CS5260 – Spring 2022
Assignment 6

Shen Jiamin
A0209166A
shen_jiamin@u.nus.edu

Abstract

In this assignment, we will study the use of LR range test, using Colossalai framework. The chosen model and dataset are LeNet5 and MNIST. You will be asked to:

1. Choose one optimizer from SGD, ADAM, ADAMW, RADAM, LARS, LAMB or other optimizer you are interested in. (Note: AdaGrad is not supposed to work with this method).
2. Choose two learning rate scheduling method from Pytorch library. (Including no scheduling) Some possible choice: Multistep, OneCycle
3. Take use of `Colossalai_lr_range_test.ipynb` to conduct LR range test for optimizer you chose. Propose several learning rates for real training.
4. Write your code (either adapt provided code or start form scratch) to train LeNet5 on MNIST with two learning rate scheduling methods you choose and proposed learning rate. A suggested epoch number is 30.
5. Observe the result, and write a brief report about what you find (optimizer and scheduling method you choose, corresponding LR region on LR range test plot. docx, pdf, ipynb are all acceptable). Remember to save necessary images or data during experiments for report writing.
6. Upload your work including requirement.txt to your github. Add the github link to your report.

The code is hosted at https://github.com/LuminousXLB/CS5260-AY2122-2/tree/main/Asg06.

1 LR Range Test

I used my script to search for appropriate ranges of learning rate for SGD (Figure 1), Adam (Figure 2), AdamW (Figure 3) and RAdam (Figure 4) optimizer. The LR range I chose is as Table 1. That is chosen by approximating the descent rate of loss and filtering the range of learning rates, where the descent of loss is above a quantile.

From the result, we can find that the effective learning rate for the SGD optimizer is much larger than those for Adam-like optimizers, where the difference is several magnitudes.

Table 1: Chosen Learning Rate Range

Optimizer	Minimun LR	Maximun LR
SGD	1.466599e-02	1.235050e-01
Adam	2.408037e-05	5.063813e-04
AdamW	2.408037e-05	4.975099e-04
RAdam	4.092044e-05	7.126536e-04

