

1158-62-79

Adriano Zanin Zambom* (adriano.zambom@csun.edu), 6651 Neddy Ave, West Hills, CA 91307, and **Gregory Matthews**. *Sure Independence Screening in the Presence of Missing Data*.

Variable selection in ultra-high dimensional data sets is an increasingly prevalent issue with the readily available data arising from, for example, genome-wide associations studies or gene expression data. When the dimension of the feature space is exponentially larger than the sample size, it is desirable to screen out unimportant predictors in order to bring the dimension down to a moderate scale. In this paper we consider the case when observations of the predictors are missing at random. We propose performing screening using the marginal linear correlation coefficient between each predictor and the response variable accounting for the missing data using maximum likelihood estimation. This method is shown to have the sure screening property. Moreover, a novel method of screening that uses additional predictors when estimating the correlation coefficient is proposed. Simulations show that simply performing screening using pairwise complete observations is out-performed by both the proposed methods and is not recommended. Finally, the proposed methods are applied to a gene expression study on prostate cancer. (Received February 21, 2020)