

Proposition d'un sujet de stage Master II en statistique appliquée

Intitulé du Sujet : Evaluation de six méthodes pour l'analyse exploratoires des données multi-blocs

Durée du stage : 6 mois, à partir de fin Mars 2020.

Gratification : 570 euros/ mois

Laboratoire d'accueil : Unité de recherche en Statistique, Sensométrie et Chimiométrie (StatSC)
Oniris, site de la Géraudière, rue de la Géraudière CS 82225, Nantes 44322 Nantes Cedex 03

Responsable scientifique : Mohamed Hanafi (envoyer CV, lettre de motivation à mohamed.hanafi@oniris-nantes.fr)

Intérêts du stage

- ⇒ Traitement de données complexes.
- ⇒ Implication dans une thématique d'actualité « analyse des données multiblocs » ouverte sur de nombreux domaines applications

Compétences requises

- ⇒ connaissances des techniques d'analyse de données multidimensionnelles en particulier l'ACP,
- ⇒ pratique de la programmation sous R ou matlab ,
- ⇒ maîtrise des notions de base en algèbre linéaire nécessaires pour comprendre rapidement les concepts des méthodes statistiques non maîtrisées initialement.

Descriptif du sujet

On se limite aux données multi blocs qui se présentent sous la forme d'une collection de tableaux quantitatives portant sur les mêmes individus mais pas nécessairement les même variables. Les travaux de recherche relatifs à l'analyse de cette classe de données structurées en blocs se sont développés d'une manière très significative lors de ces vingt dernières années conduisant à une pléthore de méthodes. Dans le cadre de ce stage on se limite à six méthodes les plus utilisées dans la pratiques.

L'objectif de ce stage est d'évaluer sur des bases objectives six méthodes pour l'analyse des données multiblocs. Il s'agit d'exhiber des critères qui peuvent aider les utilisateurs de ces méthodes de choisir telle ou telle méthode. Pour se faire, partant du fait que l'Analyse en composantes Principales est définissable à partir du problème d'approximation d'une matrice par une matrice de rang inférieur, Il s'agit de formuler les six méthodes multi blocs comme des problèmes cherchant à réaliser le même objectif mais impliquant différentes notion de rang. La mise en œuvre de l'évaluation des méthodes peut se faire par le biais de packages R ou matlab existants.