Project: Robustness of Stochastic Gradient and Newton Methods

Last updated: June 18, 2019
Goal

- Check how sensitive stochastic gradient and Newton methods are to the parameters
From our slides we see that there are many variants of Newton methods for NN. It is difficult to see which one should be used. In fact, all these variants were proposed for handling large data or deep networks. If the standard Newton is not robust enough, then probably those variants are not either. Thus we decide to use the standard Newton. This is doable because we consider only small sets.
All settings and data sets are the same as project 4

The only exception is that we add the regularization term back

Let’s solve the optimization problem

\[
\min_\theta \frac{1}{2C} \theta^T \theta + \frac{1}{l} \sum_{i=1}^{l} \xi(z^{L+1,i}(\theta); y^i, Z^{1,i})
\]

with

\[C = 0.01/l\]

This is the default setting of the code so you shouldn’t need any changes
The Newton implementation is available at https://github.com/cjlin1/simpleNN

We check the following parameters

- Percentage of data for subsampled Hessian: 20%, 50%
- With/without Levenberg-Marquardt method
Everything is the same as before:

simple stochastic gradient + momentum

We check the following initial learning rates

\[ 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} \]
Try to see the relation between training time and test accuracy

You now have

4 settings for Newton

and

4 settings for stochastic gradient

You need to design maybe figures for the comparison
The following teams 6, 1, 2, 7, 5 please do a 15-minute presentation (13-minute the contents and 2-minute Q&A)

Page limit: $\leq$ 3 pages