Topic Title

# ECE-GY 7143 Advanced Machine Learning

# Homework 3

*Author:*
Jing Wang
*NetID:*
jw5665

*Course Instructor*
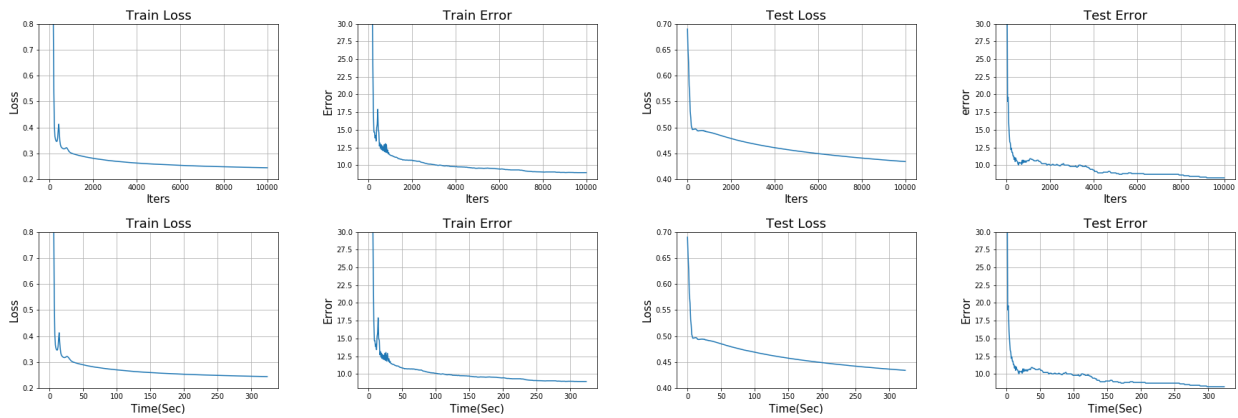Anna Choromanska

April 15, 2021

# 1 GD



Figure 1: GD with learning rate=0.005

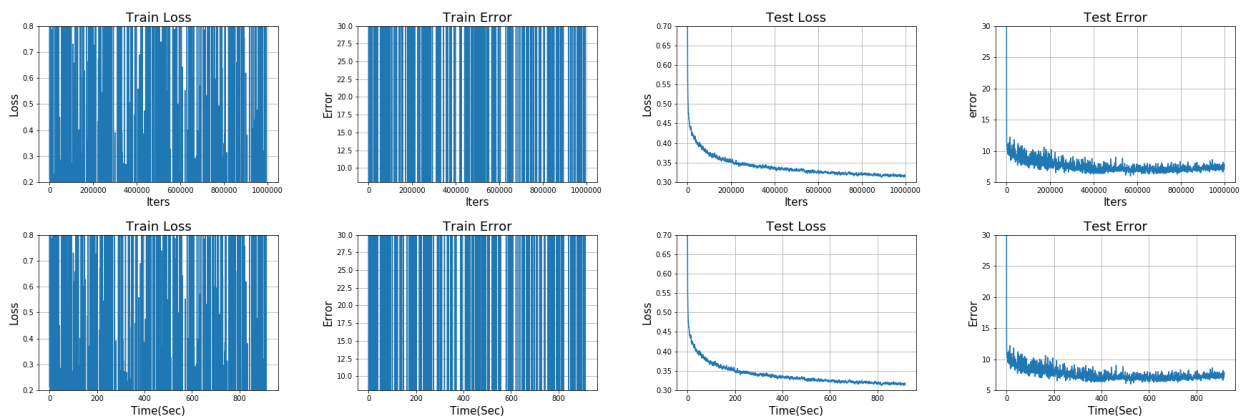# 2 SGD

## 2.1 batchsize = 1



Figure 2: SGD with batch size=1, learning rate=0.001
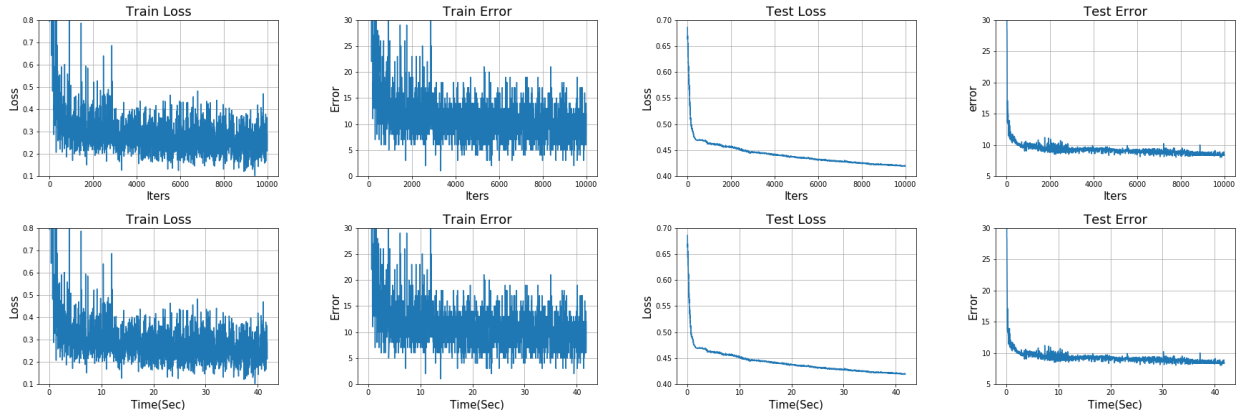
## 2.2 batchsize = 100



