

Proposition de sujet Impact

Estimation des caractéristiques spectrales d'une matrice de covariance en grande dimension

Note : ce sujet n'est pas conçu pour être couplé à un mémoire de M2 de mathématiques appliquées.

Il sera co-encadré par Mylène Maïda, Professeur, Université Lille 1, Laboratoire Paul Painlevé et Philippe Preux, Professeur, Université Lille 3, équipe Sequel du Laboratoire d'Informatique Fondamentale de Lille et de l'INRIA-Nord Europe.

Emails : `mylene.maida@math.univ-lille1.fr` et `philippe.preux@inria.fr`.

Ce sujet d'Impact est motivé par le fait que les méthodes d'apprentissage automatique et de statistiques doivent faire face aujourd'hui à d'énormes quantités de données de grande dimension et être capables de les traiter efficacement (Big Data). Pour cela, la solution ne passe pas seulement par des ordinateurs plus puissants, mais surtout par des méthodes améliorées, voire renouvelées et des algorithmes mieux pensés, tirant parti de propriétés mathématiques liées aux données traitées.

Les matrices, en particulier les matrices de covariance, jouent un grand rôle en analyse de données et la prise en compte de leurs propriétés spectrales est un élément clé dans les problèmes d'estimation. Pour différentes raisons, la simple utilisation des routines de calculs de valeurs et vecteurs propres d'une matrice peut donner des résultats sans intérêt. Pour les données qui nous intéressent ici (vivant dans des espaces de grande dimension), des estimateurs des vecteurs propres et valeurs propres de matrices de covariance basés sur des résultats pointus de la théorie des matrices aléatoires ont été proposés.

L'objectif de ce travail est de comprendre et d'étudier la mise en œuvre de ce type d'estimateurs, sur des données simulées et sur des jeux de données réels. On s'appuiera sur l'article [1] en ce qui concerne les estimateurs. Différents points sont à étudier pour en permettre la mise en œuvre concrète, qui sera de préférence réalisée en R. Même si on ne s'intéressera pas en détail à la preuve des résultats mathématiques mis en jeu, un bon niveau en mathématiques, en particulier en probabilités et statistiques, est recommandé afin d'être capable de comprendre les résultats théoriques assez en profondeur pour pouvoir les mettre en œuvre, voire d'étudier l'extension de ces résultats à des situations où les hypothèses usuelles ne sont pas vérifiées.

Bibliographie :

[1] Improved Estimation of Eigenvalues and Eigenvectors of Covariance Matrices Using Their Sample Estimates, Xavier Mestre, *IEEE Transactions on Information Theory*, Vol 54, no 11, (2008).