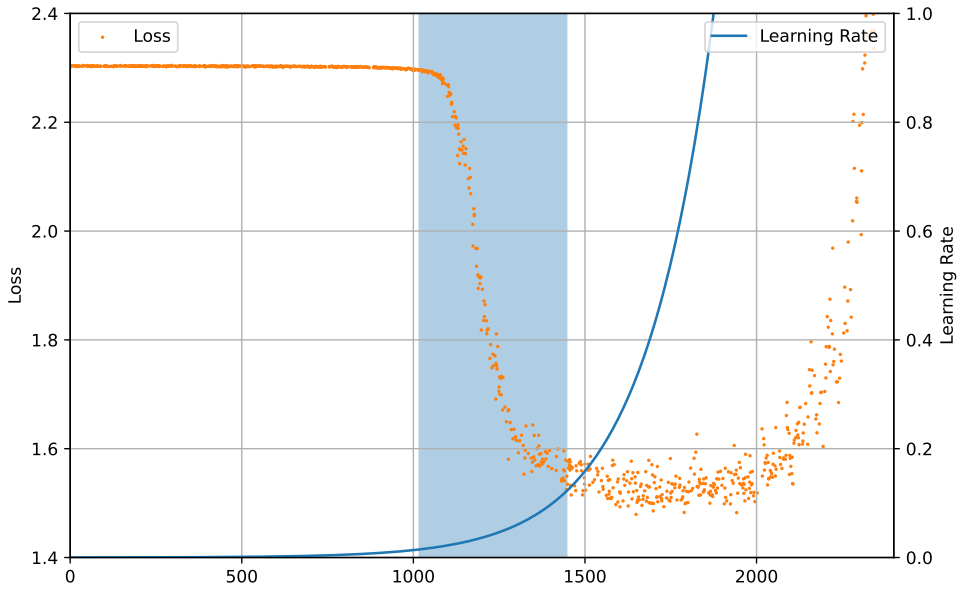

Figure 1: Learning rate with SGD optimizer

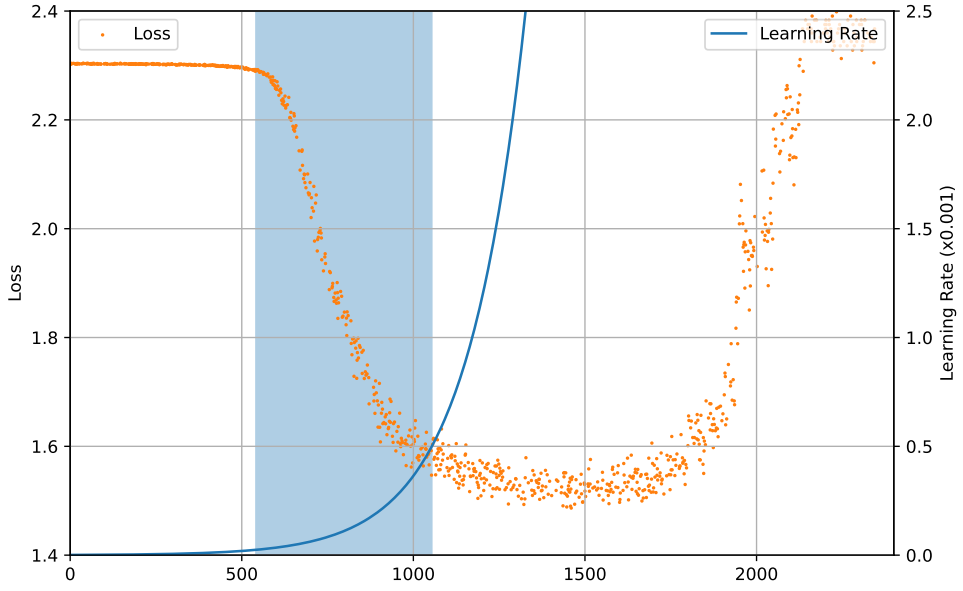

Figure 2: Learning rate with Adam optimizer

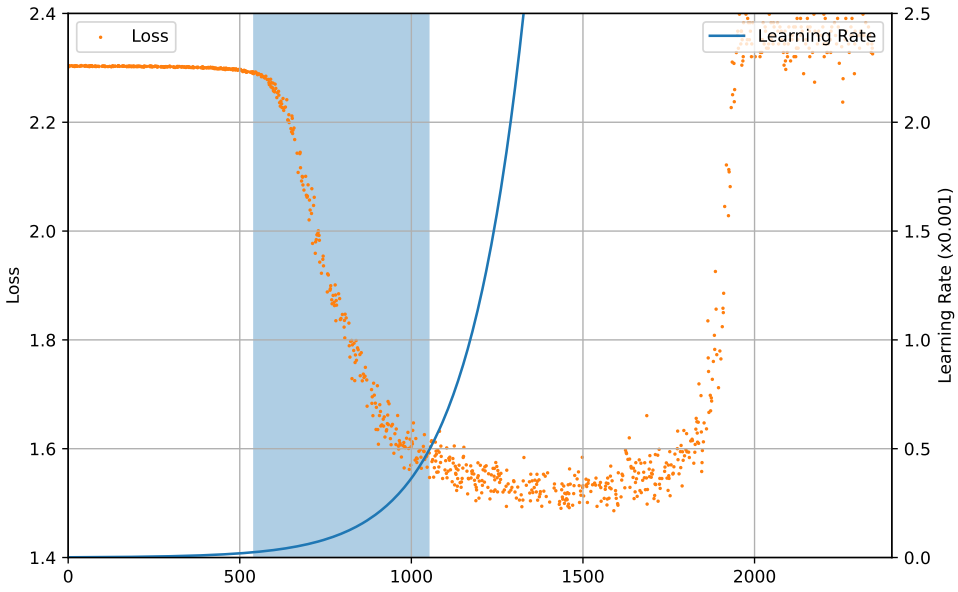

Figure 3: Learning rate with AdamW optimizer

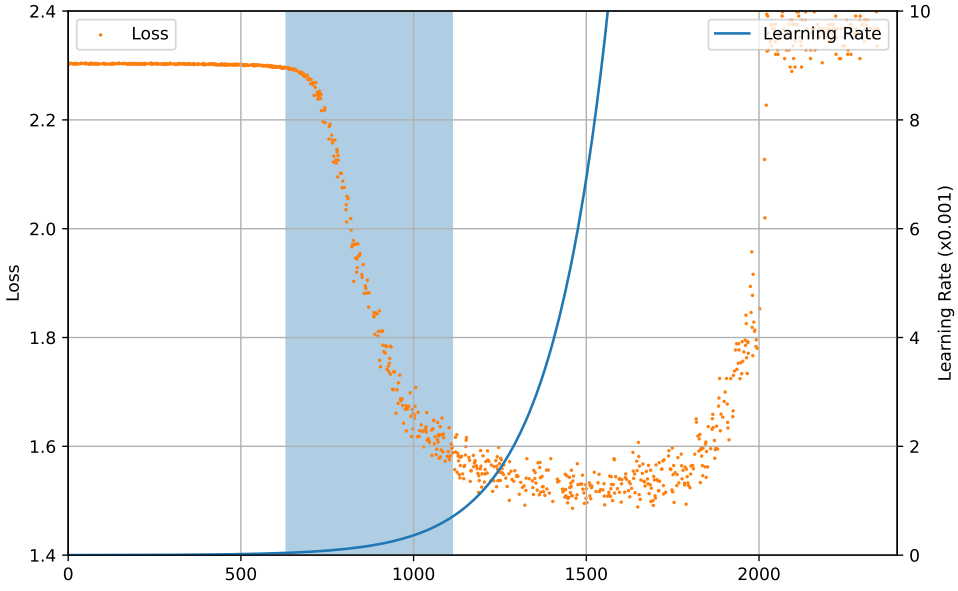

Figure 4: Learning rate with RAdam optimizer

2 LR Range Evaluation

