Visualize and Diagnose Optimization Algorithms for Objective Functions

Yali Bian
Virginia Tech
yali@vt.edu

Siyu Mi
Virginia Tech
siyu6@vt.edu

Sirui Yao
Virginia Tech
ysirui@vt.edu

1. Introduction

The optimization algorithm is one of the most important concepts in deep learning to optimize neural networks. people can get various optimization algorithms in every deep learning library easily.These algorithms, however, are often used as black-box optimizers. Thus, it's a little difficult to understand every algorithms strengths and weaknesses. We try 7 different optimization algorithms on different objective functions (strongly convex, convex, non-convex) with detailed visualizations and diagnostics. We hope this would help people comprehend the behavior of different optimization algorithms intuitively and put the algorithms to use in the certain environment..

2. Related works

Before we start the project, we read some relevant papers. Sebastian Ruder in ”An overview of gradient descent optimization algorithms” [8] provides the behavior of different algorithms that will allow using. This paper looks at different variants of the optimization algorithm, introduces the most common optimization algorithms and investigates additional strategies for optimizing gradient descent. It's pity that in its visualization part, there are almost 2D figures and they cannot interact with users (The paper also post in the author's website). "An Interactive Tutorial on Numerical Optimization"[6] is a pretty good example of visualizing optimization algorithms. However, we find it doesn't conclude some classical algorithms, such as Adagrad, Adam, and BFGS etc.

In addition to the above two articles. What is worth challenging is the learning rate. If learning rate is too small, it would lead slow convergence. And if learning rate is too large, it may be hinder convergence and cause the loss function to fluctuate around the minimum or even to diverge. So we hope to find a right learning rate for each algorithm. What's more, we may not want to apply same learning rate to all parameters when the data is sparse and the features have different frequencies. Meanwhile, we would hope to update the learning rate for particular ones of them.

3. Proposed work

For this project, we will build a GUI that compares different optimization algorithms on different objective functions (strongly convex, convex, non-convex) with detailed visualizations and diagnostics.
Optimization algorithms include but not limited to:

- Gradient Descent
- Newton's method
- BFGS
- Nesterov accelarated gradient
- Adagrad
- Adam
- RMSprop

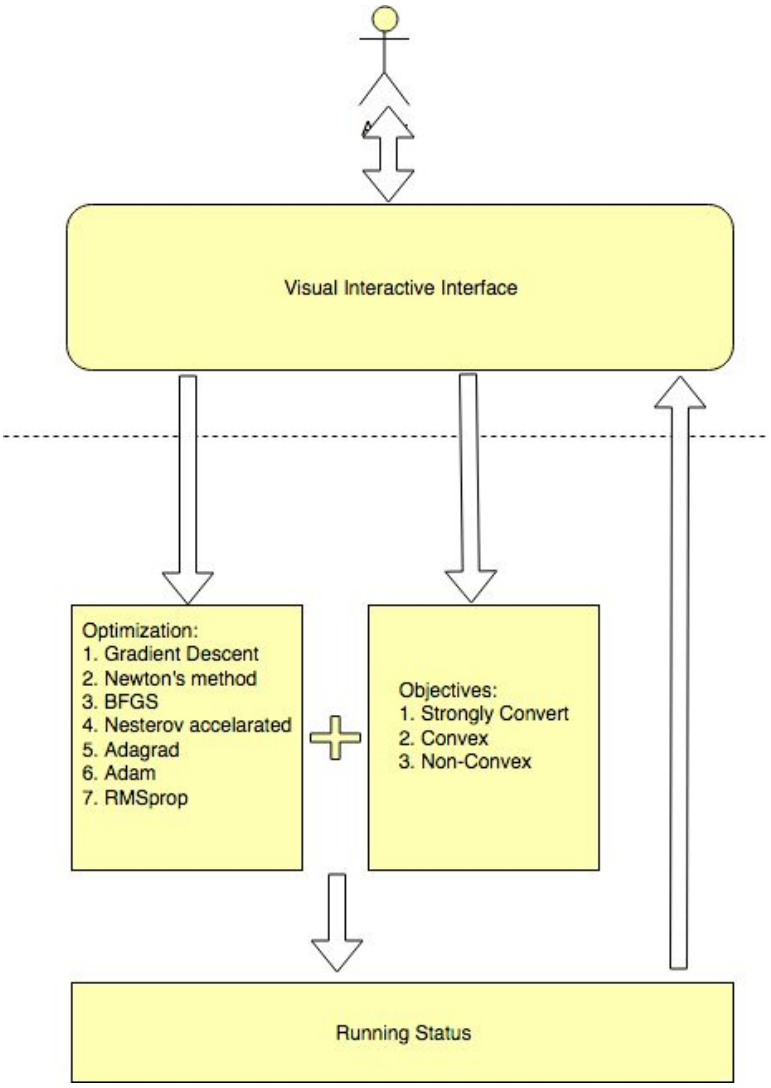
Objectives types to be tested with:

- strongly convex
- convex
- non-convex

We will run each optimization algorithm on several typical example of each objective type, for each algorithm, we will compare different settings of hyperparameters. Also, if time allows, we will support user-defined objectives/initialization and return real time results for users to try more experiments than those provided.

For the GUI part, we will provide an web-based visual analytics system to give users a clear view of the behavior and properties of different optimization algorithms on different objectives, including analysis of the influence of hyperparameters and visual diagnosis of algorithm performance. Ideally, it should also support some extend of customization. For visualization, there will be three important parts included for machine learning researchers to explore and interact with "Optimization Algorithms" and "Objective Functions".

1. More interactions and views for researchers to select optimization algorithms and input their own objective functions effectively and efficiently.
2. Multiple and coordinated views will be used to explore and compare different optimization algorithms on one specific objective function or different objective functions on one optimization algorithms.
3. More visual clues/hints will be showed to recommend researchers about the status of searching loops, which would be better for researchers to diagnose the relevant optimization algorithm.



4. Work Plan

Yali Bian will be responsible for visualization and web design, Siyu Mi and Sirui Yao will write the script for all experiments. All members will work together to link the data generated from the back end with the front end web.

5. Reference

[1] Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*(2014).

[2] Zeiler, Matthew D. "ADADELTA: an adaptive learning rate method." *arXiv preprint arXiv:1212.5701* (2012).

[3] Acerbi, L. & Ma, W. J. (2017). Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. arXiv preprint, arXiv:1705.04405.

[4] Visualize optimization algorithms in MATLAB: https://github.com/lacerbi/optimviz

[5]Visualizing and Animating Optimization Algorithms with Matplotlib:
http://tiao.io/notes/visualizing-and-animating-optimization-algorithms-with-matplotlib/
[6] An Interactive Tutorial on Numerical Optimization:
http://www.benfrederickson.com/numerical-optimization/
[7] Virtual Library of Simulation Experiments: https://www.sfu.ca/~ssurjano/optimization.html
[8] Sebastian Ruder “An overview of gradient descent optimization algorithms⋆” arXiv:1609.04747
Blog:http://ruder.io/optimizing-gradient-descent/