

Concentration du Leave-one-out autour de son espérance

Projet

Dans le contexte du machine/statistical learning, un enjeu majeur est le réglage des paramètres dont dépendent les algorithmes développés (calibration).

Très souvent, ceci est fait en pratique par validation-croisée sans avoir de réelle garantie quant à la qualité du réglage obtenu.

La première étape pour quantifier la qualité de l'approche obtenue est d'obtenir des bornes avec grande probabilité sur l'écart entre la quantité visée et l'estimateur validation-croisée.

Les objectifs du stage sont :

1. étudier les résultats de concentration exponentielle existant sur la performance de la validation-croisée leave-one-out dans différents contextes (régression, classification),
2. utilisation d'inégalités de moments pour obtenir de nouvelles bornes sur les moments du leave-one-out,
3. adapter une stratégie existante pour obtenir de nouvelles inégalités de concentration exponentielles pour l'estimateur leave-one-out.

Prérequis

Outre de bonnes compétences en probabilités et statistique, le candidat utiliser R pour réaliser des simulations afin de vérifier empiriquement la finesse des bornes obtenues.

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).

Bibliography

1. Celisse, Alain and Robin, Stéphane (2008). Nonparametric density estimation by exact leave-p-out 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