Assignment #01000

Ruiqi Liu 949919908

03/19/2022 19:14PM

1   MLP for FMNIST

We test the performance of SGDoptimizer on FMNIST dataset. Here is the average result over three trials, the **PRE** indicated the batch normalization layer is applied after pre-activation while **POST** indicated the batch normalization layer is applied after post-activation. After

BN	Train ACC	Train Loss	Train StdErr	Train Var	Valid Acc	Test Acc	Test Loss	Time(s)
Pre	0.8990	0.0021	0.0012	0.076	0.8767	0.8725	0.0948	193.10
Post	0.8970	0.0021	0.0012	0.075	0.8669	0.8632	0.0950	227.16

comparison, we can find when you applied batch normalization after post-activation the performance is a bit worse that pre-activation.
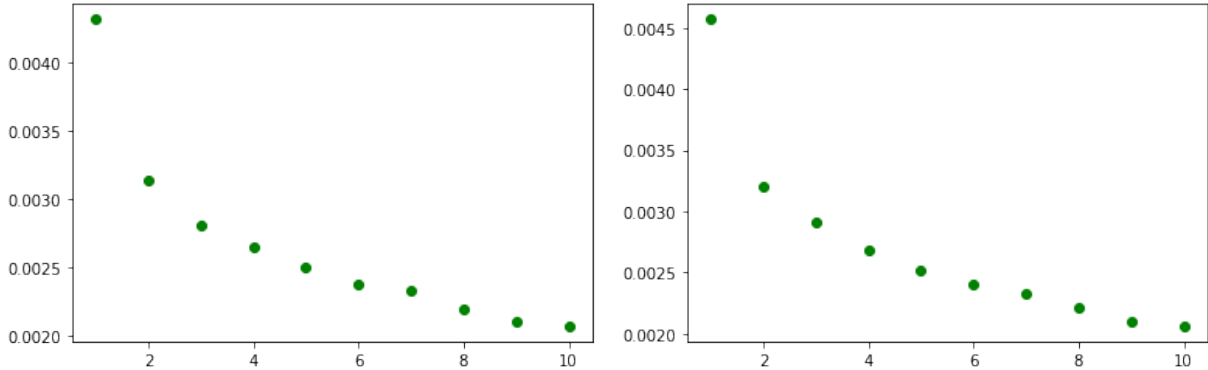

Figure 1: Left side is the Loss Curve for PRE; Right Side is the figure for POST.

1.1   Different Application for Training and Testing

We also use different method for batch normalization in training and validation/testing period. For training process, we just calculate the statistics for each mini-batch. But when comes to validation and testing, it will be a wasting of time to calculate each time, so we get statistic from the dataset and use it directly for the normalization in those steps.

2   CNN for FMNIST

We test the performance of Adam optimizer on FMNIST dataset with lr=0.01. Here is the average result over three trials, the **PRE** indicated the batch normalization layer is applied

BN	Train ACC	Train Loss	Valid Acc	Valid Loss	Test Acc	Test Loss	Time(s)
PRE	0.9115	0.1638	0.9034	0.2844	0.9011	0.3117	146.52
POST	0.9124	0.1787	0.9039	0.1786	0.9037	0.2834	208.76

after pre-activation while **POST** indicated the batch normalization layer is applied after post-activation. Although the pre-activation version have similar result with post ones, but it will always stop 1-2 epoch earlier than post one due to the application of early-stopping mechanism.