Elementary Estimators for High-dimensional Statistical Models

Eunho Yang
IBM T.J. Watson Research Center, Yorktown Heights, USA, eunhyang@us.ibm.com

Aurélie C. Lozano
IBM T.J. Watson Research Center, Yorktown Heights, USA, aclozano@us.ibm.com

Pradeep Ravikumar*
University of Texas, Austin, USA, pradeepr@cs.utexas.edu

We consider the problem of learning high-dimensional statistical models, where the number of variables could be potentially larger than the number of observations. This class of problems has attracted considerable attention over the last decade, with state of the art statistical estimators based on solving regularized convex programs. Scaling these typically non-smooth convex programs to the very large-scale problems of the Big Data era comprises an ongoing and rich area of research.

In contrast to this two-stage approach of first devising statistically efficient estimators, and then devising computationally efficient optimization methods to solve these estimators, we attempt to address this scaling issue at the source, by asking whether one can build simpler closed-form estimators, that yet come with statistical guarantees that are nonetheless comparable to regularized likelihood estimators. Surprisingly, we answer this question in the affirmative. We analyze our estimators in the high-dimensional setting, and moreover provide empirical corroboration of their statistical and computational performance guarantees.

Keywords: High-dimensional Statistical Models, Comptastical Learning, Computational Efficiency, Optimization