Figure 3: SGD with batch size=100, learning rate=0.005
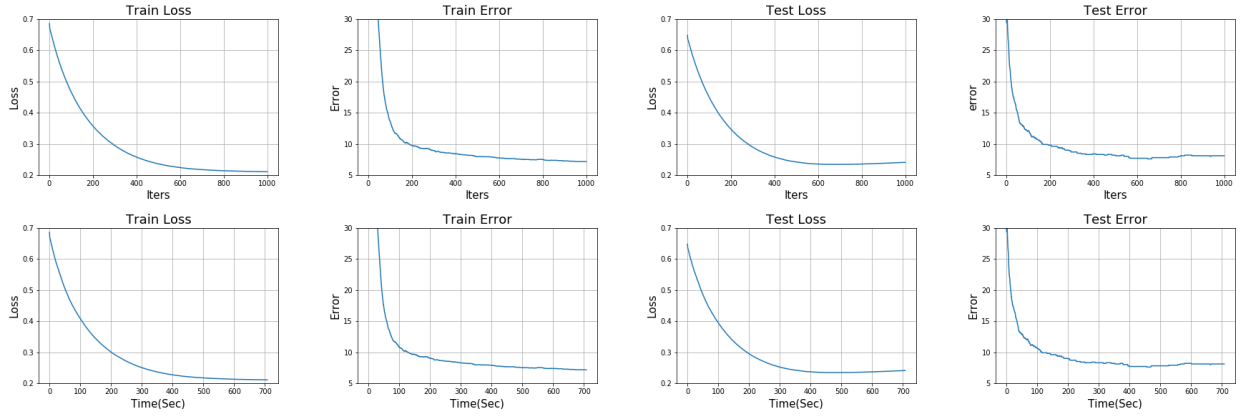
# 3 BFGS



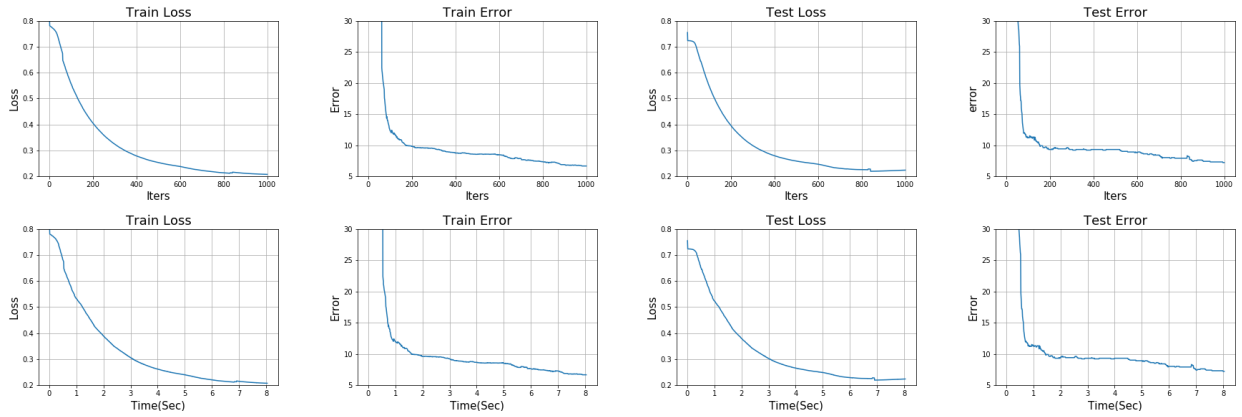Figure 4: BFGS learning rate=0.005

# 4 LBFGS



Figure 5: LBFGS learning rate=0.005, history size=10

# 5 Discussion

- GD is stable but go through the whole dataset each iteration is really time consuming.

- SGD converges faster timewise but unstable compared with GD. Batchsize=100 performs more stable than batchsize=1.

- As second-order algorithms, BFGS and LBFGS converges faster than first-order methods GD and SGD epochwise.

- LBFGS and BFGS have similar convergence performance epochwise. However, the time complexity of BFGS is $O(N^2)$ and time complexity of LBFGS is $O(mN)$, where $N$ is the number of observations and $m$ is the history size. Since $m << N$, LBFGS converges dramatically faster than BFGS in timewise perspective.

- The log(time) vs testaccu plot shown below. We can rank the performance of different algorithms as
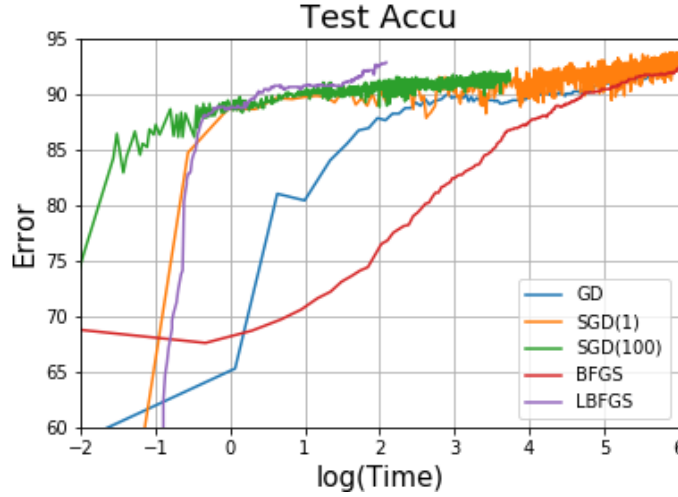
$$SGD(100) > SGD(1) \approx LBFGS > GD > BFGS$$



Figure 6: Log time and test accuracy comparison for different algorithms