Séminaire de Probabilités et Statistique

Mardi 24 mars à 14h00

Laboratoire Dieudonné

Salle de conférence - LJAD

Jaouad Mourtada

Università di Genova

Sur l'erreur minimax pour la régression linéaire et les moments négatifs de matrices de covariance aléatoires

Nous considérons le problème de la prédiction linéaire avec variables prédictives aléatoires, où il s'agit de prédire une variable scalaire Y par une fonction linéaire d'un vecteur aléatoire X dans $\mathbb{R}^d$, étant donné un échantillon de n réalisations i.i.d. de loi (X, Y) ; nous nous intéressons en particulier à l'influence de la loi des variables sur la difficulté du problème. Tout d'abord, l'estimateur des moindres carrés est exactement optimal au sens minimax dans le cas "bien spécifié" (où la fonction de régression optimale est linéaire), pour toute loi de X. Nous montrons ensuite que l'erreur minimax est caractérisée par la loi des "leviers" des différents échantillons X_i. Nous en déduisons une borne inférieure sur le risque optimal, valable pour toute loi de X, qui correspond en grande dimension à l'erreur "universelle" atteinte dans le cas de variables prédictives gaussiennes ou plus généralement indépendantes (par la loi de Marchenko-Pastur).

L'étude de bornes supérieures sur l'erreur fait apparaître la nécessité de conditions de régularité sur la loi de X, et se ramène à l'étude de moments négatifs de matrices de covariance aléatoires. Afin de contrôler de tels moments, les inégalités de concentration existantes de type "sous-gaussien" sur les matrices de covariance empiriques s'avèrent insuffisantes. Nous établissons tout d'abord une borne inférieure sur les déviations de telles matrices, valable pour toute loi. Nous établissons ensuite une borne supérieure correspondante, valable sous une hypothèse minimale sur la loi ; notre approche repose sur la technique dite "PAC-Bayésienne" permettant de contrôler des processus empiriques à partir d'une version "lissée" de ces processus. Ces résultats permettent d'obtenir des bornes non-asymptotiques fines sur les moments négatifs de matrices de covariances empiriques et l'erreur optimale en régression.