# Passage aux noyaux en classification recouvrante

Guillaume Cleuziou*,**

*LIFO, Université d'Orléans, Rue Léonard de Vinci, 45067 Orléans Cedex 2
guillaume.cleuziou@univ-orleans.fr
**GREYC, UCBN, Bd. Maréchal Juin, 14032 Caen Cedex 5

**Résumé.** La classification recouvrante correspond à un domaine d'étude très actif ces dernières années et dont l'objectif est d'organiser un ensemble de données en groupes d'individus similaires avec la particularité d'autoriser des chevauchements entre les groupes. Parmi les approches étudiées nous nous intéressons aux extensions recouvrantes des modèles de type moindres carrés et constatons les difficultés théoriques et pratiques liées à leur adaptation aux noyaux. Nous formulons alors une nouvelle définition ensembliste pour caractériser un recouvrement de plusieurs classes, nous montrons que cette modélisation permet le recours aux noyaux et nous proposons une solution algorithmique efficace pour répondre au problème de la classification recouvrante à noyaux.

## 1 Introduction

La classification recouvrante consiste à organiser de manière non-supervisée un ensemble d'individus en classes chevauchantes ou recouvrantes composées d'individus similaires. L' étude des méthodologies associées vise avant tout à répondre aux besoins réels et pratiques communs à de nombreux domaines d'application : qu'il s'agisse de classer des documents (textes, images, vidéos), de constituer des communautés de personnes sur des critères sociaux ou marketing ou encore d'exhiber des groupes de gènes ou de molécules présentant des caractéristiques structurelles ou fonctionnelles communes, il est très fréquents d'être confronté à des données qui s'organisent naturellement en classes recouvrantes. Dans ces applications, le recours à des techniques usuelles de classification stricte ou disjointe apporterait un biais préjudiciable à l'usage final de la classification.

Depuis plus d'une trentaine d'années, la classification recouvrante est identifiée comme une problématique à part entière et donne lieu à des avancées constantes au fil de l'évolution des techniques de classification traditionnelles. Partant des premiers travaux de Shepard et Arabie (1979) portant sur le clustering (recouvrant) additif, la problématique s'est ensuite orientée vers l'acquisition et la théorisation des classifications hiérarchiques recouvrantes ou empiétantes (Diday, 1987; Diatta et Fichet, 1994; Bertrand et Janowitz, 2003) avant d'être reconsidérée plus récemment et de façon plus formelle du point de vue "partitionnement" (Banerjee et al., 2005; Cleuziou, 2008; Depril et al., 2012). Parmi ces dernières avancées on notera d'une part les modèles additifs ALS (Depril et al., 2012) et son équivalent en terme de modèle de mélanges recouvrants MOC (Banerjee et al., 2005) et d'autre part le modèle OKM (Cleuziou, 2008) que l'on pourrait qualifier de modèle géométrique tant il modélise les intersections de clusters par