Stochastic Newton with Arbitrary Sampling

Igor Melnikov

MIPT

Goal

$f_i(x) = \log\left(1 + \exp(-b_i a_i^\top x)\right) + \frac{\lambda}{2}\|x\|$
Conventional sampling: $p(i) = \frac{1}{n}$
Important sampling: $p(i) = \frac{L_i}{\sum_{j=1}^{n} L_j}$

Improving NS algorithm by changing sample strategy.

Literature

Stochastic Newton and Cubic Newton Methods with Simple Local Linear-Quadratic Rates - Dmitry Kovalev, Konstantin Mishchenko, Peter Richtaric

An Incremental Quasi-Newton Method with Local Superlinear Convergence Rate - Aryan Mokhtari, Mark Eisen, Alejandro Ribeiro

Empirical Risk Minimization problem

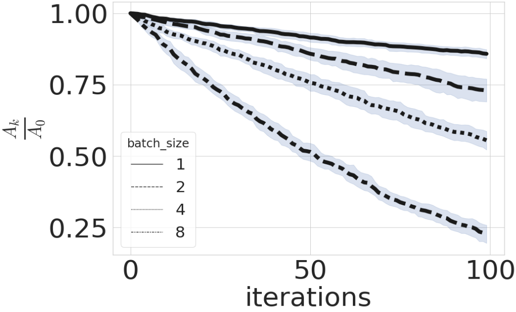
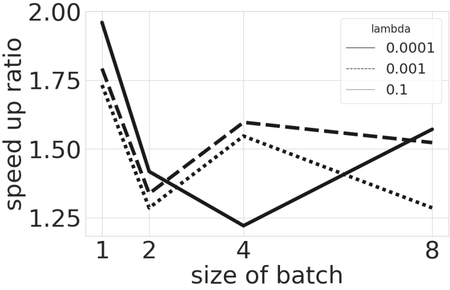


Figure 1: Convergence rate NS algorithm ($\lambda = 0.001$)

$f_i(x) = \log\left(1 + \exp(-b_i a_i^\top x)\right) + \frac{\lambda}{2}\|x\|$
Conventional sampling: $p(i) = \frac{1}{n}$
Important sampling: $p(i) = \frac{L_i}{\sum_{j=1}^{n} L_j}$

Empirical Risk Minimization problem



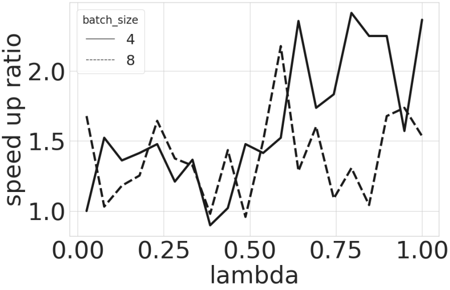


Figure 2: Speed up of NS algorithm

$$f_i(x) = \log\left(1 + \exp(-b_i a_i^\top x)\right) + \frac{\lambda}{2}\|x\|$$

Conventional sampling: $p(i) = \frac{1}{n}$

Important sampling: $p(i) = \frac{L_i}{\sum_{j=1}^{n} L_j}$

Batch size	1	2	4	8
Speed up	2.42	1.78	1.62	1.39