Fitting High-Dimensional Regression Models via Supervised Dimension Reduction

Yanzhu Lin

Department of Statistics

Purdue University

West Lafayette, Indiana 47907

Abstract

Analysis of modern biological data often involves ill-posed problems due to high dimensionality and multicollinearity. In order to solve the problems, we develop the regression-based modeling methods by constructing sparse orthogonal components through a penalized framework and then treating these components as predictors to fit the model. We propose a penalized orthogonal components regression (POCRE) for high dimensional data with continuous response variables, and a generalized orthogonal components regression (GOCRE) and a sparse-GOCRE for high dimensional data with categorical response variables. Our methods not only provide the solution to the high dimensionality issue but also have the ability to identify the variables that are highly correlated or share some common coherent pattern. Both simulated and real data analysis are presented to illustrate competitive performance of the new approaches and compare them with several other existing methods.

This is a joint work with Professor Min Zhang and Professor Dabao Zhang in the Department of Statistics at Purdue University.