The model is trained for five epochs using the LR range found above. The accuracy is plotted as Figure 5. Generally, a higher starting LR makes the accuracy grows faster in the first epoch. While the difference is not significant, OneCycleLR seems more appropriate for Adam-like optimizers than MultiStepLR.

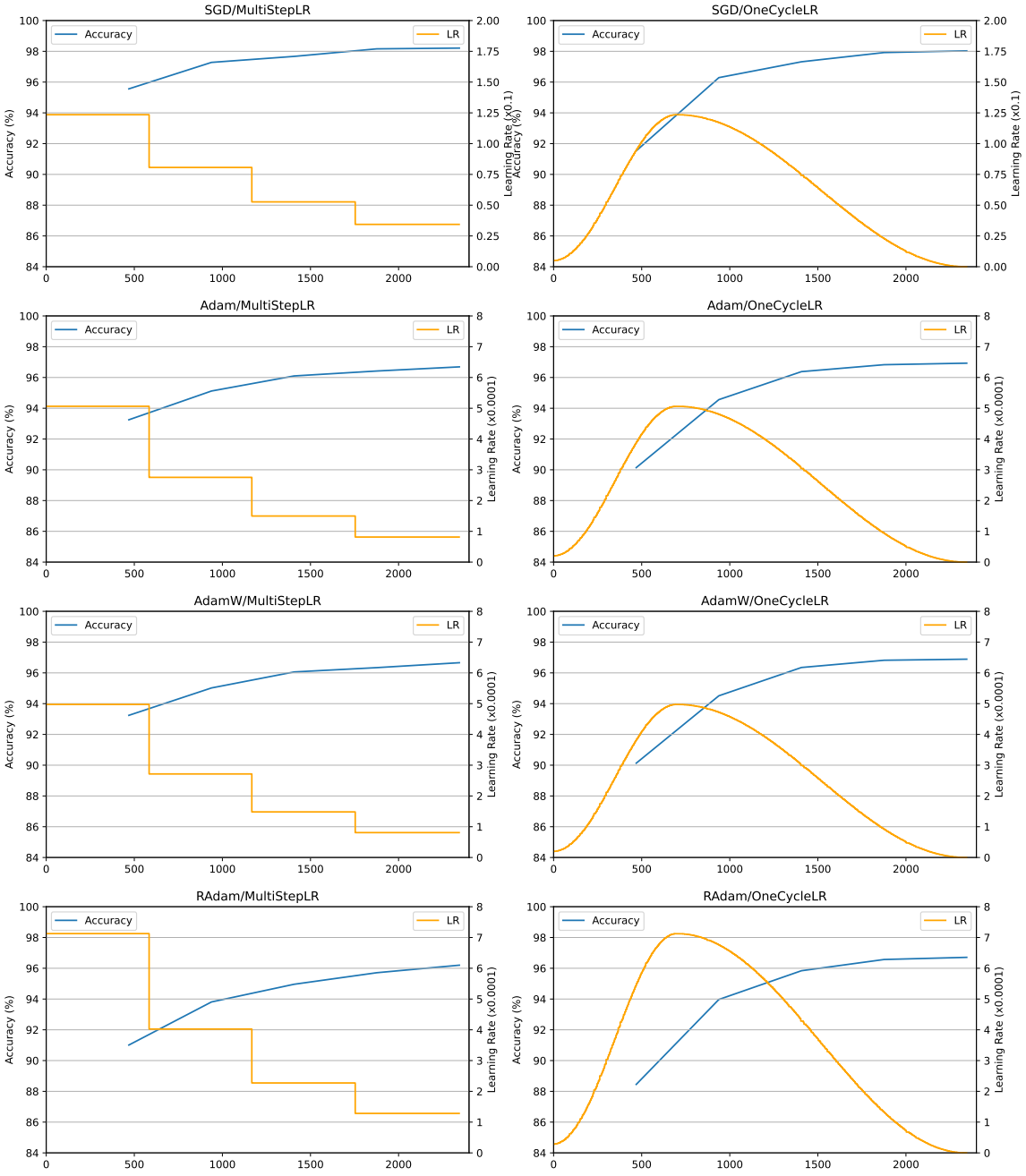

Figure 5: Training result with difference combinitions