Assignment #00100

Ruiqi Liu 949919908

03/05/2022 19:14PM

1 MLP for MNIST

We test the performance of SGD, RMSProp, Adam optimizer and custom one on MNIST dataset. Due to the different mechanisms, we cannot choose a common learning rate for all optimizers, for example, SGD converge too slow at $lr = 1e - 4$, but such learning rate will already cause Adam not converge. After adjustment and choose 1e-4 as the learning rate for Adam, 1e-2 for SGD, 1e-5 for RMS and 1e-3 for our own optimizer. The difference in learning rate will affect the converge speed, but when the epoch is large enough, the performance and the running time of each epoch can be correctly evaluated. Here is the result After comparison,

Optimizer	Train ACC	Train Loss	Train StdErr	Train Var	Valid Acc	Test Acc	Test Loss	Time(s)
Custom	0.9771	0.0006	0.0013	0.0849	0.9675	0.9711	0.0275	390.14
SGD	0.9541	0.0013	0.0013	0.0806	0.9532	0.9486	0.0421	280.51
RMS	0.9442	0.0016	0.0013	0.0806	0.9473	0.9383	0.0509	370.47
Adam	0.9712	0.0009	0.0013	0.0860	0.9670	0.9653	0.0310	340.56

we can find that our optimizer has the highest accuracy, Adam ranks second while the SGD converge fastest and ours did not performs well on speed performance. We will focus more on the accuracy of model, so we choose Adam to be the comparison optimizer.

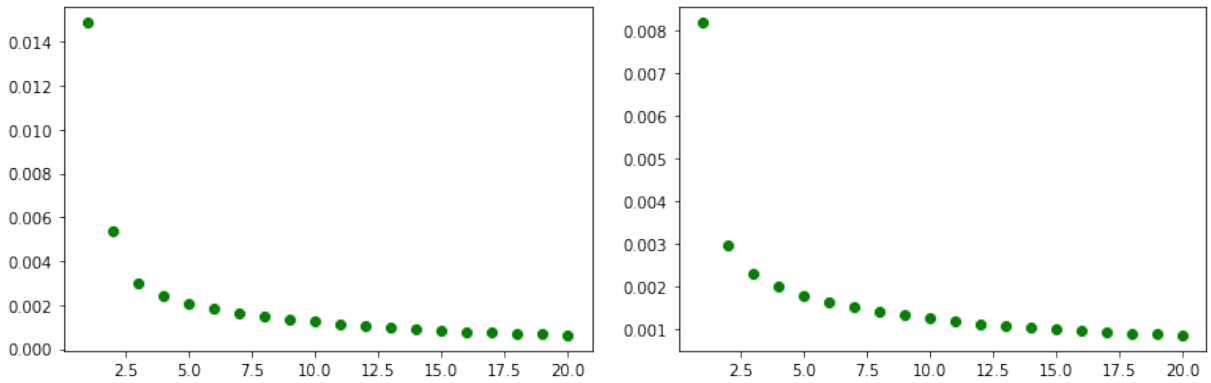

Figure 1: Left side is the Loss Curve for MNIST for our custom optimizer; Right Side is the figure for Adam.

2 MLP for FMNIST

Here is the optimizers comparison on FMNIST, We can see that our optimizer has more obviously better performance to others, and SGD still converge fastest.

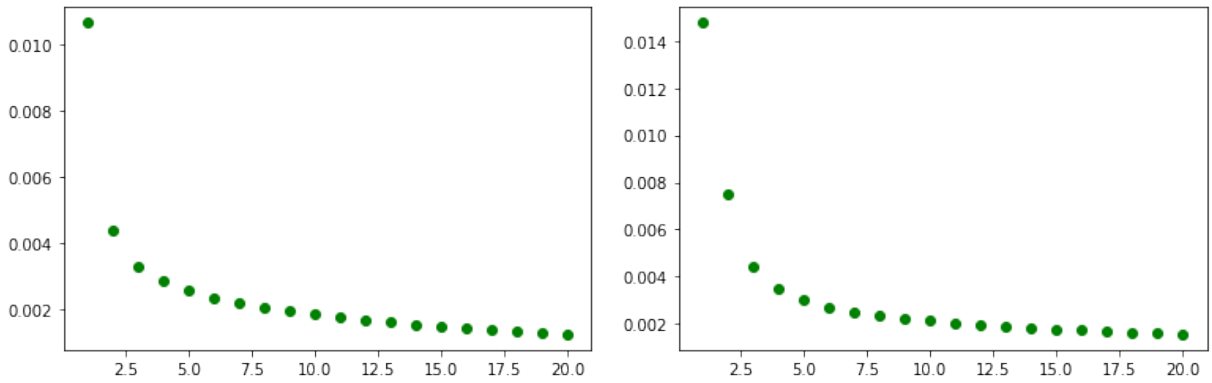

Figure 2: Left side is the Loss Curve for MNIST for our SGD optimizer; Right Side is the figure for RMS.

