IST_597 Deep Learning Assignment 00010 Report
tfv5097-Thejasvi Velaga

Inputs:

```
size_input = 784
size_hidden = [128,64]
size_output = 10
number_of_train_examples = 7000 #60000
number_of_test_examples = 1500   #10000
seed=5097
```

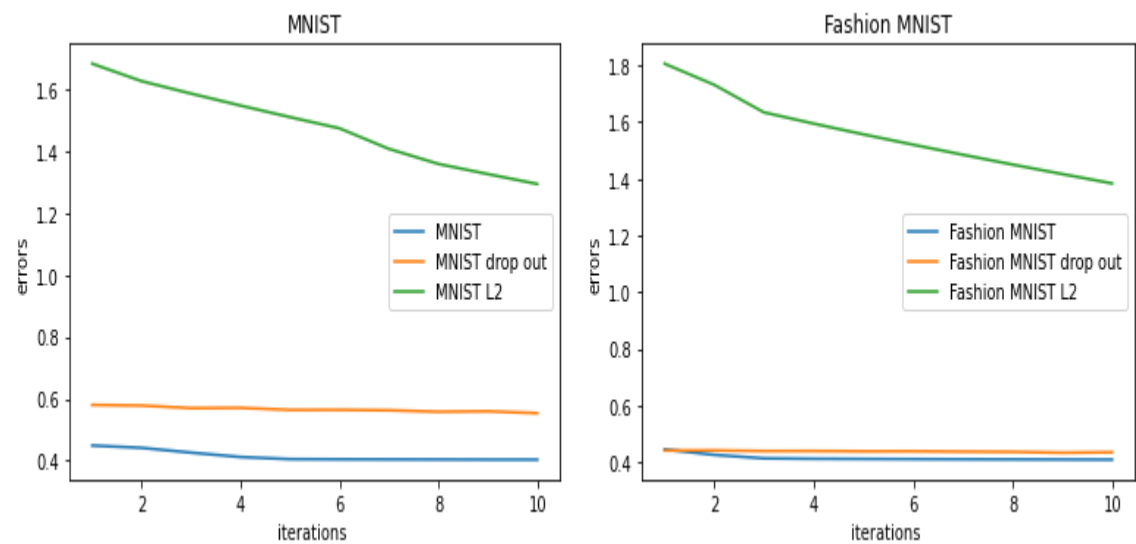| For GPU without Regularization					
MNIST			Fashion MNIST		
Epoch	Avg categorical_cross_entropy	accuracy	Epoch	Avg categorical_cross_entropy	accuracy
1	0.4494061802455357	99.91666666666667	1	0.4442196568080357	99.91666666666667
2	0.4418868582589286	100.0	2	0.4249189801897321	99.91666666666667
3	0.4264952218191964	99.95833333333333	3	0.4126382533482143	99.91666666666667
4	0.4123005022321429	99.91666666666667	4	0.4111673060825893	99.875
5	0.40541392299107143	99.83333333333333	5	0.410490234375	99.83333333333333
6	0.404906005859375	99.95833333333333	6	0.4099150390625	99.875
7	0.40455594308035714	99.91666666666667	7	0.40936146763392856	99.79166666666667
8	0.4042332589285714	99.875	8	0.4089474051339286	99.91666666666667
9	0.40388016183035713	99.75	9	0.4084927455357143	0.4084927455357143

10	0.403768310 546875	100.0	10	0.408492745 5357143	0.408492745 5357143
Total time taken (in seconds): 94.05			Total time taken (in seconds): 113.01		

For GPU without Drop Out					
MNIST			Fashion MNIST		
Epoch	Avg categorical_cross_entropy	accuracy	Epoch	Avg categorical_cross_entropy	accuracy
1	0.580371477 3995535	99.9375	1	0.440268310 546875	99.91666666 666667
2	0.578351248 6049107	99.8125	2	0.440652378 62723217	99.875
3	0.570089146 2053571	99.9375	3	0.438192313 0580357	99.91666666 666667
4	0.570954241 0714286	99.8125	4	0.438430350 1674107	99.875
5	0.564282191 6852678	99.8125	5	0.437439906 5290179	99.79166666 666667
6	0.564434500 5580357	99.875	6	0.437729422 4330357	99.875
7	0.562941824 7767857	99.75	7	0.436318743 0245536	99.875
8	0.558334437 7790178	99.9375	8	0.435490722 65625	99.79166666 666667
9	0.559834228 515625	99.875	9	0.432429966 51785715	99.875
10	0.553708844 8660714	99.75	10	0.434183314 73214283	99.83333333 333333
Total time taken (in seconds): 145.35			Total time taken (in seconds): 89.09		

