Statistics > Machine Learning

Predictive Correlation Screening: Application to Two-stage Predictor Design in High Dimension

Hamed Firouzi, Bala Rajaratnam, Alfred Hero

(Submitted on 10 Mar 2013 (v1), last revised 10 Apr 2013 (this version, v2))

We introduce a new approach to variable selection, called Predictive Correlation Screening, for predictor design. Predictive Correlation Screening (PCS) implements false positive control on the selected variables, is well suited to small sample sizes, and is scalable to high dimensions. We establish asymptotic bounds for Familywise Error Rate (FWER), and resultant mean square error of a linear predictor on the selected variables. We apply Predictive Correlation Screening to the following two-stage predictor design problem. An experimenter wants to learn a multivariate predictor of gene expressions based on successive biological samples assayed on mRNA arrays. She assays the whole genome on a few samples and from these assays she selects a small number of variables using Predictive Correlation Screening. To reduce assay cost, she subsequently assays only the selected variables on the remaining samples, to learn the predictor coefficients. We show superiority of Predictive Correlation Screening relative to LASSO and correlation learning (sometimes popularly referred to in the literature as marginal regression or simple thresholding) in terms of performance and computational complexity.

Comments:	31 pages, 9 figures, Appearing in Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)
Subjects:	**Machine Learning (stat.ML)**
Cite as:	**arXiv:1303.2378 [stat.ML]**
	(or **arXiv:1303.2378v2 [stat.ML]** for this version)

Submission history

From: Hamed Firouzi [view email]
[v1] Sun, 10 Mar 2013 21:31:37 GMT (120kb,D)
[v2] Wed, 10 Apr 2013 07:05:51 GMT (117kb,D)

Which authors of this paper are endorsers?