Project Part 3: Classification Using Neural Networks and Deep Learning

1. Introduction

The given task is to classify handwritten digits from the MNIST dataset using Convolutional Neural Networks CNN. CNN is used to classify the images, in this case digits, using the following parameter settings:

- Input layer- Handwritten digit image of dimension 28x28 pixels.

- Hidden Layer Description-
  - Kernel Size- The project focuses on two different variations of kernel size a) 3x3 and b) 5x5.
  - Number of feature maps- The project considers two different cases- first the number of feature maps for this layer is given 6. The second scenario is when a different value for number of feature maps is required and the value 20 is chosen.
  - Stride-1
- Pooling Layer-
  - Size- 2x2
  - Stride- 1
- Second Convolutional Layer-

  - Kernel Size- The project focuses on two different variations of kernel size for this layer a) 3x3 and b) 5x5.
  - Number of feature maps- The project considers two different cases- first the number of feature maps for this layer is given 16. The second scenario is when a different value for number of feature maps is required and the value 20 is chosen.
  - Stride-1
- Second Pooling Layer- The second convolutional layer is followed by second pooling layer.
  - Size- 2x2
  - Stride- 1
- Fully Connected hidden layer
  - Number of Nodes- 120
  - Activation- Rectified Linear Unit (ReLu)
- Fully Connected hidden layer
  - Number of Nodes- 84
  - Activation- Rectified Linear Unit (ReLu)
- Output Layer- SoftMax activation unit with 10 nodes representing each output.

Goal- To classify MNIST data using above described CNN, plot the training error and the testing error as a function of the learning epochs and report the final classification accuracy.

2. **Plots for the learning/testing errors and the final classification accuracy of baseline code.**

The CNN model described above is run on the baseline code, that is, kernel size 3x3 and first convolutional layer has 6 feature maps and second convolutional layer has 16 feature maps. The observed overall loss and accuracy is summarized in the given table:

Table 1: Testing error and Final Classification Accuracy for the baseline code

Parameter	Observation
Testing Error	61.79706339826807%
Final Classification Accuracy	97.96%

Plots for the training error and testing error as a function of the learning epochs for the baseline code is illustrated by Fig 1.

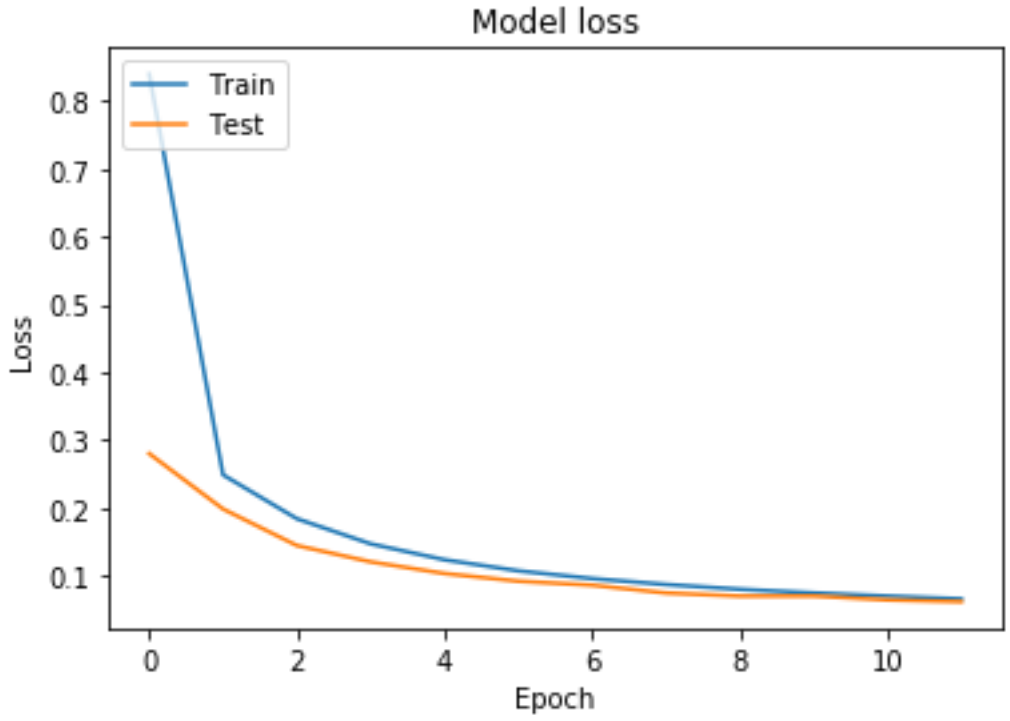

Fig 1- Error rate of training and testing data vs the number of epochs

3. Plots for the learning/testing errors and the final classification accuracy of the case where kernel size is changed to 5x5.

The CNN model described above is run with kernel size 5x5 and first convolutional layer has 6 feature maps and second convolutional layer has 16 feature maps. The observed overall loss and accuracy is summarized in the given table:

Table 2: Testing error and Final Classification Accuracy where kernel size is 5x5

Parameter	Observation
Testing Error	59.06998049425893%
Final Classification Accuracy	98%

Plots for the training error and testing error as a function of the learning epochs for the case with kernel size 5x5 is illustrated by Fig 2.

