



ALGORITHMES D'APPRENTISSAGE ET MODÈLES STATISTIQUES

ERIC LACOURSE ET CLEMENTINE COURDI

Le 16 juin, 2021

Cette présentation vise à introduire les différents concepts associés à l'apprentissage statistique/automatique, plus principalement les concepts de régularisation/pénalisation ainsi que les concepts associés à la validation croisée. La régularisation et la validation croisée sont des techniques statistiques qui peuvent permettre de sélectionner un modèle statistique/algorithmique « optimal » pour résoudre différents problèmes de recherche, dont ceux en grandes dimensions (jusqu'à des millions de variables, récemment) où le nombre de variables/dimensions p peut largement dépasser le nombre d'unités n dans l'échantillon. Une particularité de ces techniques est que celles-ci peuvent s'appliquer à presque tous les modèles statistiques/algorithmiques en permettant d'estimer les paramètres de ces derniers, alors qu'il serait difficile, voire impossible de le faire avec des modèles multivariés plus « classiques ». Ce sont des outils robustes de sélection de modèle qui peuvent parfois être plus performants que les indices d'ajustement (AIC, BIC, etc.), favorisant potentiellement une meilleure reproductibilité de la recherche scientifique. [GitHub](#)

ERIC LACOURSE est professeur titulaire au Département de sociologie à l'Université de Montréal. Il s'intéresse dernièrement à l'intégration des approches en apprentissage automatique aux données complexes. Il arbore une perspective multidisciplinaire qui intègre des concepts théoriques provenant de la psychologie développementale et de l'éducation, de la psychiatrie, de la criminologie et de la sociologie.