Simultaneous Dimension Reduction and Prediction Optimization: Method and Application to High Dimensional Data

Joseph Ryan Lansangan*
University of the Philippines, Quezon City, Philippines – jglansangan@up.edu.ph

Erniel Barrios
University of the Philippines, Quezon City, Philippines – ebbarrios@up.edu.ph

A constrained optimization method is developed to address estimation problems when dealing with high dimensional input in regression. The method simultaneously considers dimension reduction (among the input variables) while maintaining relatively high predictive ability (in the fitted target variable). The method uses an alternating and iterative solution via soft thresholding, and yields fitted models with sparse regression coefficients. Results of the simulation studies show that the method may outperform other constrained regression methods in terms of predictive ability and selection of input variables. That is, the method selects a smaller set of input variables that both captures the dimensionality of the inputs (high retained variability from the inputs) and gives the most predictive model for the target variable (lowest squared prediction error). The method is applied to model cross-country quality of life (with emphasis on mortality). Environment, lifestyle, health care, health status, health policy, and morbidity indicators are considered as inputs. Results from the empirical data show that quality of life may be explained primarily by health condition of women, welfare of children, and government spending on health.

Keywords: high dimensional data; regression; variable selection; quality of life.