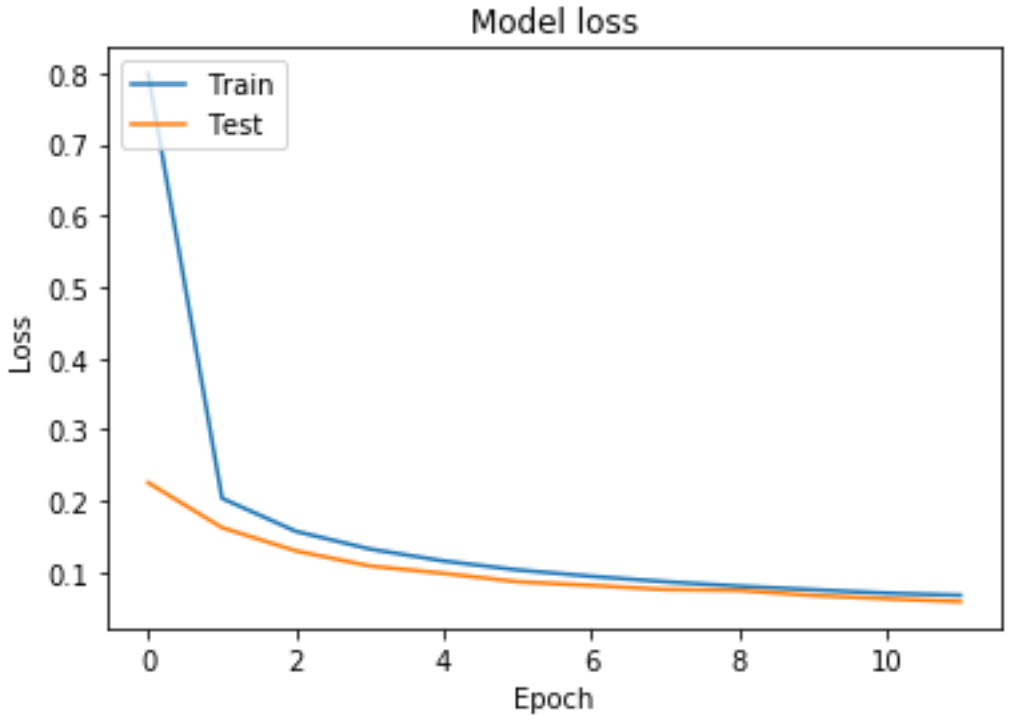

Fig 2. Error rate of training and testing data vs the number of epochs for 5x5 kernel

4. Plots for the learning/testing errors and the final classification accuracy of the case where the number of feature maps is modified to 20.

The CNN model described above is run with kernel size 5x5 and first convolutional layer has 20 feature maps and second convolutional layer has 20 feature maps. The observed overall loss and accuracy is summarized in the given table:

Table 3: Testing error and Final Classification Accuracy where number of feature maps is 20

Parameter	Observation
Testing Error	49.53636176921427%
Final Classification Accuracy	98.39%

Plots for the training error and testing error as a function of the learning epochs for the case with number of feature maps modified to 20 is illustrated by Fig 3.

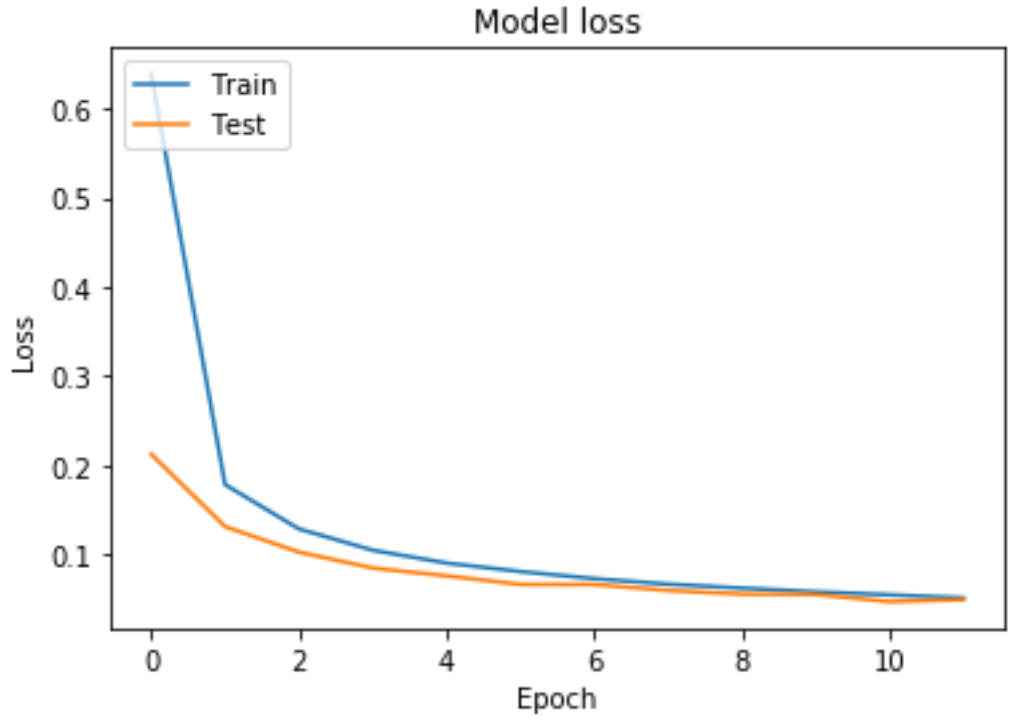

Fig 3- Error rate of training and testing data vs the number of epochs where number of feature maps is 20.