

---

---

# LFA

# Fuzzy Clustering Library

— Eléonore d'Agostino —  
Benoît Zuckschwerdt

---

---

# Introduction

- Fuzzy clustering
- Méthodes développées
- Résultats
- Difficultés
- Conclusion

# Fuzzy clustering

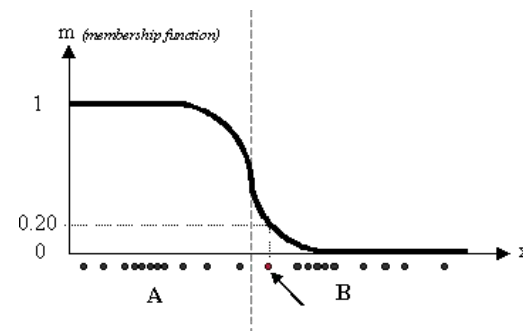
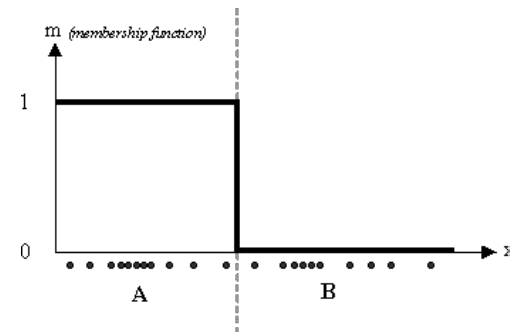
- Le contraire du “Hard Clustering”, où les points appartiennent à un seul cluster
- Alias “Soft clustering”, un point à un degré d'appartenance à un ou plusieurs cluster
- On cherche à créer des clusters les plus distincts les uns des autres, mais permettant quand même des appartenances multiples

# Méthodes développées - outils

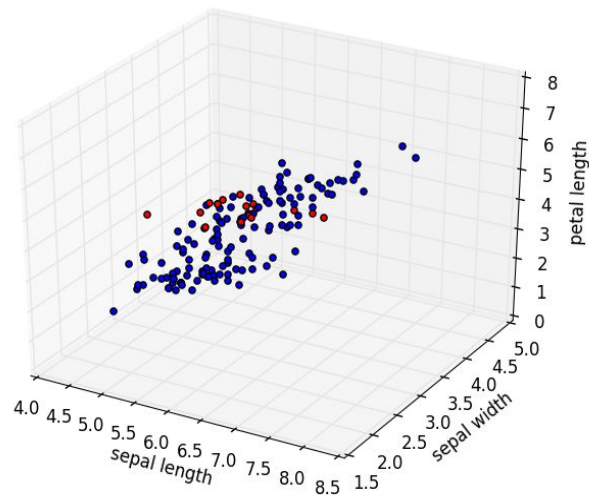
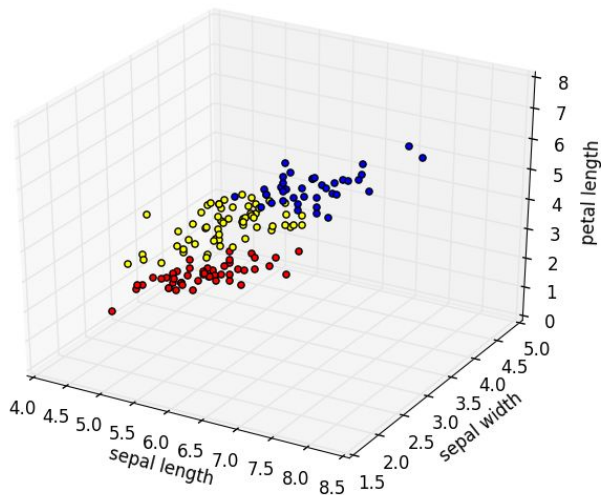
- **Sub-clustering** permet de décider d'un nombre de clusters pour un set de données
- **K-Means ++** permet de trouver de meilleurs seeds pour K-means, permettant à l'algorithme de finir plus vite (et demandant moins d'entrées utilisateur)

