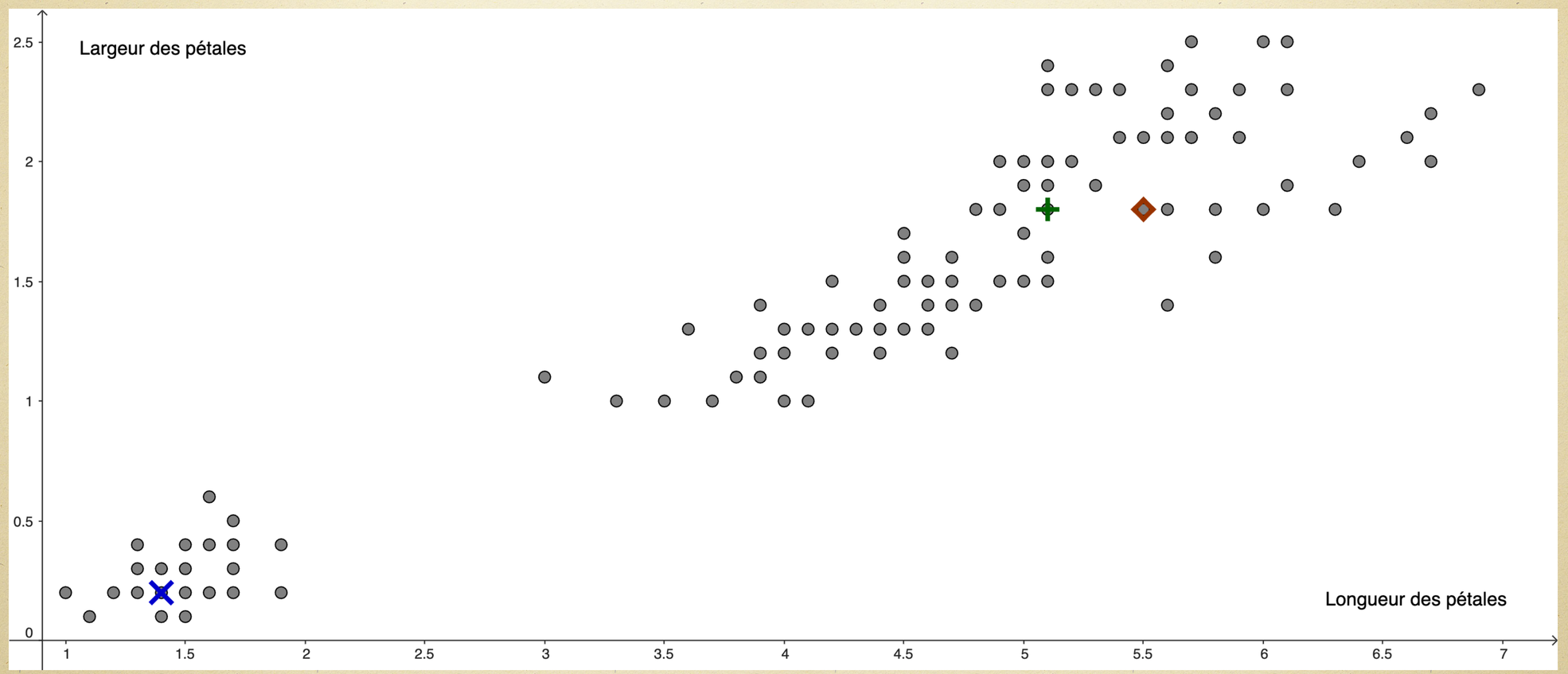
Algorithme des k-moyennes

Idée : partitionner un ensemble de données en k clusters (groupes) distincts, de manière à ce que les éléments au sein d'un même cluster soient plus similaires entre eux qu'avec ceux des autres clusters.

Chaque cluster est représenté par le centroïde (barycentre des points qui lui sont associés).