Optimizer	Train ACC	Train Loss	Train StdErr	Train Var	Valid Acc	Test Acc	Test Loss	Time(s)
Custom	0.8944	0.0024	0.0012	0.0733	0.8752	0.8706	0.0918	354.39
SGD	0.8648	0.0030	0.0012	0.0686	0.8586	0.8521	0.1058	196.76
RMS	0.8618	0.0031	0.0012	0.0688	0.8490	0.8470	0.1101	353.96
Adam	0.8636	0.0031	0.0012	0.0691	0.8538	0.8467	0.1092	363.26

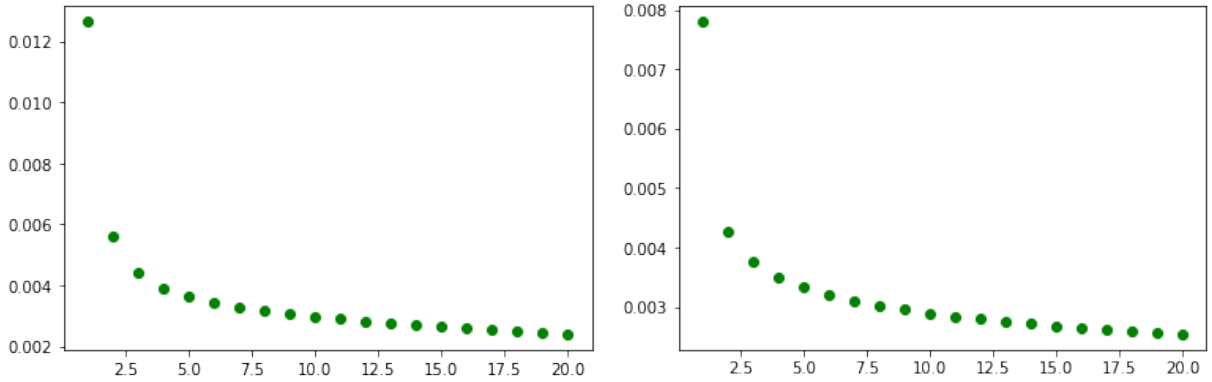

Figure 3: Left side is the Loss Curve for FMNIST for our custom optimizer; Right Side is the figure for Adam.

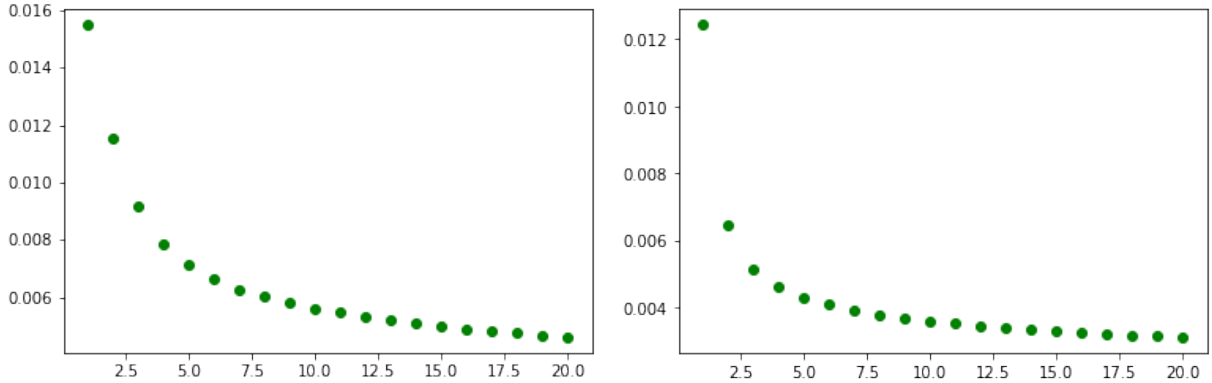

Figure 4: Left side is the Loss Curve for FMNIST for our SGD optimizer; Right Side is the figure for RMS.

3 Comparing in 10 Trials

Then we choose the optimizer which has the best performance among the Keras in-build optimizer, Adam, to compare with our own one. We run ten trials and get the average performance metric, the results are displayed below(data can be found in log folder),

Optimi- zer	Aveage Train Loss	Aveage Train Var	Aveage Test Acc	Aveage Test Loss	Aveage Time(s)
Custom(MNIST)	0.0006	0.0850	0.09675	0.0273	391.62
Adam(MNIST)	0.0009	0.0860	0.09622	0.0341	374.33
Custom(FMNIST)	0.0024	0.0730	0.8697	0.0934	340.75
Adam(FMNIST)	0.0026	0.0737	0.8651	0.0995	355.21

We can find that our optimizer has lower loss in both Train and Test phase, which means better robustness. Also the lower average variance means better stability. The time performance is similar with Adam and RMS but much lower than SGD.