

Approximation par une matrice de faible rang et espaces de projection

Projet

Dans le contexte du machine/statistical learning, les approches basées sur les noyaux semi-définis positifs sont très répandues.

Elles présentent entre autres l'avantage de pouvoir traiter des types variés de données (graphes, séquences d'ADN, ...).

Une limitation de ces approches basées sur les noyaux est qu'elles induisent un coût algorithmique souvent trop important lié au calcul et stockage de la matrice de Gram de taille $n \times n$.

Les objectifs du stage sont :

1. étudier plusieurs techniques d'approximation de la matrice de Gram par une matrice de faible rang (de type Nystrom par exemple),
2. développer une nouvelle analyse théorique de ce type d'approche en se basant sur le papier de Baraud, Giraud et Huet (2014) "Estimator selection in the Gaussian setting",
3. proposer une nouvelle stratégie de choix de paramètres basée sur cette analyse,
4. mettre en oeuvre l'approche globale sur un exemple.

Prérequis

Outre de bonnes compétences en probabilités et statistique, le candidat utiliser R pour réaliser des simulations afin de vérifier empiriquement ses résultats.

Encadrant : Alain Celisse, maître de conférences.

Length: 4-6 mois.

Opportunity: Le stage pourrait donner lieu à une thèse.

Laboratory: MODAL équipe-projet Inria, Lille.

Contact: Alain Celisse (celisse@math.univ-lille1.fr).

Suite: Le sujet proposé pourrait faire l'objet d'une thèse.

Bibliography

1. Celisse, Alain and Robin, Stéphane (2008). Nonparametric density estimation by exact leave-p-or cross-validation. *Computational Statistics and Data Analysis*, 52(5), 2350-2368.
2. Alain Celisse, Tristan Mary-Huard. New upper bounds on the k-nearest neighbor rule for classification. *ArXiv*. 2015