Dynamic Sample Size Selection in Optimization Methods for Machine Learning

Statistical Learning, Stochastic Optimization and Active-Set Methods for Nonlinear Programming

Jorge Nocedal
Northwestern University
2145 Sheridan Rd, Evanston, ILL 60208

Abstract

We present a methodology for using varying sample sizes in mini-batch optimization methods for large scale machine learning problems. The overall goal is to keep computational cost to a minimum while ensuring a fast rate of convergence and allowing for the parallel evaluations of functions, gradients and Hessian-vector products.

The first part of the talk deals with the delicate issue of dynamic sample selection in the evaluation of the function and gradient. We propose a rigorous criterion for increasing the sample size based on variance estimates computed during the computation of gradient approximations. We establish an $O(1/\epsilon)$ complexity bound on the *total cost* of the algorithm. The second part of the talk discusses the use of sub-samples in the Hessian-vector computation of a Newton-CG method for L_1 regularized problems. Numerical tests on computer vision, document classification and speech recognition problems illustrate the performance of the algorithms.

We conclude by discussing the application of our approach in the context of deep learning, and present results for a recursive neural net.