Stochastic Quasi Newton methods for Neural Networks

Konstantin Yakovlev
Optimization Class Project. MIPT

Introduction

The project considers a stochastic quasi-Newtonian method. This project draws heavily on the work of [1]. The algorithm given here uses the classical BFGS update formula in its bounded form. It is also efficient, robust, scalable, and has promising prospects in machine learning. In this paper, we propose to implement the above algorithm and compare it with the SGD algorithm on the example of logistic regression. In addition, we will compare two optimizers of a neural network – SQGN [2] and SGD.

Problem statement

Problem statement from [1]. I'll try to compress it.

Stochastic quasi-Newton method

Description from [1].

Algorithm

Two algorithms from [1]

Convergence

Theorem from [1]

Basic experiment

Logistic regression: dataset description, results + plot(from colab).

Main experiment

Dataset description + neural network description + parameters of SGD and SQGN.

Results

Results of the nain experiments: plot + table of metrics(execution time, loss value depending on the epoch. For example, 30, 50, 70, 100).

Conclusion

References

[1] Richard H. Byrd, S. L. Hansen, Jorge Nocedal, and Yoram Singer. A stochastic quasi-newton method for large-scale optimization. *SIAM J. Optim*, 26(2):1008–1031, 2016.

[2] Christopher Thiele, Mauricio Araya-Polo, and Detlef Hohl. Deep neural network learning with second-order optimizers - a practical study with a stochastic quasi-gauss-newton method. *CoRR*, abs/2004.03040, 2020.