

Recherche de groupes parallèles en classification non-supervisée

Lionel Martin*, Matthieu Exbrayat*, Teddy Debroutelle**,
Aladine Chetouani**, Sylvie Treuillet**, Sébastien Jesset***

*LIFO, Batiment IIIA, rue Léonard de Vinci, B.P. 6759,
F-45067 Orléans cedex 2, prenom.nom@univ-orleans.fr,
<http://www.univ-orleans.fr/lifo/>

**Laboratoire PRISME, Université d'Orléans, 12 rue de Blois,
F-45067 Orléans cedex 2, prenom.nom@univ-orleans.fr,
<http://www.univ-orleans.fr/prisme/>

***Service Archéologique Municipal d'Orléans, 13 bis rue de la Tour Neuve,
45000 Orléans, sjesset@ville-orleans.fr

Résumé. Dans cet article, nous nous intéressons à une situation de classification non supervisée dans laquelle nous souhaitons imposer une "forme" commune à tous les clusters. Dans cette approche, la "forme" commune sera caractérisée par un hyperplan qui sera le même pour tous les groupes, à une translation près. Les points sont donc supposés être distribués autour d'hyperplans parallèles. La fonction objectif utilisée peut naturellement s'exprimer comme la minimisation de la somme des distances de chaque point à son hyperplan. Comme pour le cas de k-means, la résolution est effectuée par l'alternance de phases d'affectation de chaque point à l'hyperplan le plus proche et de phases de calcul de l'hyperplan qui ajuste au mieux l'ensemble des points qui lui sont affectés. L'objectif étant d'obtenir des hyperplans parallèles, cette phase de calcul est menée simultanément pour tous les hyperplans, par une méthode de régression.

1 Introduction

Le clustering, tâche classique d'apprentissage automatique, consiste à scinder un ensemble d'objets en quelques sous-groupes cohérents. Les critères de cohérence peuvent varier, mais reposent en général sur les distances ou (dis)similarités entre les objets observés. La littérature propose plusieurs classifications des méthodes de clustering ; nous suggérons au lecteur de consulter par exemple (Aggarwal et Reddy (2013)) pour une vue d'ensemble.

Dans cet article, nous nous intéressons au cas où nous possédons une connaissance a priori sur la forme des groupes ; en particulier, les groupes sont supposés être distribués autour d'hyperplans parallèles. Divers jeux de données, notamment en reconnaissance de formes, présentent une structuration spatiale des objets en couches parallèles. La figure 1 donne un exemple de distribution spatiale en frises parallèles. Sur cette figure on observe à gauche le relevé de motifs sur des tessons de poteries carolingiennes. Ces motifs ont été imprimés lors de