Robust estimation of high dimensional Generalized Linear Models

Eunho Yang*
University of Texas, Austin, USA - eunho@cs.utexas.edu

Ambuj Tewari
University of Michigan, Ann Arbor, USA - tewaria@umich.edu

Pradeep Ravikumar
University of Texas, Austin, USA - Pradeepr@cs.utexas.edu

We study *robust high-dimensional estimation* of generalized linear models (GLMs); where a small number k of the n observations can be arbitrarily corrupted, and where the true parameter is high dimensional in the "$p \gg n$" regime, but only has a small number s of non-zero entries. There has been some recent work connecting robustness and sparsity, in the context of linear regression with corrupted observations, by using an explicitly modeled outlier response vector that is assumed to be sparse. Interestingly, we show, in the GLM setting, such explicit outlier response modeling can be performed in *two distinct ways*. For each of these two approaches, we give ℓ_2 error bounds for parameter estimation for general values of the tuple (n, p, s, k). The first approach, based on modeling gross errors in the parameter space, leads to a convex problem but requires more stringent conditions for the bounds to hold. The second approach, based on modeling gross errors in the output space, requires more benign conditions but leads to a non-convex problem in general. However, it turns out that non-convexity is not a big impediment and a simple projected gradient method converges to one of the global optima up to statistical precision.

Keywords: GLM; high dimension; sparse estimation; robustness.