EPSTEIN INSTITUTE SEMINAR ▪ ISE 651

On Fast Rates In Empirical Risk Minimization Beyond Least-Squares

ABSTRACT - This talk will be focused on ``fast'' learning rates for empirical minimization (M-estimation in the statistical terminology) with some convex losses beyond the square loss. I will show that in the parametric setup, where the model is specified up to a finite-dimensional parameter, the excess risk essentially decreases at the rate O(d/n) -- as guaranteed in the asymptotic regime by the central limit theorem -- whenever the sample size n grows linearly with the parameter dimension d. Furthermore, this result can be naturally extended to misspecified models by replacing d with a suitable notion of ``effective dimension'', and can also be generalized to the non-parametric setup. In contrast to most of the existing results in the literature, this result only requires local assumptions on the distribution of the loss gradient and Hessian at the optimal predictor -- namely, this distribution must be sufficiently light-tailed. The absence of global assumption is achieved by requiring that the loss function satisfies a version of self-concordance, allowing one to control the precision of local quadratic approximations of the population and empirical risks. Self-concordant functions were first introduced by Nesterov and Nemirovski in their seminal work on interior-point methods (1994). In 2010 this property were brought to the attention of statistical learning theorists by Francis Bach, who pointed out that a modified version of self-concordance holds for some losses arising in generalized linear models (most notably logistic regression) and robust estimation, and used it to prove fast statistical rates in such M-estimators. We go beyond this earlier work in several directions. First, and most importantly, we note that the earlier results of Bach required the quadratic growth of n with respect to d, and then suggest a sharper analysis allowing for O(d) sample size. Second, we compare the classes of canonically self-concordant (à la Nesterov and Nemirovski) and ``pseudo'' self-concordant (à la Bach) losses, and show that the critical sample size for the latter ones is somewhat larger. Motivated by this finding, we construct the analogues of the logistic and Huber losses that are canonically self-concordant. Finally, we will briefly discuss the extension of these results to l2-regularized ERM obtained in the joint work with Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi (COLT 2019). In conclusion, if time permits, I will briefly discuss our recent result with Alessandro Rudi on robust covariance estimation. This construction has applications in linear regression with heavy-tailed random design, and potentially could allow to study M-estimators under heavy-tailed data distributions.

Dr. Dmitrii Ostrovskii
Postdoctoral Scholar
USC Viterbi School of Engineering

SPEAKER BIO - Dmitrii Ostrovskii has defended his MSc at the Moscow Institute of Physics and Technology (MIPT) in 2014 and PhD at the University of Grenoble Alpes under the supervision of Anatoli Juditsky and Zaid Harchaoui in 2018. His work during PhD has been concerned with structure-adaptive signal recovery problems. He then did a postdoc at Inria Paris, hosted by Francis Bach, where he worked on statistical learning theory, robust estimation, and efficient algorithms for large-scale optimization. Dmitrii recently joined the USC Viterbi School of Engineering as a postdoc, hosted by Dr. Meisam Razaviyayn.

USC Viterbi
School of Engineering
Daniel J. Epstein Department of Industrial and Systems Engineering

TUESDAY, OCTOBER 8, 2019
3:30 PM – 4:50 PM
USC ANDRUS GERONTOLOGY CENTER (GER), ROOM 206