Statistics > Methodology

Homogeneity in Regression

Tracy Ke, Jianqing Fan, Yichao Wu

(Submitted on 29 Mar 2013)

This paper explores the homogeneity of coefficients in high-dimensional regression, which extends the sparsity concept and is more general and suitable for many applications. Homogeneity arises when one expects regression coefficients corresponding to neighboring geographical regions or a similar cluster of covariates to be approximately the same. Sparsity corresponds to a special case of homogeneity with a known atom zero. In this article, we propose a new method called clustering algorithm in regression via data-driven segmentation (CARDS) to explore homogeneity. New mathematics are provided on the gain that can be achieved by exploring homogeneity. Statistical properties of two versions of CARDS are analyzed. In particular, the asymptotic normality of our proposed CARDS estimator is established, which reveals better estimation accuracy for homogeneous parameters than that without homogeneity exploration. When our methods are combined with sparsity exploration, further efficiency can be achieved beyond the exploration of sparsity alone. This provides additional insights into the power of exploring low-dimensional strucuture in high-dimensional regression: homogeneity and sparsity. The newly developed method is further illustrated by simulation studies and applications to real data.

Subjects: **Methodology (stat.ME)**
Cite as: **arXiv:1303.7409 [stat.ME]**
(or **arXiv:1303.7409v1 [stat.ME]** for this version)

Submission history

From: Tracy Ke [view email]
[v1] Fri, 29 Mar 2013 14:34:10 GMT (95kb,D)

Which authors of this paper are endorsers?