UNIVERSITÉ LUMIÈRE LYON 2
M1 MIASHS

Apprentissage non supervisé

Projet sous le logiciel R - 2023-2024

L'objectif du projet est d'implémenter un algorithme EM pour l'estimation d'un modèle de mélange gaussien sous R.

L'algorithme Les données sont supposées vivre dans $\mathbb{R}^p$, et le modèle considéré est un modèle de mélange à K composantes gaussiennes. L'inférence sera réalisée par un algorithme EM. Votre fonction de clustering devra prendre en entrée :

- les données,
- le nombre K de composantes,
- le nombre d'initialisation aléatoires (20 par défaut).

Pour chaque estimation de modèle, des initialisations aléatoires multiples seront implémentées en interne de la fonction, et la solution de meilleure vraisemblance sera conservée.

En sortie, l'algorithme devra retourner :

- les probabilités d'appartenances de chaque donnée à chaque composante,
- l'affectation des données aux composantes du mélange,
- la valeur de la vraisemblance tout au long des itérations de l'algorithme (pour la meilleure initialisation conservée),
- les paramètres du modèle estimé.

Expérimentation Vous devrez expérimenter votre algorithme sur deux jeux de données :

- un jeu de données simulées suivant un modèle de mélange gaussien (dont vous choisirez les paramètres), en vérifiant que votre algorithme permet de bien retrouver les paramètres utilisés pour la simulation,
- les données iris.

Livrable (par binôme pour le 03/03/2024)

- un script Rmarkdown (ainsi que sa compilation pdf) qui contiendra l'ensemble des éléments demandés : le code de votre algorithme EM, les expérimentations, etc...