

IST 597: Foundations of Deep Learning

Assignment #00100

Prabhav Borate

PSU ID: 990654897

Date: 03/02/2022

Time: 02:29:59 pm

Objectives

The objective of the assignment is to build a custom stochastic algorithm to update the model weights. Also, the custom optimizer will be compared with Keras in build optimizer (SGD, RMSprop, and ADAM) with performance across ten trials.

Custom Optimizer:

The custom stochastic optimizer is implemented as discussed in the HW. Following optimizer parameters are used.

- 1) learning rate: $1e-3$
- 2) Beta1: 0.9
- 3) Beta2: 0.999
- 4) Beta3: 0.999987
- 5) Epsilon: $10e-8$

The custom optimizer and Keras inbuild optimizers are tested across 10 trials.

MNIST Dataset:

The performance of the custom optimizer with MNIST dataset is as shown in figure 1. Where the base, L1 and L2 regularized models are compared for the mean and variance of the accuracy. Figure 2, 3, and 4 displays the performance of the Keras in build optimizer Adam, RMSProp and SGD respectively.

Fashion MNIST Dataset:

The performance of the custom optimizer with Fashion MNIST dataset is as shown in figure 5. Where the base, L1 and L2 regularized models are compared for the mean and variance of the accuracy. Figure 6, 7, and 8 displays the performance of the Keras in build optimizer Adam, RMSProp and SGD respectively.

Discussion

- 1) Time taken for training (sec): SGD(~ 300), Adam(~ 350), RMSprop(~ 400), Custom Optimizer(~ 800)
- 2) Adam: Works with the first and second momentum (takes best of RMSprop and SGD). In Adam the learning rate is adaptive & computation of mean and variance of gradients takes place. Faster computation & convergence is noticed and results shows that low variance is observed (better performance).
- 3) SGD: It is variant of a standard gradient descent optimizer. In this method the model parameters are updates for every training sample (the update is frequent). So the least convergence time is observed.

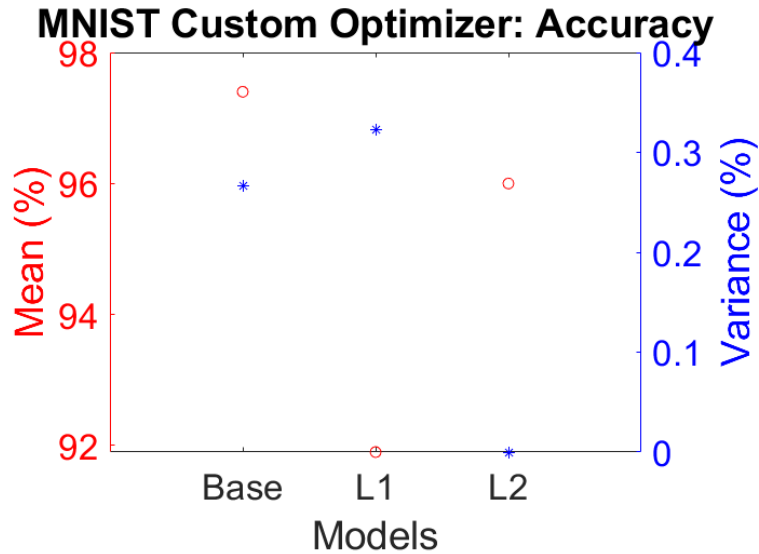


Figure 1: MNIST Custom Optimizer Model: With 10 Inference

4) RMSprop: The aim of the RMSprop (root mean square prop) as it's name suggest is to get moving average of the squared gradient of the weights. The observed convergence time with RMSprop method is close with the time needed for Adam method to converge.

5) Custom Optimizer: When compared with the inbuild keras optimizer discussed above the custom optimizer shows highest convergence time. This is due the presence of more model parameters compared to the inbuild optimizer. So the time needed for parameter update and overall training is increased compared to the inbuild optimizers.

GitHub Link

https://github.com/prabhavorate92/IST597_Spring_2022_Assignment_00100.git

The files are under the "master branch".

