Deep Learning Assignment 2 Report

Aalo Majumdar
SR#: 16116
M.Tech Research, CSA, IISc

Task 1: Multilayer Perceptron Network

I have used a neural network with one hidden layer and one output layer. The number of hidden units in the hidden layer is 150 (as this was giving the most accuracy). The input image is first flattened into a 784 (28*28) sized vector and passed to the network.

Learning rate = 0.01
Optimiser used: SGD
L2 regulariser was used with weight decay = 1e-5

The network was trained for 100 epochs with a batch size of 100.
Training took 624.7 sec.

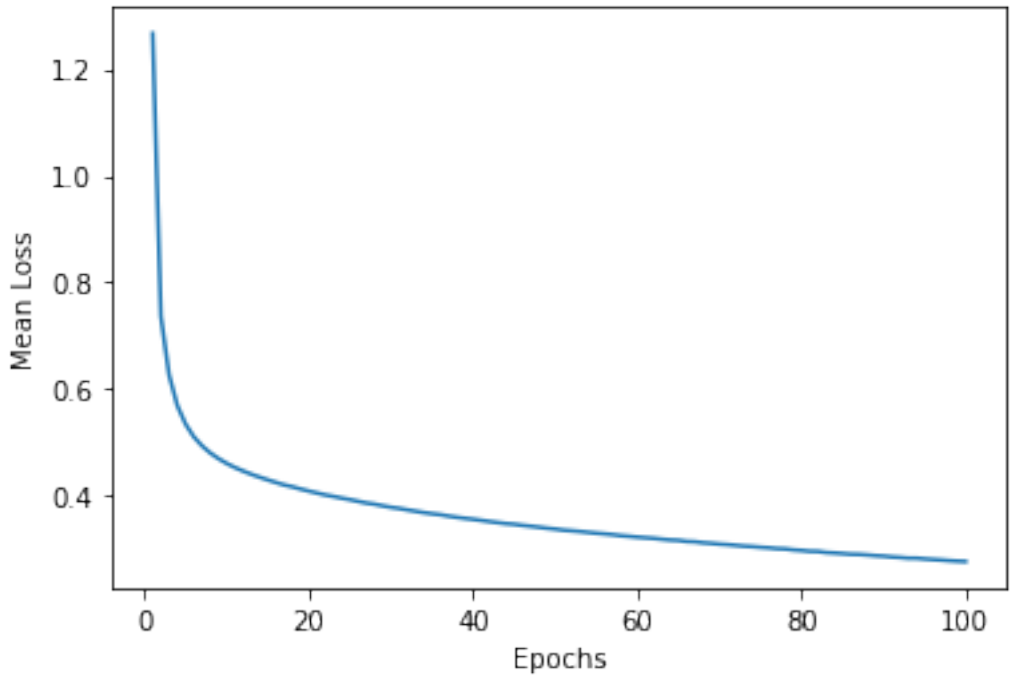

Illustration 1: Mean loss (averaged over all batches) vs No. of Epochs

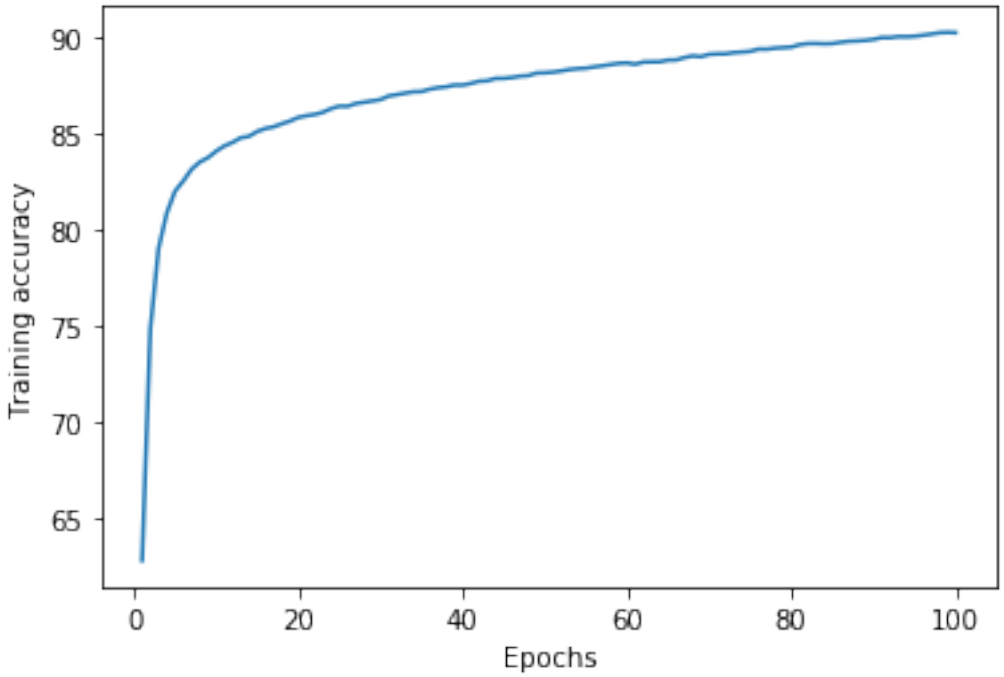

Illustration 2: Training Accuracy vs No. of Epochs

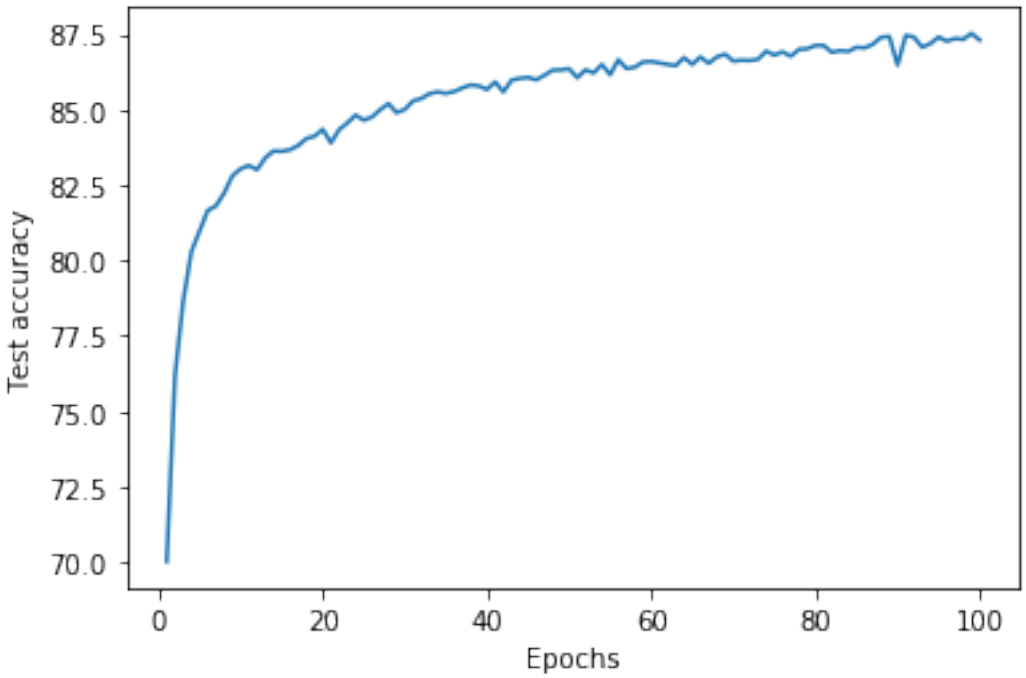

Illustration 3: Test accuracy vs No. of Epochs

Task 2: CNN

Two models were trained.

Model 1:

1. Input image of size 28*28 is passed to the 1st convolutional layer, with 1 in-channel, 6 out-channels, a kernel size of 5, stride of 1 and a padding of 1.
2. Max pool layer of kernel size 2 and stride 2
3. 2nd convolutional layer with 6 in-channels, 12 out-channels, kernel size of 5 and padding of 1.
4. Max pool layer of kernel size 2 and stride 3
5. A fully connected layer with 120 units.
6. 2nd fully connected layer with 60 output features.
7. Output layer has 10 output features.

Model 2:

1. Input image of size 28*28 is passed to the 1st convolutional layer, with 1 in-channel, 6 out-channels, a kernel size of 5.
2. Max pool layer of kernel size 2 and stride 2
3. 2nd convolutional layer with 6 in-channels, 12 out-channels, kernel size of 5.
4. Max pool layer of kernel size 2 and stride 2
5. A fully connected layer with 120 units.
6. 2nd fully connected layer with 60 output features.
7. Output layer has 10 output features.