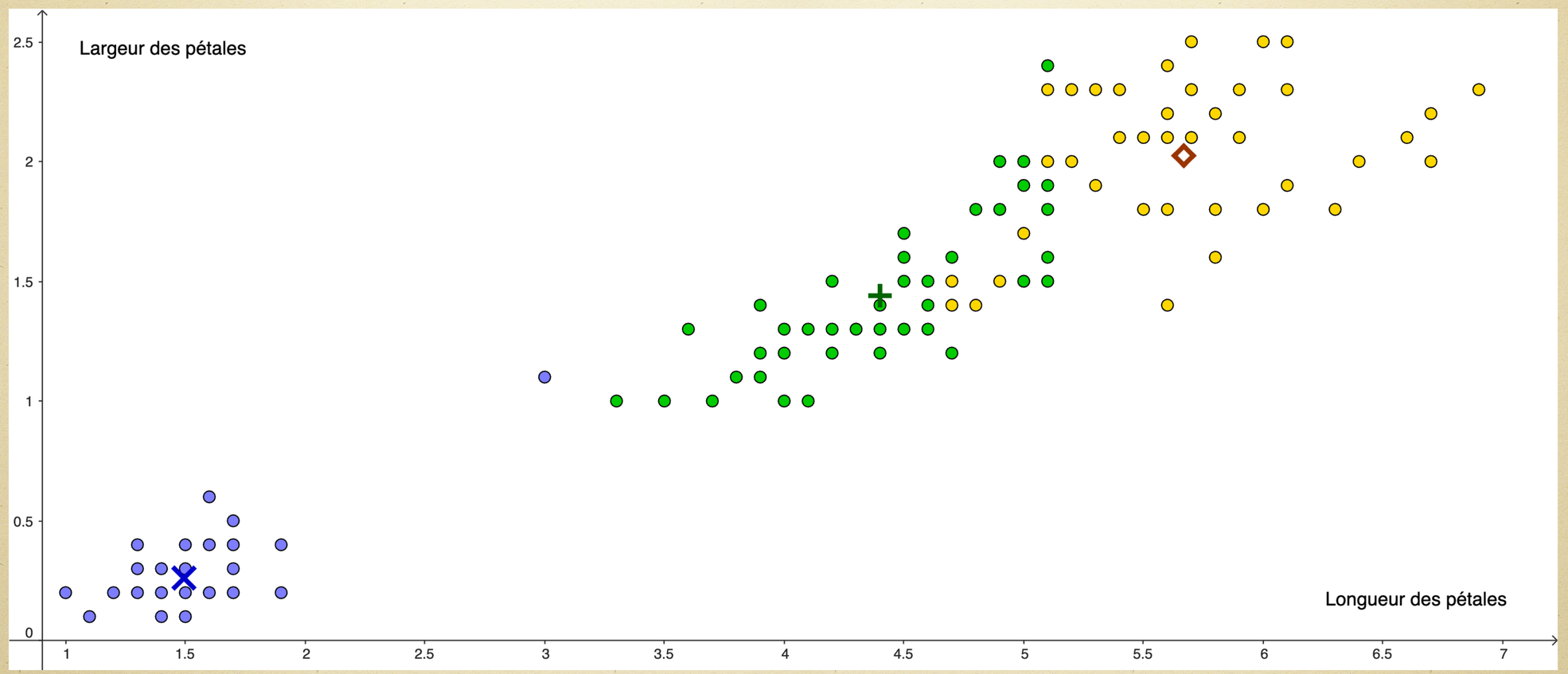
choix : distance euclidienne

Algorithme des k-moyennes



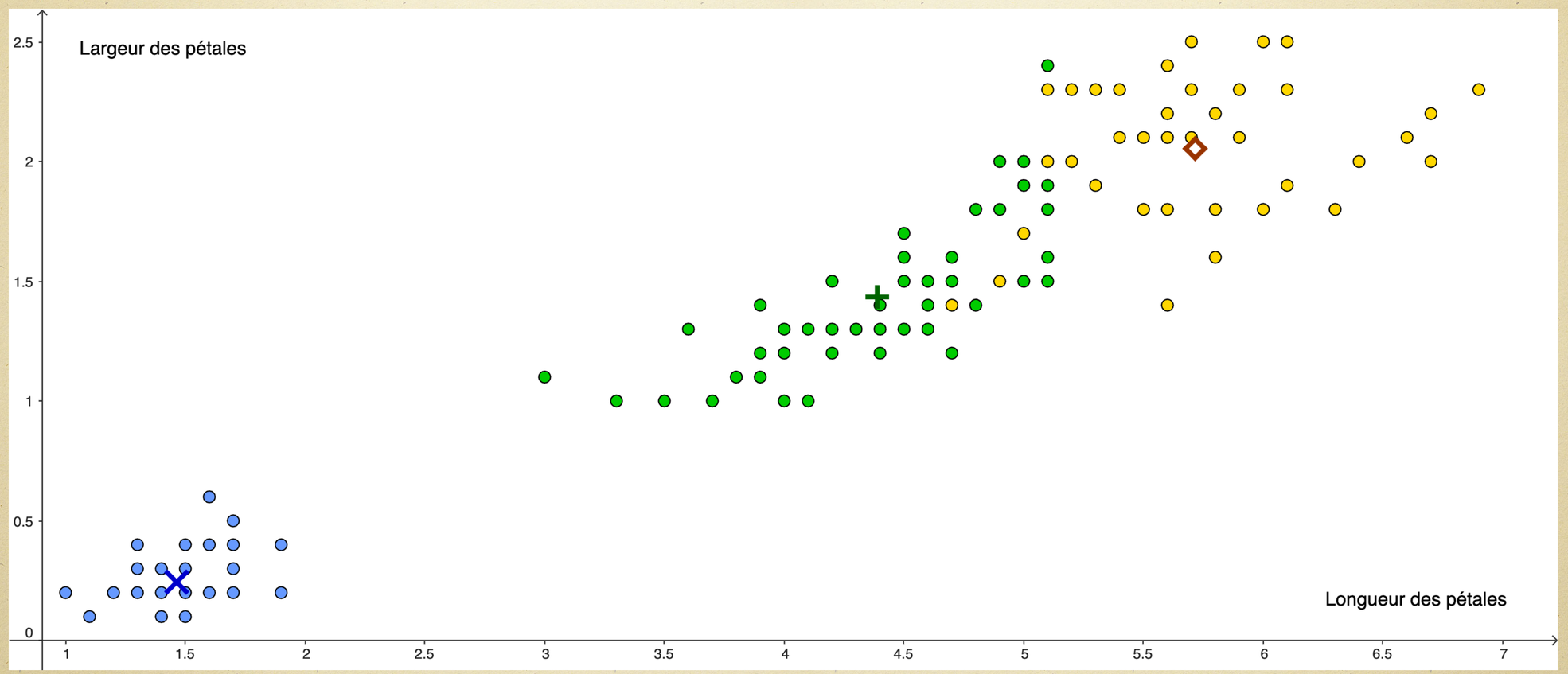
initialisation avec les trois