Stochastic gradient descent with Polyak’s learning rate.

Summary: Stochastic gradient descent (SGD) for strongly convex functions converges at the rate $O(1/k)$. However, achieving good results in practice requires tuning the parameters (for example the learning rate) of the algorithm. In this paper we propose a generalization of the Polyak step size, used for subgradient methods, to stochastic gradient descent. We prove a non-asymptotic convergence at the rate $O(1/k)$ with a rate constant which can be better than the corresponding rate constant for optimally scheduled SGD. We demonstrate that the method is effective in practice, and on convex optimization problems and on training deep neural networks, and compare to the theoretical rate.

MSC:
90Cxx Mathematical programming
65Kxx Numerical methods for mathematical programming, optimization and variational techniques
68Txx Artificial intelligence

Keywords:
stochastic gradient descent; learning rate; Polyak’s learning rate; optimization; strong convexity

Full Text: DOI

References: