A0232295E Assignment6

Zheng Jingnan

April 2022

1 Report

I choose to use Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments.[1] First we try Adam with MultiStepLR. This scheduler decays the learning rate by gamma once the number of epoch reaches one of the milestones. For instance, when epoch number reaches 5, 10, 15, 20, 25 and 30, the learning rate equals to 1.5 times the original learning rate, starting from lr = 1e-5. We can choose learning rate 2e-5 with Adam optimier since the training and testing loss reduce rapidaly in the corresponding period. Second we try Adam with OneCycleLR. This scheduler sets the learning rate according to the 1cycle learning rate policy. The 1cycle policy anneals the learning rate from an initial learning rate to some maximum learning rate and then from that maximum learning rate to some minimum learning rate much lower than the initial learning rate. We set the maximum learning rate to 1e-4. Result is shown below. Learning rate 2e-5 is an appropriate parameter, according to the plot.

2 References

[1]Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

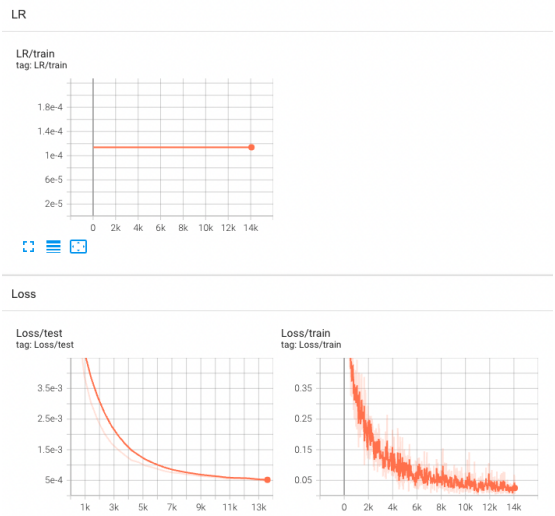

Figure 1: MultiStepLR Adam

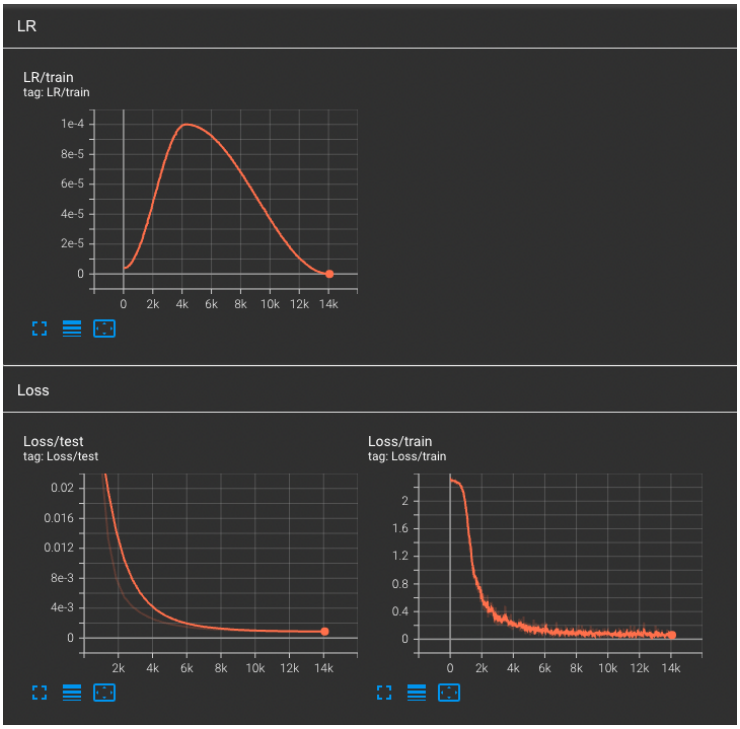

Figure 2: OneCycleLR Adam