centroïdes choisis au hasard

Algorithme des k-moyennes



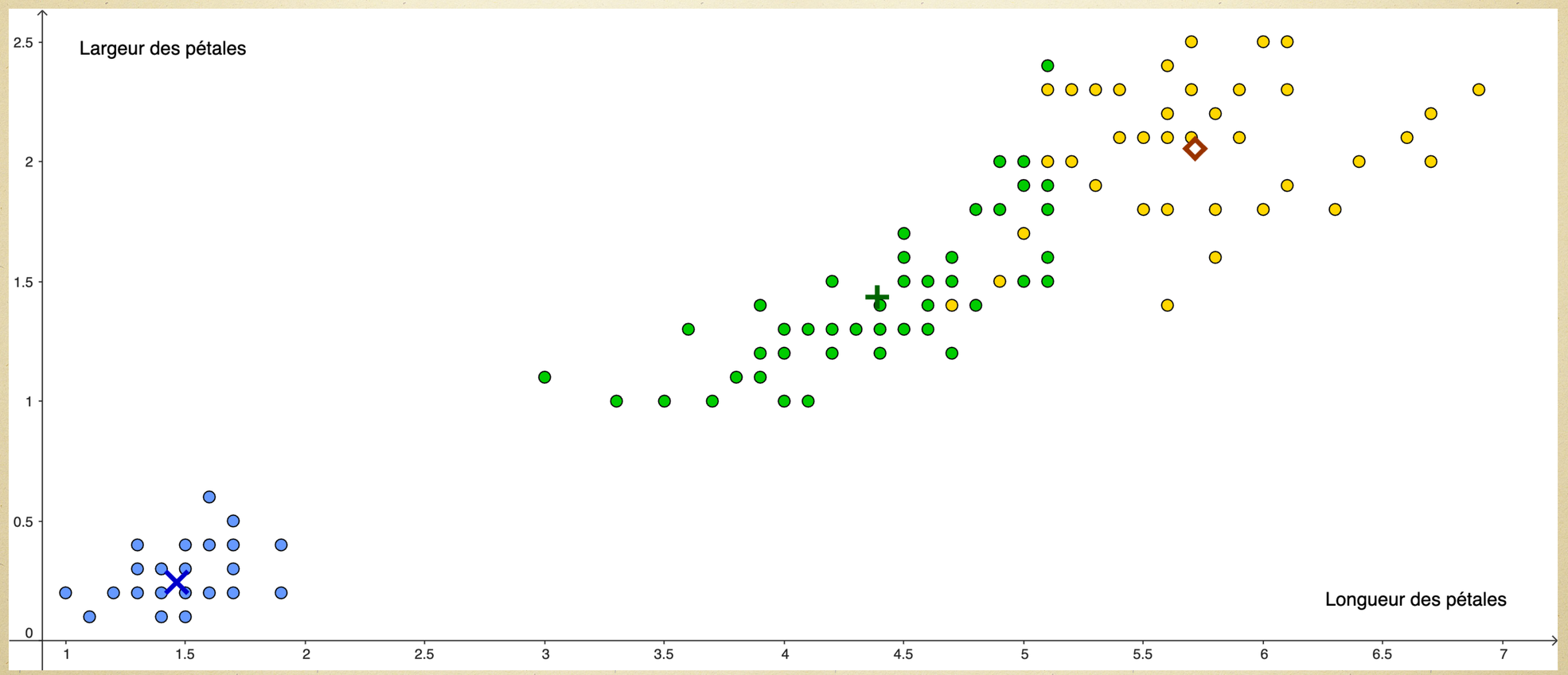
répartition des individus après l'itération 0 et nouveaux centroïdes

