Sparse Canonical Correlation Analysis

Zongming Ma
University of Pennsylvania, USA –zongming.ma@gmail.com

Chao Gao
Yale University, New Haven, CT, USA - chao.gao@yale.edu

Harrison Zhou*
Yale University, New Haven, CT, USA - huibin.zhou@yale.edu

Zhao Ren
University of Pittsburgh, Pittsburgh, PA, USA - zren@pitt.edu

Canonical correlation analysis (CCA) is a widely used multivariate statistical technique for exploring the relation between two sets of variables. In this talk we consider the problem of estimating the leading canonical correlation directions in high dimensional settings. Recently, under the assumption that the leading canonical correlation directions are sparse, various procedures have been proposed for many high dimensional applications involving massive data sets. However, there has been few theoretical justification available in the literature. In this talk, we first establish rate-optimal non-asymptotic minimax estimation with respect to an appropriate loss function for a wide range of model spaces. Two interesting phenomena are observed. First, the minimax rates are not affected by the presence of nuisance parameters, namely the covariance matrices of the two sets of random variables, though they need to be estimated in the canonical correlation analysis problem. Second, we allow the presence of the residual canonical correlation directions. However, they do not influence the minimax rates under a mild condition on eigengap. In addition, we discuss a new efficient algorithm for sparse CCA with theoretical guarantees. This talk is based on joint work with Chao Gao, Zhao Ren and Harrison Zhou.

Keywords: sparsity; low rank matrix estimation.