COMPARISON OF THE PERFORMANCE OF TWO MULTI-LAYER PERCEPTRONS IN MNIST DATASET

Submitted by,

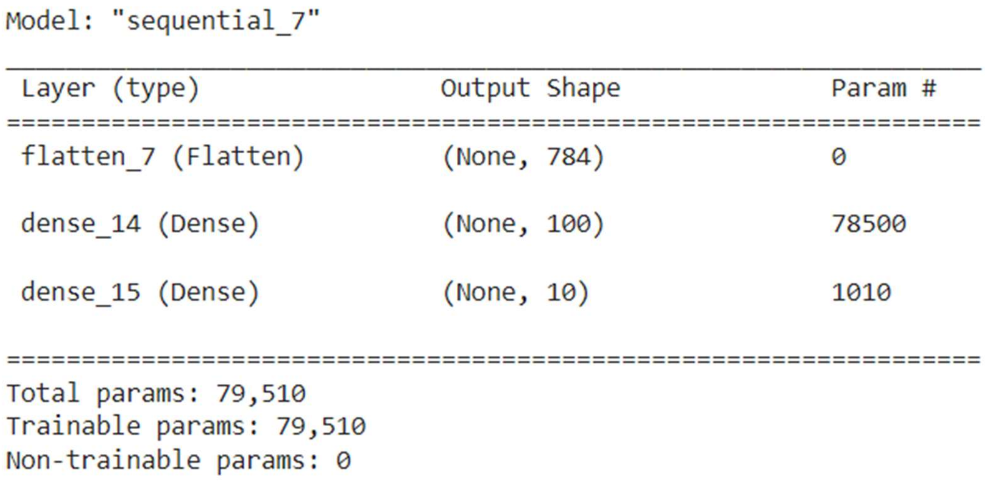
Thomas Ajai
S5 CSE
Roll no: 62
TVE19CS064

Source code

The Source code used to train the model can be found here(colab link) and github link

I trained the MNIST dataset in two models. The model summary of the two models is as shown in the images below

Model 1



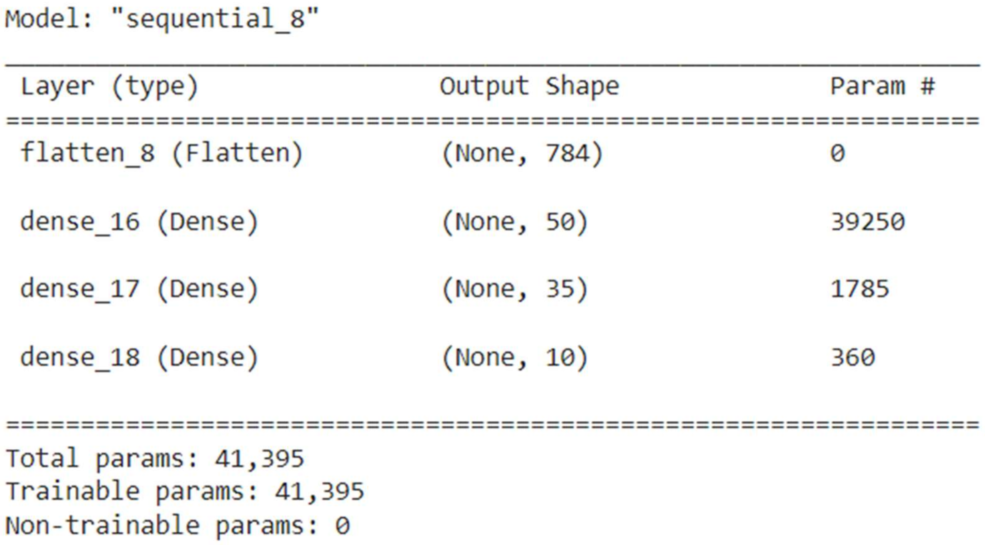
```
Model: "sequential_7"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 flatten_7 (Flatten)         (None, 784)               0

 dense_14 (Dense)            (None, 100)               78500

 dense_15 (Dense)            (None, 10)                1010

=================================================================
Total params: 79,510
Trainable params: 79,510
Non-trainable params: 0
```

