Sparse orthogonal factor regression and its extensions

Kun Chen
Department of Statistics, University of Connecticut , US – kun.chen@uconn.edu

Many modern statistical problems in such diverse areas as genetical genomics and financial econometrics can be cast in the framework of multivariate regression, where both the predictors and responses are possibly of high dimensionality. A sparse singular value decomposition (SVD) of the regression coefficient matrix, which is of low rank and with sparse singular vectors, can play a key role for simultaneous dimension reduction and variable selection. We introduce sparse orthogonal factor regression (SOFAR), a unified approach to regularized multivariate regression, to estimate such a sparse SVD structure. We formulate the regularization procedure as an orthogonality constrained optimization problem and employ sparsity-inducing penalties of a general, flexible form, which specializes to some important cases. The SOFAR methodology is connected to a variety of multivariate techniques, including biclustering, sparse principal components, factor analysis, and vector autoregression, and yields useful new methods in these contexts. We derive nonasymptotic error bounds for the regularized estimator under conditions that control the degrees of nonconvexity and nonidentifiability in the SVD problem. An efficient optimization algorithm using the alternating direction method of multipliers is developed, and its convergence properties are established. Simulation studies show that the SOFAR methodology substantially outperforms some existing methods, and its usefulness is demonstrated by an analysis of yeast expression quantitative trait loci data. Extensions to other structured reduced-rank estimation as well as Bayesian sparse and low-rank modeling will also be discussed. This is a joint work with Wei Lin and Jinchi Lv.

Keywords: dimension reduction; low-rank matrix approximation; multivariate regression; singular value decomposition.