For GPU with L2					
MNIST			Fashion MNIST		
Epoch	Avg categorical_cross_entropy	accuracy	Epoch	Avg categorical_cross_entropy	accuracy
1	1.6820062779017857	99.75	1	1.8063604910714286	99.875
2	1.6255678013392858	99.875	2	1.7316092354910715	1.7316092354910715
3	1.5853204520089286	99.83333333333333	3	1.6341934988839286	99.75
4	1.546716517857143	99.66666666666667	4	1.594478515625	99.75
5	1.5093159877232143	99.58333333333333	5	1.5565185546875	99.75
6	1.4731943359375	99.70833333333333	6	1.520040736607143	99.79166666666667
7	1.4067544642857144	99.70833333333333	7	1.4844732142857142	99.83333333333333
8	1.358173828125	99.70833333333333	8	1.4498067801339285	99.75
9	1.325374302455357	99.5	9	1.4162222377232143	99.75
10	1.2933842075892856	99.45833333333333	10	1.4162222377232143	99.70833333333333
Total time taken (in seconds): 97.17			Total time taken (in seconds): 84.02		

One Step Inference						
	Default Mode for MNIST			Default Mode for Fashion MNIST		
	Without Regularization	Drop out	L2	Without Regularization	Drop out	L2
categorical_cross_entropy	0.3997	0.4395	1.7845	0.4154	0.4330	1.8341
Accuracy	99.95833333333333	99.875	100.0	99.91666666666667	99.875	99.875

Graphs:

One Step Inference table

For GPU without Regularization					
MNIST			Fashion MNIST		
Seed	Avg categorical_cross_entropy	accuracy	Seed	Avg categorical_cross_entropy	accuracy
5097	0.3997	99.95833333333333	5097	0.4154	99.91666666666667
1234	0.5557	99.9375	1234	0.4532	99.95833333333333
1111	0.4258	99.95833333333333	1111	0.4313	99.95833333333333
1212	0.4430	99.95833333333333	1212	0.4525	99.875
98933	0.4166	99.91666666666667	98933	0.4499	99.875
737387532	0.4061	99.83333333333333	737387532	0.4533	99.95833333333333
5473	0.4479	100.0	5473	0.4533	99.91666666666667
900000	0.4556	99.91666666666667	900000	0.4544	99.79166666666667
65838	0.4574	99.83333333333333	65838	0.4499	99.79166666666667
757493755	0.4542	99.95833333333333	757493755	0.4545	99.875

For GPU with Drop Out					
MNIST			Fashion MNIST		
seed	Avg categorical_cross_entropy	accuracy	seed	Avg categorical_cross_entropy	accuracy
5097	0.4395	99.875	5097	0.4330	99.875

1234	0.5915	100.0	1234	0.4293	99.95833333 333333
1111	0.4343	99.91666666 666667	1111	0.4359	99.875
1212	0.4399	99.95833333 333333	1212	0.4384	99.83333333 333333
98933	0.4452	99.95833333 333333	98933	0.4466	99.83333333 333333
737387532	0.4316	99.95833333 333333	737387532	0.4267	99.875
5473	0.4442	100.0	5473	0.4381	99.875
900000	0.4437	99.95833333 333333	900000	0.4513	99.79166666 666667
65838	0.4484	99.91666666 666667	65838	0.4457	99.875
757493755	0.4522	99.875	757493755	0.4254	99.95833333 333333

For GPU Regularization L2					
MNIST			Fashion MNIST		
seed	Avg categorical_c ross_entropy	accuracy	seed	Avg categorical_c ross_entropy	accuracy
5097	1.7845	100.0	5097	1.8341	99.875
1234	2.4137	99.9375	1234	1.7933	99.91666666 666667
1111	1.8593	99.875	1111	1.8621	99.95833333 333333
1212	1.8010	99.91666666 666667	1212	1.8380	99.95833333 333333
98933	1.7937	99.91666666 666667	98933	1.7833	99.91666666 666667

737387532	1.7774	99.91666666 666667	737387532	1.8853	99.83333333 333333
5473	1.8094	99.875	5473	1.8211	100.0
900000	1.8224	99.83333333 333333	900000	1.8591	1.8591
65838	1.7984	99.95833333 333333	65838	1.7565	99.91666666 666667
757493755	1.8758	99.95833333 333333	757493755	1.8102	99.95833333 333333

Graphs for one step inference:

Fashion MNIST

— MLP **— MLP with drop out** **— MLP with L2**
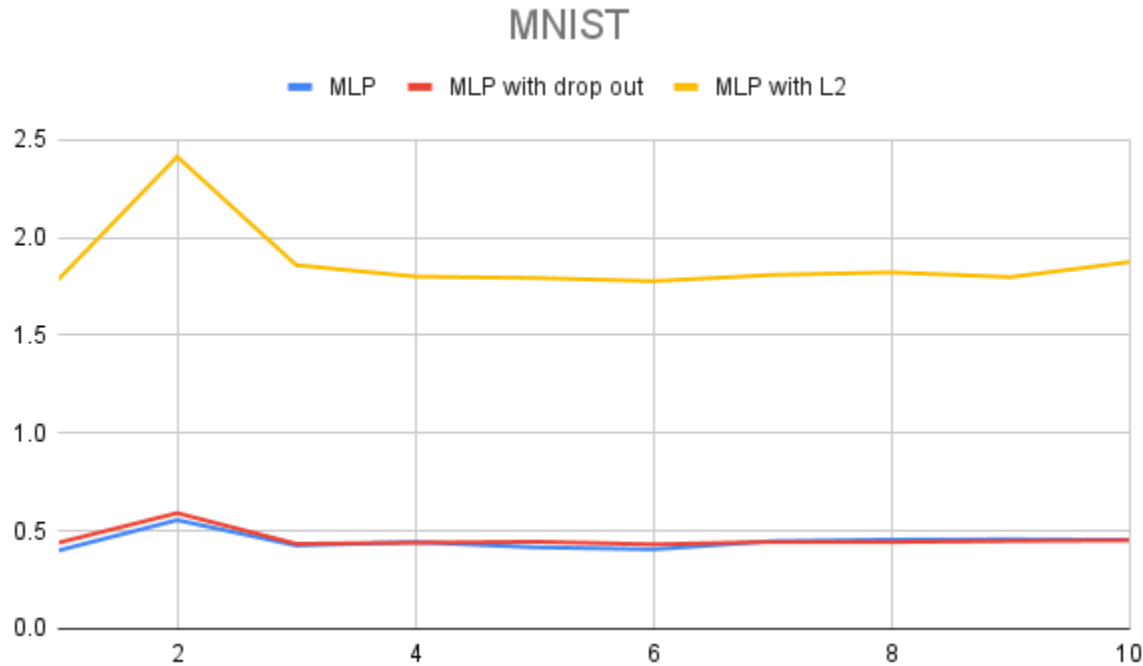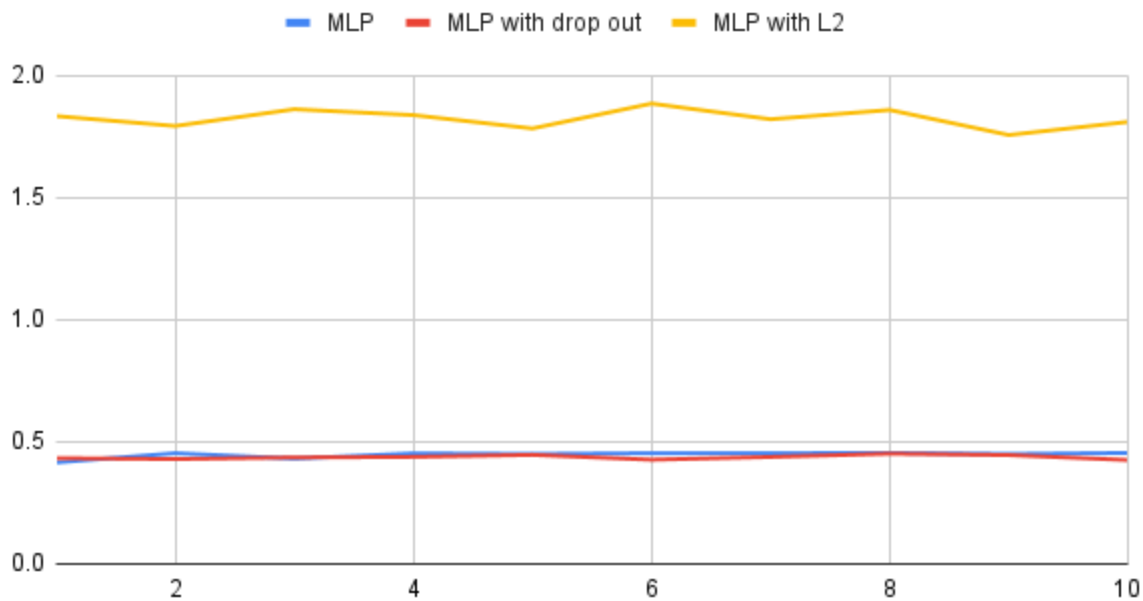
Findings:

When changing the hyper parameters from SGD to ADAM, I observed that ADAM has better performance when compared to SGD when we are considering categorical_crossentropy as a loss equation.

Instead of mean squares using categorical_crossentropy would help to increase the performance of the model as there are 10 categories of MNIST.

Adding the drop out and L2 regularization, I observed that there is drop when using L2 and boost in performance when using drop out.

As part of the hyper-parameter optimization, I changed the batch sizes to 24,32 etc and used a different learning rate for adam.