

Nouvelle stratégie de calibration efficace du Lasso

Projet

Dans le contexte du machine/statistical learning, les approches basées sur le Lasso fournissent des solutions “sparse” qui s’avèrent très utiles dans un contexte de grande dimension $d \gg n$ (plus de prédicteurs que d’observations).

Toutefois un inconvénient de ces approches tient la calibration du paramètre de régularisation $\lambda > 0$ dont elles dépendent. Une stratégie couramment utilisée pour cette calibration de λ repose sur la validation-croisée. Mais cette dernière est très coûteuse algorithmiquement, surtout lorsque n est grand.

Les objectifs du stage sont :

1. étudier les performances des approches classiques de calibration de λ reposant sur la validation-croisée,
2. étudier une nouvelle approche de calibration reposant l’introduction de nouvelles “slack-variables”, permettant de réécrire le Lasso comme une régression ridge,
3. mettre en oeuvre l’approche globale proposée sur des exemples afin d’en évaluer les performances.

Prérequis

Outre de bonnes compétences en probabilités et statistique, le candidat utilisera R ou matlab 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.