



THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine Learning and Statistics Seminar

Room 1 ,Ziskind Building
on Wednesday, Nov 16, 2016
at 11:15

Dan Garber
Toyota Technological Institute at Chicago

Faster Projection-free Machine Learning and Optimization

Abstract:

Projected gradient descent (PGD), and its close variants, are often considered the methods of choice for solving a large variety of machine learning optimization problems, including empirical risk minimization, statistical learning, and online convex optimization. This is not surprising, since PGD is often optimal in a very appealing information-theoretic sense. However, for many problems PGD is infeasible both in theory and practice since each step requires to compute an orthogonal projection onto the feasible set. In many important cases, such as when the feasible set is a non-trivial polytope, or a convex surrogate for a low-rank structure, computing the projection is computationally inefficient in high-dimensional settings. An alternative is the conditional gradient method (CG), aka Frank-Wolfe algorithm, that replaces the expensive projection step with a linear optimization step over the feasible set. Indeed in many problems of interest, the linear optimization step admits much more efficient algorithms than the projection step, which is the reason to the substantial regained interest in this method in the past decade. On the downside, the convergence rates of the CG method often fall behind that of PGD and its variants.

In this talk I will survey an ongoing effort to design CG variants that on one hand enjoy the cheap iteration complexity of the original method, and on the other hand converge provably faster, and are applicable to a wider variety of machine learning settings. In particular I will focus on the cases in which the feasible set is either a polytope or a convex surrogate for low-rank matrices. Results will be demonstrated on applications including: LASSO, video co-localization, optical character recognition, matrix completion, and multi-class classification.