Hyperparameters in both cases:
Learning rate = 0.01
Optimiser used: SGD
L2 regulariser was used with weight decay = 1e-5

The network was trained for 100 epochs with a batch size of 100.
Training time for model 1: 2348.7 sec
Training time for model 2: 1922.6 sec

Model 1 performed better on test data. Hence, model 1 was chosen.

	Train Accuracy	Test Accuracy
Model 1	92.5833	89.26
Model 2	92.4933	88.64

Comparing performance of Model 1 using SGD and Adam optimiser. Model 1 performs better with SGD optimiser

Optimiser	Train Accuracy	Test Accuracy
SGD	92.5833	89.26
Adam	91.3467	87.95

Graphs for CNN model 1

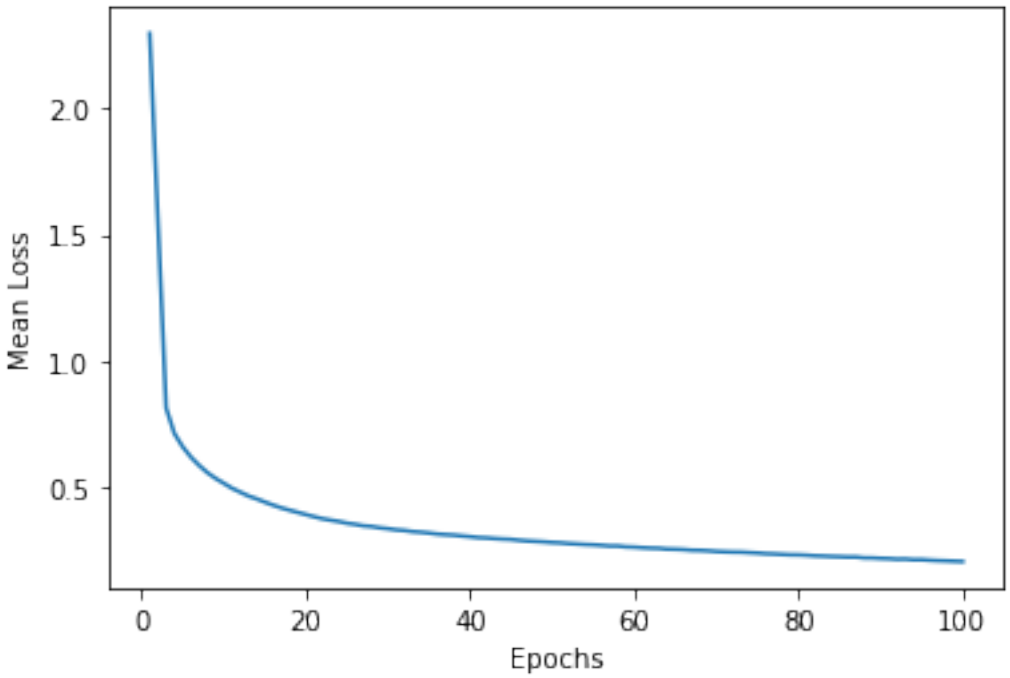

Illustration 4: Mean loss (averaged over all batches) vs No. of Epochs

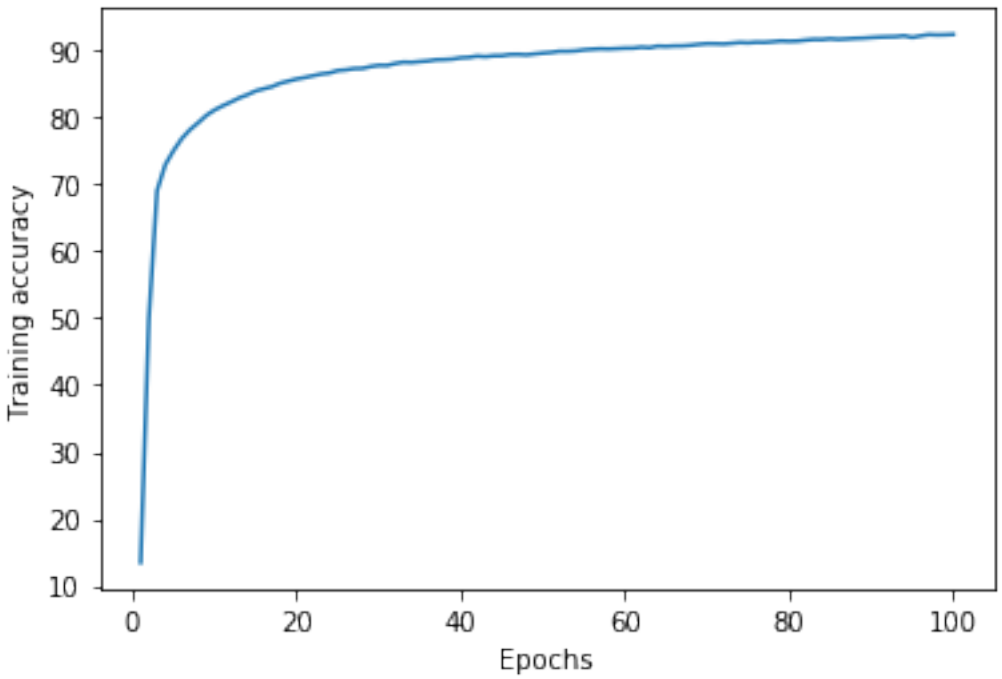