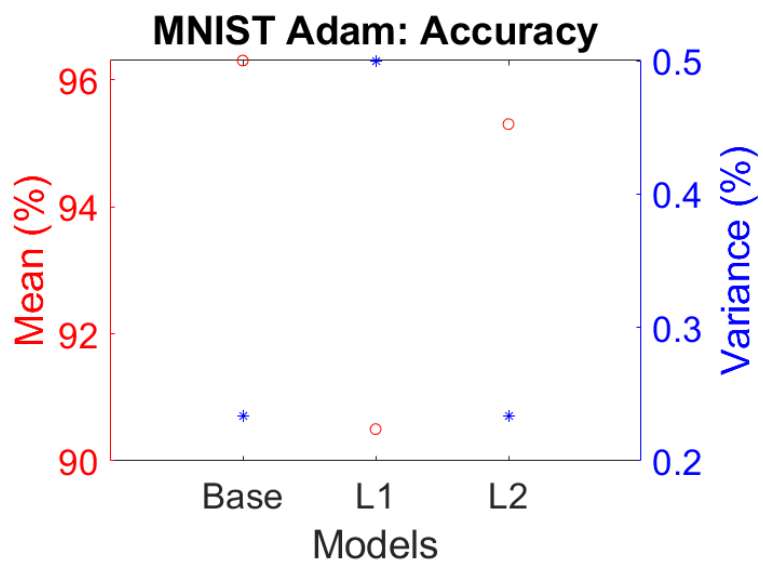


Figure 2: MNIST Adam Model: With 10 Inference

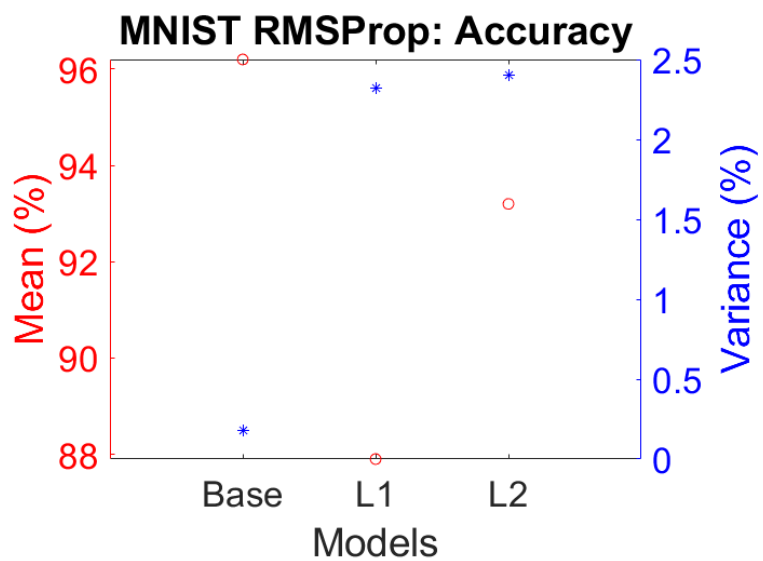


Figure 3: MNIST RMSProp Model: With 10 Inference

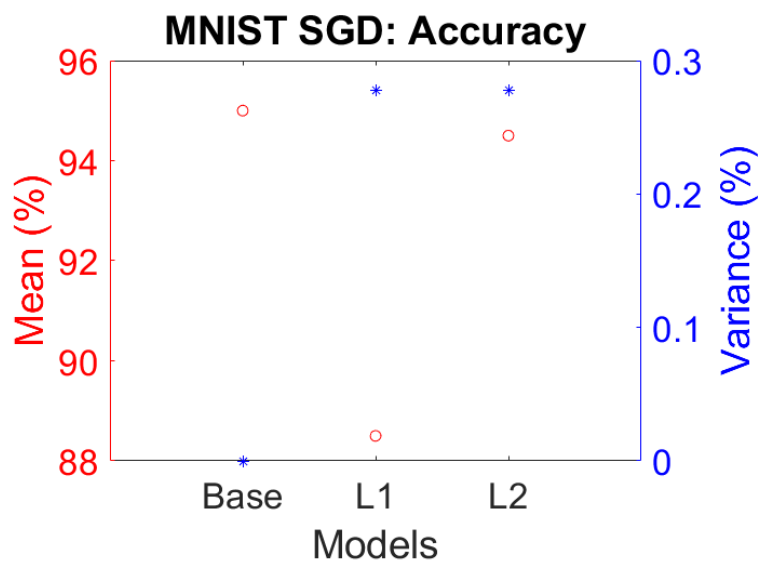


Figure 4: MNIST SGD Model: With 10 Inference

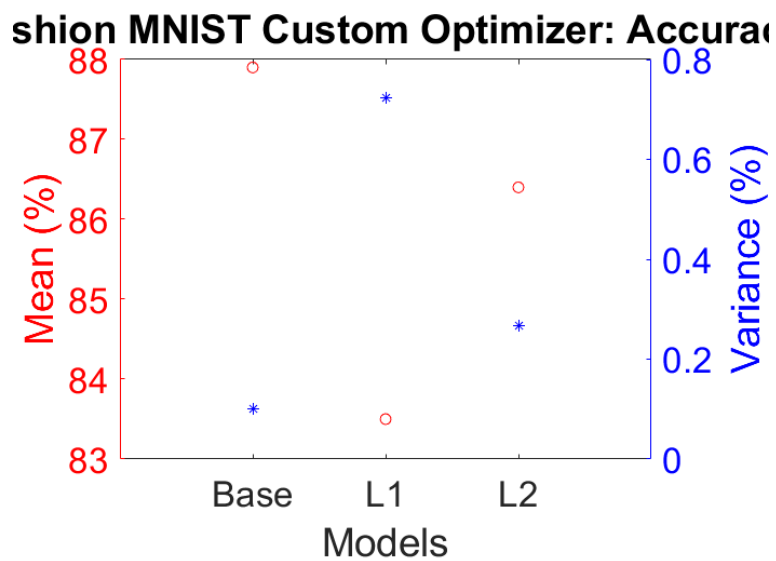


