

Sure Independence Screening with NP-dimensionality

by

Jianqing Fan

Princeton University

Department of Operations Research and Financial Engineering

Princeton, NJ 08544

`jqfan@princeton.edu`

Abstract

Ultrahigh dimensional variable selection plays an increasingly important role in contemporary scientific discoveries and statistical research. A simple and effective method is the correlation screening. For generalized linear models, we propose a more general version of the independent learning with ranking the maximum marginal likelihood estimates or the maximum marginal likelihood itself. We show that the proposed methods possess the sure screening property with vanishing false selection rate. We quantify explicitly the extent to which the dimensionality can be reduced by independence screening, which depends on the covariance matrix of covariates and true parameters. An iterative version of large-scale screening and moderate-scale selection is introduced to deal with the difficult situation where independence screening might fail. The effectiveness of the methods is demonstrated by several simulation examples and case studies.