Illustration 5: Training Accuracy vs No. of Epochs

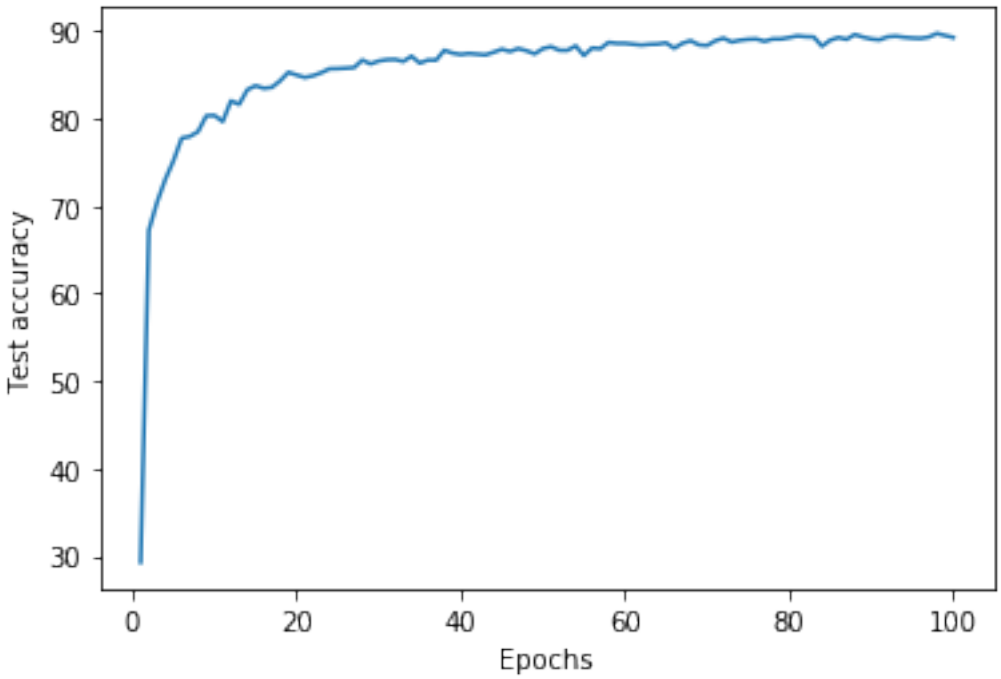

Illustration 6: Test accuracy vs No. of Epochs

Comparing the 2 models of CNN with neural network

	Train Accuracy	Test Accuracy
Model 1	92.5833	89.26
Model 2	92.4933	88.64
NN	90.4967	87.31

Confusion Matrices

CNN

884	3	16	17	5	1	61	0	13	0
5	972	1	16	2	0	1	0	3	0
15	0	841	13	77	0	51	0	3	0
30	2	5	901	38	0	21	0	3	0
3	0	102	32	826	0	36	0	1	0
0	0	0	0	0	966	1	23	4	6
217	0	111	24	105	0	529	0	14	0
0	0	0	0	0	6	0	963	2	29
3	2	6	4	4	2	10	4	965	0
1	0	0	0	0	7	0	42	2	948

Neural Network

862	0	10	22	6	2	92	0	6	0
5	968	0	20	2	0	4	0	1	0
23	0	749	15	129	0	84	0	0	0
22	9	11	880	41	0	32	0	5	0
1	2	58	25	839	0	72	0	3	0
0	0	0	1	0	949	0	35	2	13
131	2	62	23	65	0	710	0	7	0
0	0	0	0	0	20	0	967	0	13
7	2	4	9	6	3	9	7	953	0
1	0	0	0	0	9	0	56	0	934