

**# Brief report**

I choose SGD for optimizer. I choose ConstantLR and MultiStepLR (milestones are step 10 and step 20, decay rate is 0.1) for scheduler.

From the LR range test, we can observe from **\*\*Figure 1\*\*** that the usable LR range is approximately from **\*\*step 1000\*\*** to **\*\*step 2200\*\***. In this LR range, the training loss keeps dropping. After this LR range, the training loss tends to explode. So I use this LR range to propose a list of LRs that are used for the real training.

In the real training, I conducted the training procedures with the two selected schedulers and the proposed LRs. **\*\*Figure 2\*\*** is the training loss for ConstantLR scheduler with different LRs, while **\*\*Figure 3\*\*** is for MultiStepLR scheduler.

For ConstantLR scheduler, we can see that the larger LRs correspond to faster convergence. So these larger LRs are more suitable for this training configuration because they converge much faster. It is also notable that the largest **\*\*LR 4.281e-001\*\*** has a more unstable curve than **\*\*LR 1.952e-001\*\*** (need to turn off the irrelevant curves to see this), which indicates that the most suitable LR should be in the end of the dropping phase during LR range test, and it should not be too large to stabilize the training. Also, we can see that the small LRs still converge, although at a much slower pace. This is because the ConstantLR scheduler never changes the LR during training, so the optimizer can slowly find its way to the optimal point.

For MultiStepLR scheduler, the differences among the LRs are more evident. Still, we can see that larger LRs correspond to faster convergence. However, the MultiStepLR scheduler is more sensitive to the choice of initial LR. Especially, the training loss from the smallest initial LR does not converge. The reason might be that the MultiStepLR scheduler decays the LR every 10 steps, so the small LRs quickly become insufficient for the training to converge. In other words, large LRs are more favored in this training configuration. This implies that we need to be careful with the initial LR selection when using certain shedulers.

In summary, I first conducted the LR range test with exponentially growing LR to obtain a usable LR range for this task. Then I conducted the real training with the two selected schedulers and the proposed LRs. I observe that even for the same LR range found by LR range test, the schedulers behave very differently. ConstantLR scheduler favors large LR but becomes unstable when LR is too large. MultiStepLR scheduler favors large LR and the convergence rate is very sensitive to the initial LR selection. Nonetheless, it is clear that with the LR range test algorithm, we are able to quickly find some initial LRs that can converge much faster than others selections.