Assignment 6 – Srinivasan Nandakumar A0228510N

Report:

Basic baseline(SGD)

Chart, line chart

Description automatically generated

The loss starts increasing at around the 2000th iteration and the corresponding learning rate is 0.12(approx). We use this learning rate for two schedulers MultistepLRand OneCycleLR.

Chart, line chart, box and whisker chart

Description automatically generated

Baseline with Adam Optimizer. The learning rate after which the training loss increases is around 11k steps. Adam optimizer shows greater fluctuation and higher errors as the learning rate increases when compared to the SGD optimizer.

MultistepLR:

Chart

Description automatically generated

With decaying learning rate at epochs 10 , 20 the train converges fast but is slightly slower when compared to OneCycleLR in the earlier stages of the training which can be explained due to the decaying learning rates only post 10 and 20th epoch.

OneCycleLR:

With the minimum learning rate at 0.1 and max learning rate 0.125 , we get quick convergence using the OneCycleLR scheduler.

Chart, line chart

Description automatically generated