Model 2



```
Model: "sequential_8"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 flatten_8 (Flatten)         (None, 784)               0

 dense_16 (Dense)            (None, 50)                39250

 dense_17 (Dense)            (None, 35)                1785

 dense_18 (Dense)            (None, 10)                360

=================================================================
Total params: 41,395
Trainable params: 41,395
Non-trainable params: 0
```

A comparison on both the model training is tabulated below:

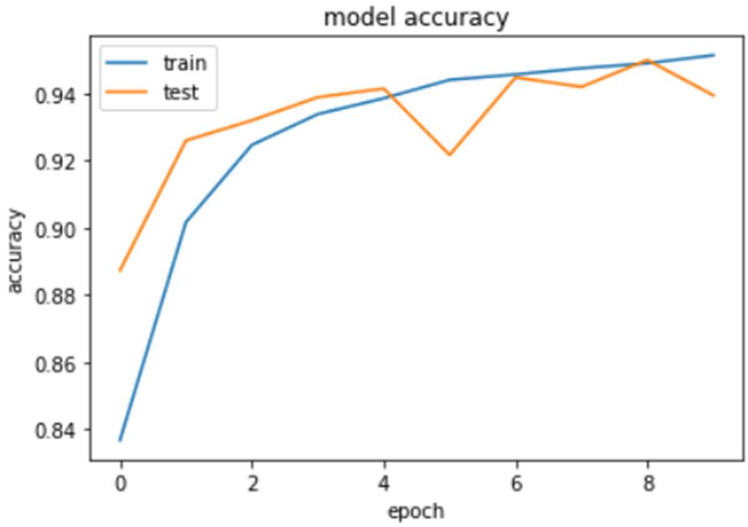
Criteria	Model 1	Model 2
# of Hidden layers	1	2
# of parameters	63610	42,310
Total # of neurons	80	110

Both the models were trained with same Loss function (cross entropy), same optimization function (Adam) and the same number of epochs(ten).

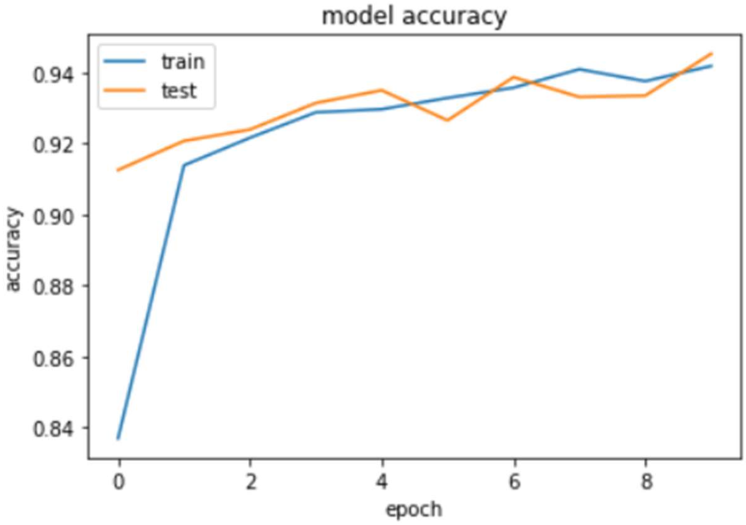
Validation Accuracy of Model 1 = 93.95%
Validation Accuracy of Model 2 =94.54%

Accuracy vs Epoch Plots

Model 1



Model 2



Inference

The 2nd Model is more accurate than the 1st model.
Both the models are trained though 10 epochs.
The 1st model has more parameters than the 2nd model.
But the second model has more hidden layers and so it is able to give better predictions with better accuracy.
The 1st model gets overfitted towards the end of the 10th epoch and tends to lose its generalization ability.