Box-Cox Transformations in High-Dimensional Linear Models

The LASSO is a popular tool for analyzing high dimensional data, but its performance depends on how well the relationship between the explanatory and response variables can be described by a linear model. In cases where this relationship is strongly non-linear, it is often possible to find a simple transformation of the response that makes things more linear. One such family of transformations that is widely used is due to Box and Cox (1964). The Box-Cox transformation has been studied extensively in the traditional linear model framework (i.e. full rank design matrix). However, little has been written about how this transformation performs when applied in a high-dimensional setting. We use simulation and theoretical work to describe the difficulties that arise when the Box and Cox methodology is applied directly to models fit by the LASSO. We then give suggestions for how to address some of these difficulties.