CSE 6740: Final Project Proposal

Team Members: Zixuan Wang, Mingzheng Michael Huang

1 Motivation

The goal of our project is to analyze the effectiveness of different optimization methods on the handwritten digit classification (MNIST) problem.

In deep learning models, we always use first-order optimization, such as Stochastic Gradient Descent, to calculate the gradient of the objective function. It's widely used since its computational cost is relatively lower than that of second-order optimization, and it is suitable for large-scale optimization problems.

On the contrary, second-order optimization uses both the first and second derivatives of the objective function to find the optimal solution. Second-order methods can often converge in fewer iterations compared to first-order methods and help with non-convex problems.

We aim to experiment with various optimization methods (both first-order and second-order) to find out the trade-off between computational cost and convergence speed on a medium-scale problem, e.g., handwritten digit classification.

2 Approach

- Use PyTorch framework and MNIST dataset to train our shallow model, with several linear layers and non-linear activation.
- Implement and experiment with first-order optimization methods (Gradient Descent) and second-order optimization methods (Hessian, Gauss-Newton, BFGS, LBFGS, AdaGrad, Shampoo, etc.).
- Draw the training loss and testing accuracy curve for every method and theoretically analyze what causes the performance disparity between different methods.
- Calculate the FLOPs for every method and draw the conclusion of the suitable scenario for each method.

3 Inferences

- Learn both theories and code implementation of first-order optimization and second-order optimization.
- Learn the motivation of design behind different second-order optimization methods.
- Gain deeper insight into the trade-off between computational cost and convergence speed during the optimization process.