Master's Thesis

Efficient Implementation of Stochastic Gradient Descent Method with Variance Reduction

Guidance

Professor Nobuo YAMASHITA

Takahiro YAMAGUCHI

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2017

Abstract

In this paper, we consider large-scale optimization problems whose objective function is expressed as the sum of a huge number of convex functions. Such problems have applications in diverse fields, like stochastic optimization, machine learning, statistical reasoning and signal processing. One way to solve these problems is to use the stochastic gradient descent (SGD) method, which at each iteration takes just one component of the objective function and uses its gradient as the search direction. However, since it is different from the gradient of the full objective function, the method converges slowly. To overcome this drawback, the stochastic variance reduction gradient (SVRG) descent method has been proposed. It calculates the true gradient of the objective function only sporadically, using it to improve the search direction in a large number of iterations. It achieves convergence faster than SGD method. However, when the current iterate is far from the point where the true gradient was evaluated, SVRG tends to be slow. Moreover, even if the gradient of each function is sparse, SVRG doesn't exploit its sparsity. This is because the gradient of the whole objective function is usually dense.

Here, we propose a new and efficient implementation of SVRG, and prove that it has the same theoretical properties of the original SVRG method. By introducing a certain parameter, the new method includes SVRG and SGD as special cases. As a result, we can use the SGD method at the beginning and then switch to the SVRG method at latter iterations. We show that the proposed method achieves linear convergence like the SVRG, when the parameters are updated properly. Moreover, when the gradient is sparse, we propose a way to reduce computational complexity while generating the same sequence of points as the new SVRG method. Finally, we report some numerical experiments and conclude the proposed method is promising.