Figure 5: Fashion MNIST Custom Optimizer Model: With 10 Inference

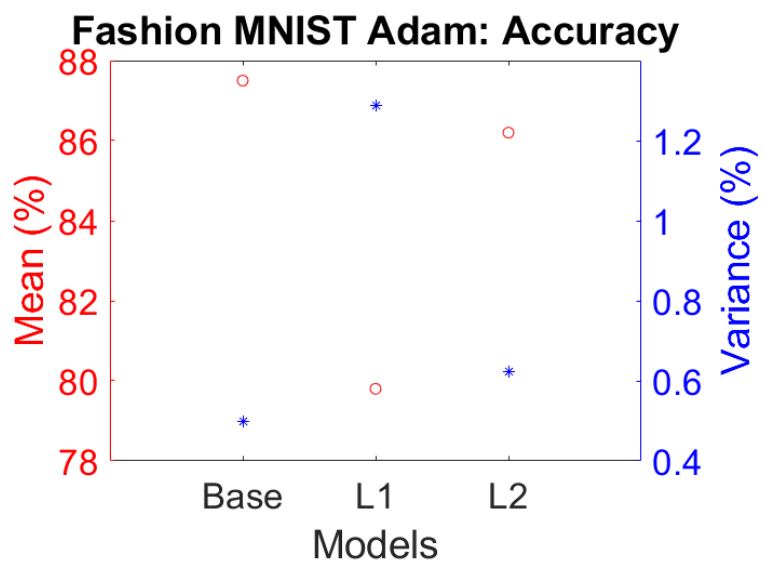


Figure 6: Fashion MNIST Adam Model: With 10 Inference

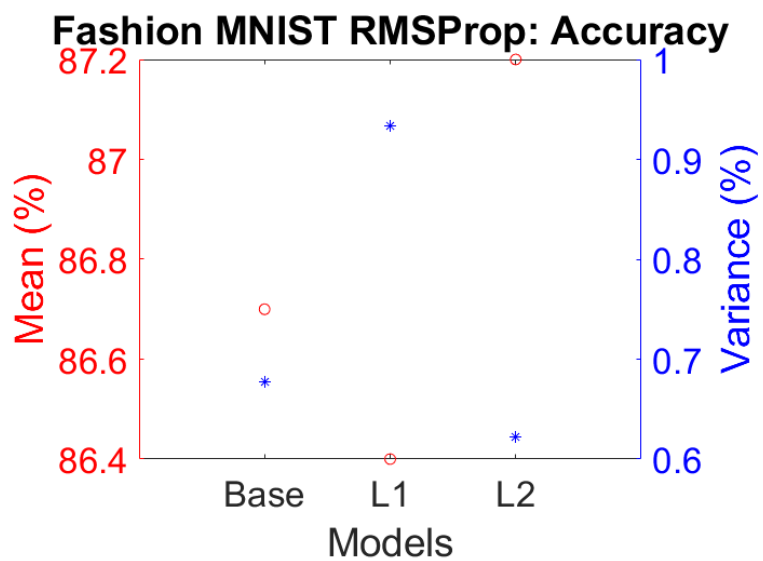


Figure 7: Fashion MNIST RMSProp Model: With 10 Inference

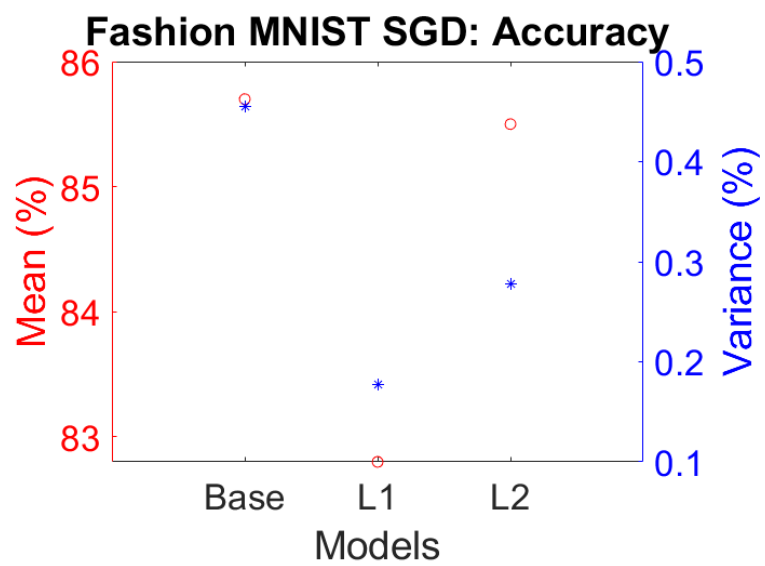


Figure 8: Fashion MNIST SGD Model: With 10 Inference