

Identification de classes non-disjointes ayants des densités différentes

Hela Masmoudi*, Chiheb-Eddine Ben N’Cir**, Nadia Essoussi***

*LARODEC, ISG Tunis, Université de Tunis
hela.masmoudi@gmail.com

**LARODEC, ISG Tunis, Université de Tunis
chiheb.benncir@isg.rnu.tn

***LARODEC, ISG Tunis, Université de Tunis
nadia.essoussi@isg.rnu.tn

Résumé. La classification recouvrante correspond à un enjeu important en classification non-supervisée en permettant à une observation d’appartenir à plusieurs clusters. Plusieurs méthodes ont été proposées pour faire face à cette problématique en utilisant plusieurs approches usuelles de classification. Cependant, malgré l’efficacité de ces méthodes à déterminer des groupes non-disjoints, elles échouent lorsque les données comportent des groupes de densités différentes car elles ignorent la densité locale de chaque groupe et ne considèrent que la distance Euclidienne entre les observations. Afin de détecter des groupes non-disjoints de densités différentes, nous proposons deux méthodes de classification intégrant la variation de densité des différentes classes dans le processus de classification. Des expériences réalisées sur des ensembles de données artificielles montrent que les méthodes proposées permettent d’obtenir de meilleures performances lorsque les données contiennent des groupes de densités différentes.

1 Introduction

La classification non-supervisée a pour but de regrouper les observations proches dans un même groupe, tandis que les observations éloignées doivent être affectées à des groupes différents. Cette définition pourrait être insuffisante dans de nombreuses applications de regroupement dans lesquelles un objet peut appartenir à la fois à plusieurs classes. Ce type de problématique est appelée classification recouvrante ou encore classification non-exclusive. Plusieurs applications réelles nécessitent d’utiliser ce type de schéma de classification tels que le regroupement de documents où chaque document peut aborder plusieurs thèmes, la classification de vidéos où un film peut avoir différents genres (Snoek et al., 2006), la détection d’émotions où un morceau de musique peut évoquer plusieurs émotions distinctes (Wieczorkowska et al., 2006).

Afin de produire des classes non-exclusives, divers méthodes ont été proposées utilisant des approches hiérarchique (Diday, 1984), de partitionnement (Fu et Banerjee, 2008; Banerjee