

Optimisation exacte pour la classification non supervisée

Arnaud Knippel (LMI)

7e journée de la Fédération Normandie-Mathématiques,
Le Havre, 12 juin 2015

De nombreuses méthodes existent pour la résolution approchée des problèmes de classification non supervisée, ou partitionnement de données. Ce type de problème se modélise typiquement comme un problème d'optimisation combinatoire NP-difficile, pour lequel des formulations mathématiques ont été proposées sous la forme de programmes linéaires ou quadratiques en variables binaires.

Ce type de méthode est bien adapté à la classification en grande dimension, mais est encore limité par le nombre d'éléments à classer : des résultats exacts ont déjà été produits dès les années 1990 dans de nombreux domaines pour des données constituées de quelques dizaines ou centaines d'éléments ayant un nombre quelconque de caractéristiques.

Une piste sérieuse d'amélioration consiste à réduire la symétrie de ces formulations. Nous proposons une nouvelle formulation sans aucune symétrie et dont la relaxation continue se compare bien à celles des formulations existantes - c'est un point important pour l'efficacité des méthodes de type *Branch and Cut*. Cette formulation peut être étendue pour améliorer encore cette relaxation.

Cet exposé s'appuie sur les travaux de thèse de Zacharie Alès et des résultats récents.