Li, Gaorong; Peng, Heng; Zhang, Jun; Zhu, Lixing
Robust rank correlation based screening. (English) Zbl 1257.62067

Summary: Independence screening is a variable selection method that uses a ranking criterion to select significant variables, particularly for statistical models with nonpolynomial dimensionality or “large \( p \), small \( n \)” paradigms when \( p \) can be as large as an exponential of the sample size \( n \). We propose a robust rank correlation screening (RRCS) method to deal with ultra-high dimensional data. The new procedure is based on the Kendall \( \tau \) correlation coefficient between response and predictor variables rather than the Pearson correlation of existing methods. The new method has four desirable features compared with existing independence screening methods.

First, the sure independence screening property can hold only under the existence of a second order moment of predictor variables, rather than exponential tails or alikeness, even when the number of predictor variables grows as fast as exponentially of the sample size. Second, it can be used to deal with semiparametric models such as transformation regression models and single-index models under monotonic constraint to the link function without involving nonparametric estimation even when there are nonparametric functions in the models. Third, the procedure can be largely used against outliers and influence points in the observations. Last, the use of indicator functions in rank correlation screening greatly simplifies the theoretical derivation due to the boundedness of the resulting statistics, compared with previous studies on variable screening. Simulations are carried out for comparisons with existing methods and a real data example is analyzed.

MSC:
62H20 Measures of association (correlation, canonical correlation, etc.)
62G08 Nonparametric regression and quantile regression
62G35 Nonparametric robustness
62J12 Generalized linear models (logistic models)
62F35 Robustness and adaptive procedures (parametric inference)

Keywords:
variable selection; dimensionality reduction; semiparametric models; large \( p \) small \( n \); SIS

Software:
Excel

Full Text: DOI arXiv Euclid

References:


Edited by FIZ Karlsruhe, the European Mathematical Society and the Heidelberg Academy of Sciences and Humanities © 2022 FIZ Karlsruhe GmbH


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.