

# TS326 : Reconnaissance des Formes

*Filière Electronique - option TSI, 3<sup>E</sup> année*

Rémi Giraud

*remi.giraud@enseirb-matmeca.fr*

2021-2022

## TP Classification non supervisée

### 1 Introduction

---

L'objectif de ce TP est d'implémenter plusieurs techniques populaires de classification non supervisée. Plus particulièrement, il s'agira d'implémenter les méthodes des centres mobiles ou  $K$ -moyennes, ainsi que la méthode de classification ascendante hiérarchique. Ces algorithmes seront appliqués dans un premier temps au clustering spatial de nuages de points 2D, puis à des descripteurs extraits d'images d'épi de blé.

#### 1.1 Travail demandé

Le but de ce TP est de compléter les codes Matlab fournis pour réaliser l'implémentation des algorithmes  $K$ -moyennes et de classification ascendante hiérarchique. Dans un premier temps, nous chercherons à grouper les données de nuage de points 2D en  $K = 4$  classes par ces deux méthodes. Les mesures d'inerties intra classes, inter classes et totale seront données à chaque étape. Ensuite, nous pourrons tenter d'appliquer ces algorithmes aux données issues d'images d'épi de blé, en considérant une distance multi-critères et un nombre de classes  $K = 2$ .

#### 1.2 Données

- Trois ensembles de données spatiales 2D sont fournis dans les fichiers "cloud\_data\_#1-3.mat". Les données sont rangées sous la forme  $X = (y, x, classe)$  et sont représentées comme nuages de points en Figure 1, .

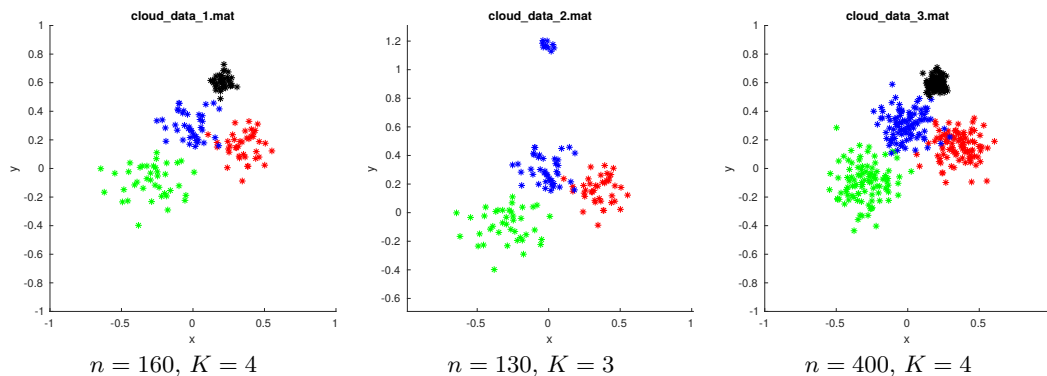


Figure 1 : Nuages de points 2D.

- Dans la partie 3, nous nous intéresserons aux données d'épi de blé vue en cours. Des images sont extraites d'une image d'épi et classifiées manuellement en deux classes : épis de blés (classe = 1) et feuilles / fond (classe = 2). Quatre descripteurs statistiques sont issus des matrices de cooccurrences et sont fournis dans le dossier "ble\_data.mat".

## 2 Algorithmes de classification non supervisée

---

### 2.1 Algorithme $K$ -moyennes

1. Implémenter l'algorithme  $K$ -moyennes pour un nombre de classes  $K$  quelconque.  
Les centres seront d'abord choisis aléatoirement dans l'espace des données.  
On utilisera comme distance la norme euclidienne  $L_2$ .
2. Quelles solutions envisager pour arrêter l'algorithme automatiquement ?
3. Calculer les inerties intra classes, inter classes et totale à chaque itération.
4. Visualiser l'évolution de l'algorithme aux données du nuage de points "cloud\_data\_1.mat" pour  $K = 4$ .  
Arrive-t-il d'avoir un nombre de classe variable en sortie de l'algorithme ?
5. Implémenter la méthode d'optimisation de l'initialisation  $K$ -moyennes++.  
Vérifier que l'algorithme converge plus rapidement.
6. Observer la difficulté d'obtenir un bon clustering sur le nuage "cloud\_data\_2.mat".  
Implémenter la méthode  $K$ -médoides (les centres sont les points avec la faible distance moyenne aux autres points de la même classe).
7. Comparer les coûts temporels des méthodes sur le nuage "cloud\_data\_3.mat"

### 2.2 Algorithme de classification ascendante hiérarchique

1. Implémenter l'algorithme de classification ascendante hiérarchique.  
On propose d'utiliser les structures de données suivantes :
  - Données initiales  $data$  de taille  $n \times p$  (ici  $n = 160$  et  $p = 2$  pour "cloud\_data\_1.mat")
  - Tableau de classification  $classif\_exp$  de taille  $n \times n$  à modifier à chaque itération :  
(ici à l'itération 1, les clusters 1 et 3 ont fusionné)

Itération	1	2	3	...	i	...	n
0	1	2	3	...	i	...	n
1	1	2	1	...	i	...	n
...							
n	i	i	i	...	i	...	i
  - Barycentres  $G_i$  updatés à chaque nouvelle association
2. Implémenter et comparer les quatre stratégies d'agrégation (simple, complet, moyen, méthode de Ward).

## 3 Multi-critères

---

On peut dans un dernier temps appliquer les algorithmes implémentés à des données réelles provenant d'images d'épis de blé.

1. Tenter d'obtenir un clustering pertinent en utilisant un des quatre descripteurs.
2. Proposer une stratégie pour utiliser tous les descripteurs à la fois dans le calcul de distance, sachant qu'ils ne représentent pas la même information.