Robust regression revisited

Po-Ling Loh
Cambridge, United Kingdom
pll28@cam.ac.uk

This talk will discuss two projects on robust estimation in the presence of contaminated data, bringing new ideas beyond the framework of traditional M-estimation.

In the first part of the talk, we study the problem of linear regression in a setting where both the covariates and responses may be heavy-tailed and/or adversarially contaminated. We show how to modify the Huber regression estimator by first applying an appropriate "filtering" procedure to the data based on the covariates, and show that in low-dimensional settings, the filtered Huber regression estimator achieves near-optimal error rates. We further show that the commonly used least trimmed squares and least absolute deviation estimators may similarly be made robust to contaminated covariates via the same covariate filtering step.

In the second part of the talk, we study a variant of Newton's method for robust empirical risk minimization, where at each iteration of the optimization algorithm, we replace the gradient and Hessian of the objective function by robust estimators taken from literature on robust mean estimation for multivariate data. After proving a general theorem about the convergence of successive iterates to a small ball around the population-level minimizer, we study consequences of our theory in generalized linear models, when data are generated from Huber's epsilon-contamination model and/or heavy-tailed distributions. We also propose an algorithm for obtaining robust Newton directions based on the conjugate gradient method, which may be more appropriate for high-dimensional settings, and provide conjectures about the convergence of the resulting algorithm. Our algorithm enjoys the fast rates of convergence for successive iterates often achieved by second-order algorithms for convex problems, i.e., quadratic convergence in a neighborhood of the optimum, with a stepsize which may be chosen adaptively via backtracking linesearch.

Joint work with Ankit Pensia (IBM Research), Varun Jog (Cambridge), Eirini Ioannou (Edinburgh) and Muni Sreenivas Pydi (Paris Dauphine - PSL).