

Un outil général de modélisation linéaire généralisée en grande dimension: les composantes supervisées

Tuesday, 13 November 2018 14:20 (50 minutes)

L'estimation des modèles linéaires généralisés en grande dimension nécessite leur régularisation. Pour que le prédicteur linéaire soit interprétable, il est en outre indispensable que celui-ci soit relié à un nombre réduit de dimensions d'interprétation simple. La pénalisation LASSO permet certes de réaliser régularisation et réduction dimensionnelle, les dimensions trouvées étant les variables sélectionnées. Mais LASSO est instable lorsque les colinéarités sont nombreuses parmi les variables explicatives, et risque de sélectionner, à la place d'une variable véritablement explicative, une variable qui lui est très corrélée. A contrario, nous proposons de conserver a priori toutes les variables explicatives, mais de chercher un nombre restreint de dimensions, appelées composantes supervisées, qui soient à la fois très liées aux variables explicatives (donc facilement interprétables) et fortement prédictives des réponses. La recherche de ces dimensions permet également de régulariser le modèle. La méthodologie des composantes supervisées étend, en l'assouplissant, la régression PLS à tous les modèles linéaires généralisés. Les intervalles de confiance bootstrap sur les coefficients permettent une sélection a posteriori des variables participant à la prédiction. Nous présentons l'adaptation de cette méthodologie aux GLMM et au modèle de Cox, en montrant qu'elle constitue un pont pratique entre exploration et modélisation des données.

Co-authors: TROTTIER, Catherine (Inst. Montpellier); CHAUVET, Jocelyn (Inst. Montpellier)

Presenter: BRY, Xavier (Inst. Montpellier)