Algorithme des k-moyennes



répartition des individus après l'itération 1 et nouveaux centroïdes

Algorithme des k-moyennes



répartition des individus après l'itération 2 et nouveaux centroïdes

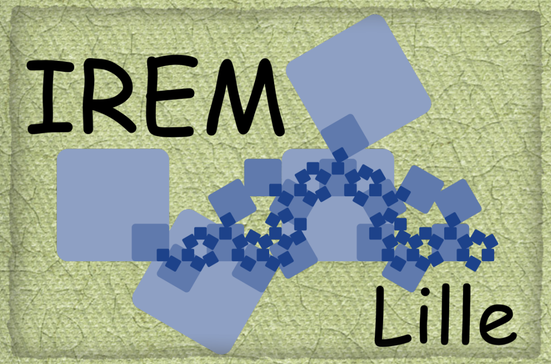
Algorithme des k-moyennes

Arrêt :

- le nombre maximum d'itération est atteint
- les centroïdes n'évoluent plus que d'une quantité inférieure à un ε fixé.

- Apprentissage supervisé
  - Algorithme des k-plus proches voisins
  - Algorithme ID3
- Apprentissage non supervisé
  - Classification hiérarchique ascendante
  - Algorithme des k-moyennes

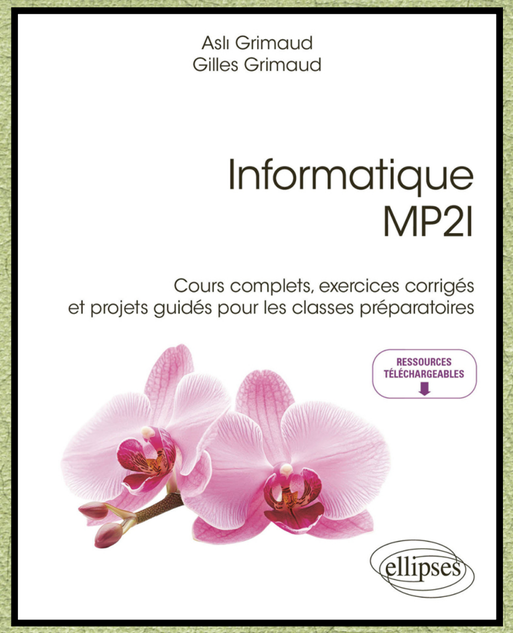
- programmes IA
- application de l'algorithme
- mémoire / capacité de calcul
- nombreux choix
- calculs, heuristiques
- solutions pas trop mauvaise
- temps raisonnable
- sentiment d'intelligence



Groupe Info

IREM de Lille

Asli Grimaud



ISBN : 9782340103849

IA et Apprentissage Automatique

MERCI

Mars - Avril 2025

asli.grimaud@ac-lille.fr