
Etude de la robustesse de RMixmod (package de classification par modèles de mélanges) en cas de chevauchement de classes

Florent Langrognet^{*1}

¹Laboratoire de Mathématiques de Besançon (LM-Besançon) – CNRS : UMR6623, Université de Franche-Comté – UFR Sciences et techniques 16 route de Gray 25 030 Besançon cedex, France

Résumé

Les modèles de mélanges offrent un cadre probabiliste flexible et efficace pour traiter des problématiques de classification supervisée ou non supervisée. L'objectif du projet MIXMOD est de diffuser un ensemble logiciel de classification des données par modèles de mélanges à un large spectre d'utilisateurs via plusieurs composants logiciels. La bibliothèque de calcul mixmodLib (C++) en est la pierre angulaire, résultat d'un travail de près de 15 ans sur la robustesse et la rapidité de calcul. Le package RMixmod, ensemble de fonctions pour R, interfacé avec mixmodLib (grâce à RCPP) est devenu un outil de référence pour la classification des données.

Intégrant de nombreuses fonctionnalités (algorithmes de type EM, critères de sélection, modèles parcimonieux, stratégies d'initialisation, ...), cet ensemble logiciel permet de traiter des données quantitatives, qualitatives et mixtes, y compris dans des situations complexes. L'une des difficultés en classification des données réside dans la capacité à donner des bons résultats en cas de chevauchement entre plusieurs classes : trouver le bon nombre de classes, les bons paramètres et les bonnes affectations des individus à ces classes (labels). L'étude consiste à tester RMixmod sur des jeux de données simulées en contrôlant le degré de chevauchement entre les classes (grâce au package MixSim) sur des données quantitatives.

*Intervenant