

Question clé

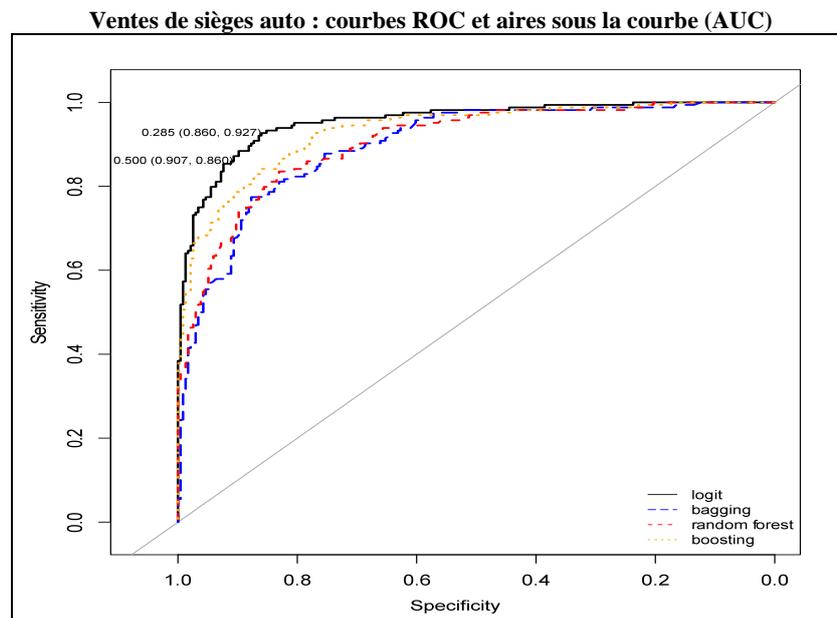
Cet article vise à présenter les techniques dites d'apprentissage machine (*machine learning*) et à comparer leurs performances aux méthodes économétriques traditionnelles. L'objet n'est pas uniquement de tester leurs performances respectives selon les types de problèmes mais également de voir dans quelle mesure ces méthodes peuvent être complémentaires. Il est ainsi montré dans quel cadre les méthodes d'apprentissage machine peuvent être utilisées pour améliorer la modélisation économétrique.

Méthodologie

Des problèmes standard de la littérature d'apprentissage machine sont ré-analysés à partir de l'exploitation des bases de données initialement étudiées. Plusieurs méthodes d'apprentissage machine sont appliquées ainsi que des méthodes économétriques de régression linéaire. Les performances des différentes méthodes sont ensuite comparées.

Principaux résultats

Comme le montrent les applications sur données réelles, les modèles économétriques traditionnels, bien spécifiés, présentent le plus souvent des performances aussi bonnes – voire meilleures – que les modèles d'apprentissage statistique. Toutefois, l'analyse des résultats des méthodes d'apprentissage permettent d'améliorer la spécification des modèles économétriques. La validation croisée offre une alternative aux développements asymptotiques, et les techniques d'agrégation par *bagging* permettent de mieux appréhender d'éventuelles non linéarités. Enfin, les méthodes de « pénalisation » offrent une méthode simple de choix de variables, problème qui devient vite complexe en grande dimension.



Données simulées de 400 points de vente de sièges auto pour bébé.
Données de James *et al.* (2013), jeu de données *carseats*, <https://CRAN.R-project.org/package=ISLR>

Message

Bien qu'historiquement développées dans des contextes très différents, les méthodes économétriques et d'apprentissage machine ont de plus en plus de points communs : la performance algorithmique de l'apprentissage permet de résoudre aujourd'hui des problèmes rencontrés classiquement en économétrie ; une vision probabiliste des techniques d'apprentissage permet d'avoir à disposition des outils qui peuvent éclairer des prévisions d'un œil neuf. Sans remettre en cause la pratique des économètres, ceux-ci gagneraient à s'approprier les nouvelles techniques de l'apprentissage machine afin d'améliorer les prévisions et renforcer certaines interprétations.