# Méthodes développées - clustering

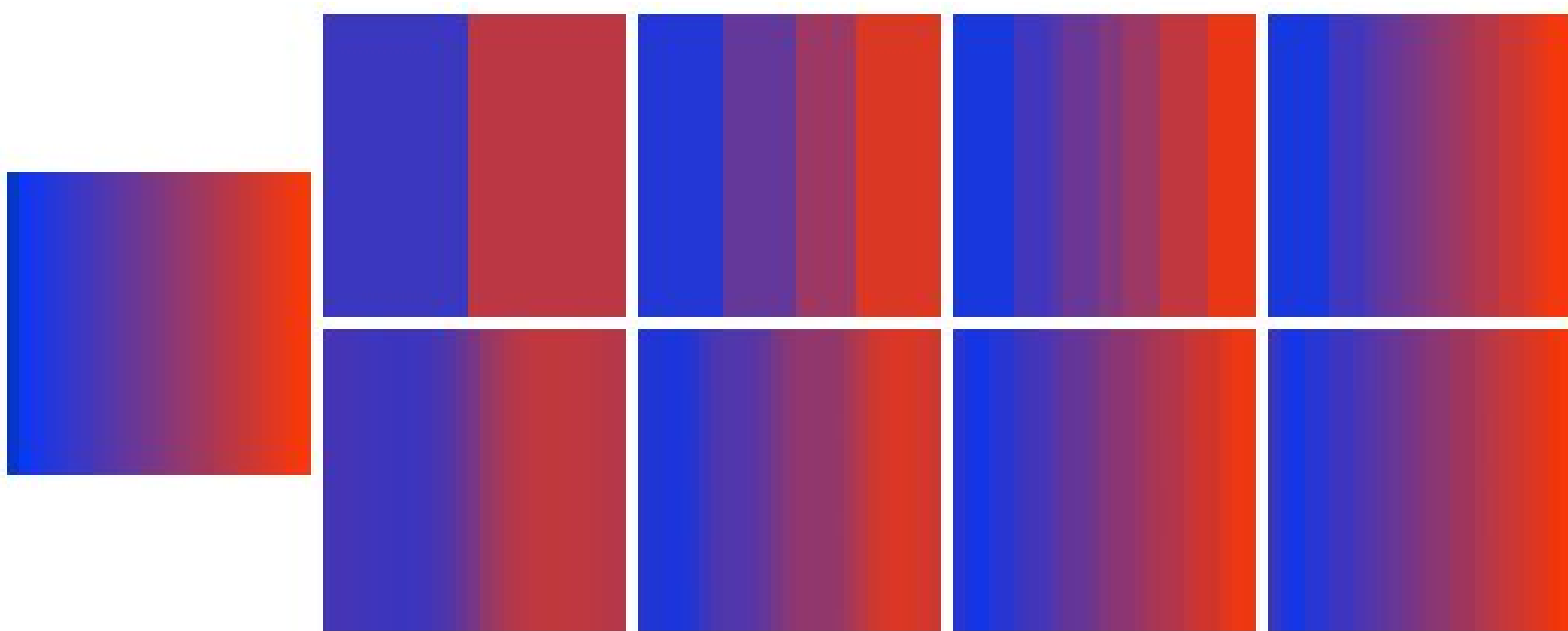
- **K-Means** permet de séparer un set de données en  $k$  groupes distincts
- **C-Means** (ou Fuzzy C-Means) est similaire à K-Means, mais produits des groupes pouvant se chevaucher: chaque point a un certain pourcentage d'appartenance à chaque groupe



# Résultats - Iris Dataset



# Résultats - Images (dégradé)



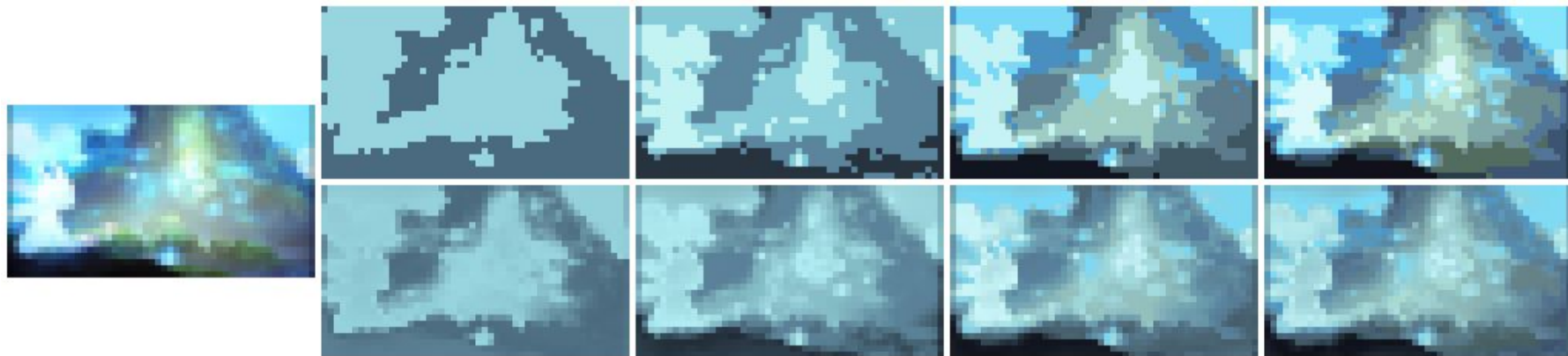
## Résultats - Images (peu de couleurs)





# Résultats - Images (beaucoup de couleurs)

- 



# Difficultés

- Implémentation de C-means
  - Difficile à débbug comparé à K-means
- Exemples d'implémentations faciles à trouver, mais difficiles de trouver des formules
  - Particulièrement pour C-means

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} / d_{ik})^{(2/m-1)}$$

C'était  $2/(m-1)$  et pas  $(2/m)-1$

- On a perdu 2 jours là-dessus

# Conclusion

- Bonne compréhension du fonctionnement interne de K-means et C-means
- Implémentation pauvre en performances surtout sur des grands sets de données
  - L'optimisation serait un plus
  - ~2min de calculs pour du C-means à 16 clusters sur 2500 points (image 50x50 RGBA)