Computer
SCIENCE

College of Engineering, Trivandrum
NNDL - ASSIGNMENT 1
20 February 2022

Divina Josy

Roll No. 28

TVE19CS028

Semester: 5

1   Source Code

The source code can be found in the link below:

https://github.com/diina7/Divina-Josy_NNDL1.git (Github repository link)

2   Project

We use the MNIST Handwritten Data Dataset as our training and test datasets.
Here are some samples:

Two models were trained on the MNIST dataset.

The model summary of the two models is as shown in the images below:

MODEL 1

MODEL 2

```
Model: "sequential_2"
_____
 Layer (type)                   Output Shape                 Param #
=================================================================
 flatten_2 (Flatten)            (None, 784)                  0

 dense_4 (Dense)                (None, 128)                  100480

 dense_5 (Dense)                (None, 89)                   11481

 dense_6 (Dense)                (None, 50)                   4500

 dense_7 (Dense)                (None, 10)                   510


=================================================================
Total params: 116,971
Trainable params: 116,971
Non-trainable params: 0
_____
```

A comparison on both the model training is tabulated below:

Criteria	Model 1	Model 2
No. of hidden layers	101,770	116,971
Total number of layers	138	277

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 = 95.45%
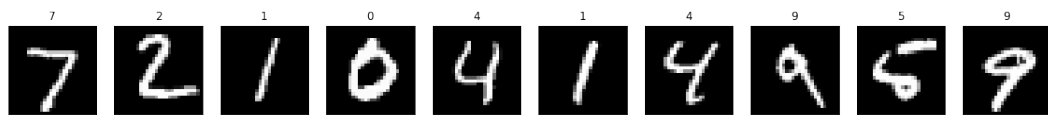Validation Accuracy of Model 2 = 98.22%

2.1   Predictions

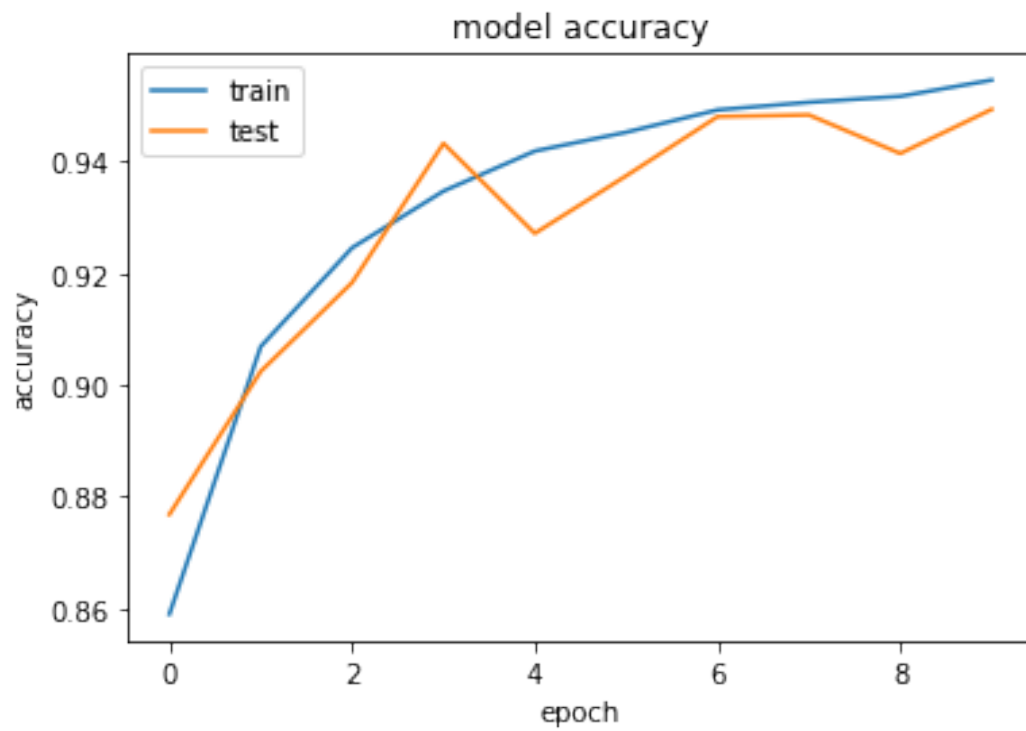Predictions done using these models are as follows:
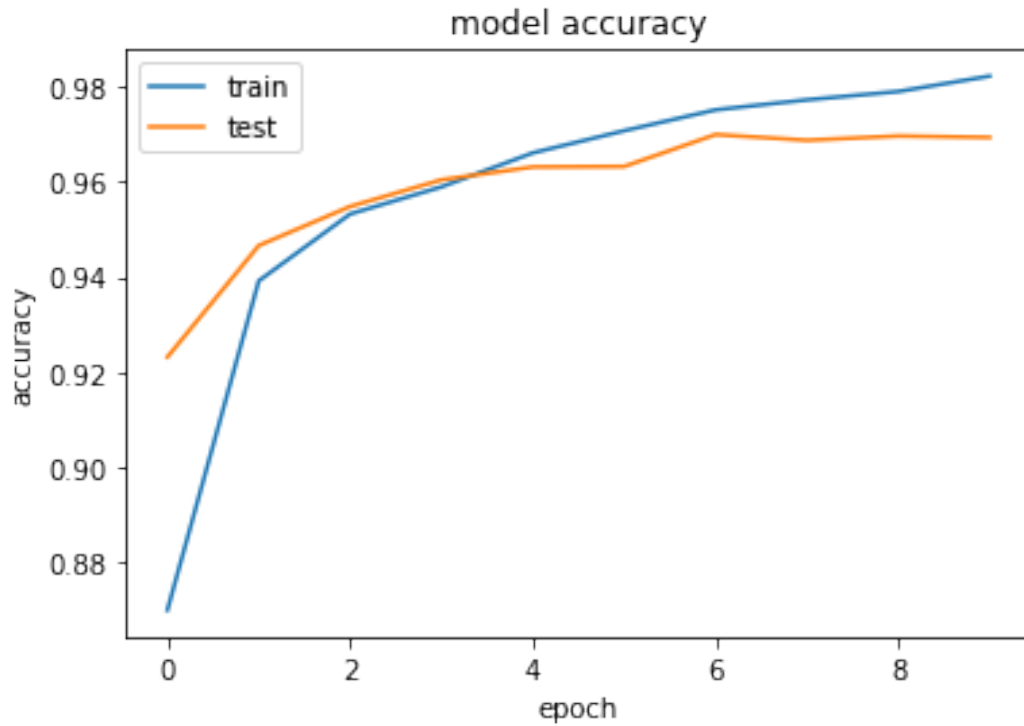
MODEL 1

MODEL 2

2.2   Accuracy vs Epoch Plots

MODEL 1

MODEL 2

2.3 Inference

Clearly the second model is a more complex model compared to the first one as it contains more hidden layers, more neurons, and hence more trainable parameters.

From the plots, we can infer that the second model is getting overfitted towards the end epochs. Thus Model 1 shows slightly less accuracy than Model 2.

The second model is getting overfitted because it is even learning the noise in the training samples and is losing its generalization ability.

It is also observed that there was a slight difference in the time required to train these models.

Model 1 took around 5 seconds for each epoch, whereas Model 2 took around 6 seconds for the same. This can be explained by the more amount of trainable parameters in Model 2 as compared to Model 1.