CSE 575: STATISTICAL MACHINE LEARNING

Project 3: Classification Using Neural Networks and Deep Learning

Raktim Mukhopadhyay

ASU ID: 1217167380

Overview

- **Dataset Description**
  We are working with MNIST dataset in this project. The MNIST dataset is an acronym that stands for the Modified National Institute of Standards and Technology dataset. It is a dataset of 60,000 small square 28×28 pixels grayscale images of handwritten single digits between 0 and 9.

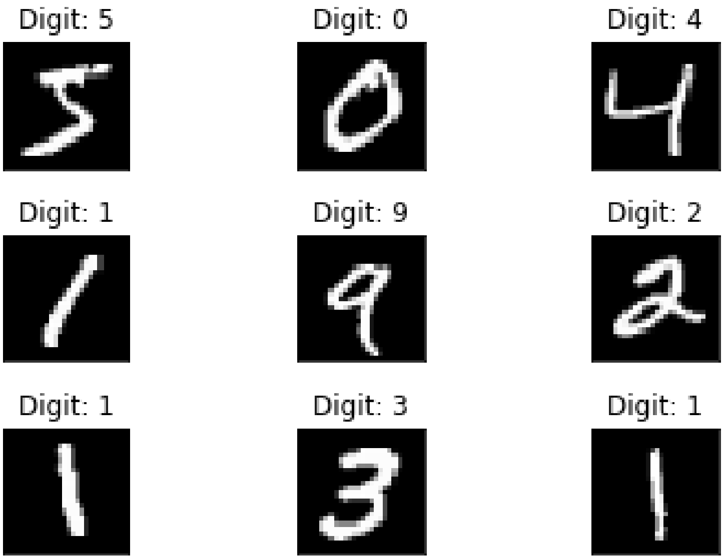

Fig 1: Sample images of some digits from the dataset

- **Convolution Neural Network**
  A Convolution Neural Network is a deep learning algorithm which can take in an input, assign weights to various features and be able to differentiate the inputs into various categories or classes. The architecture of CNN is analogous to that of the connectivity pattern of neurons in human brain and was inspired heavily by the visual cortex.

  In this project the baseline model has the following features -
  1. The first convolution layer has 6 feature maps and the convolution kernels are of size 3x3. This layer uses stride equal to 1. Stride denotes the number of pixels shifts over the input matrix. When stride is 1, the filters are moved by 1 pixel at a time.
  2. The convolution layer is followed by a max pooling layer. The poling is 2x2 with stride equal to 1.
  3. The max pooling layer is connected to the next convolution layer with 16 feature maps. The size of the kernel is 3x3 and this layer too uses stride equal to 1.
  4. The second convolution layer is followed by a max pooling layer. The pooling is 2x2 and uses a stride equal to 1.
  5. The max pooling layer is fully connected to the next hidden layer with 120 nodes and ReLU as the activation function.
  6. The fully connected layer is followed by another fully connected layer with 84 nodes and ReLU as the activation function, then connected to a SoftMax layer with 10 output nodes which corresponds to 10 classes/categories present in the data.

Objectives

1. Run the baseline code provided and report the accuracy on the testing dataset.
2. Change the kernel size from 3x3 to 5x5, report the testing error and accuracy on the testing dataset and plot training error and testing error vs epoch.
3. Change the number of the feature maps in the first and second convolutional layers, plot the training error and testing error along with the epoch, and report the testing error and accuracy on the test set.

Baseline Model

This uses a 3x3 kernel for both the first and second convolution layer. The number of feature maps in the baseline model is 6 for the first and 16 for the second convolution layer.

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.78	0.23	0.17	0.14	0.12	0.11	0.10	0.09	0.08	0.08	0.07	0.07

Table 1: Epochs and the corresponding training error for the Baseline Model

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.26	0.18	0.13	0.12	0.11	0.09	0.08	0.08	0.07	0.07	0.07	0.06

Table 2: Epochs and the corresponding testing error for the Baseline Model

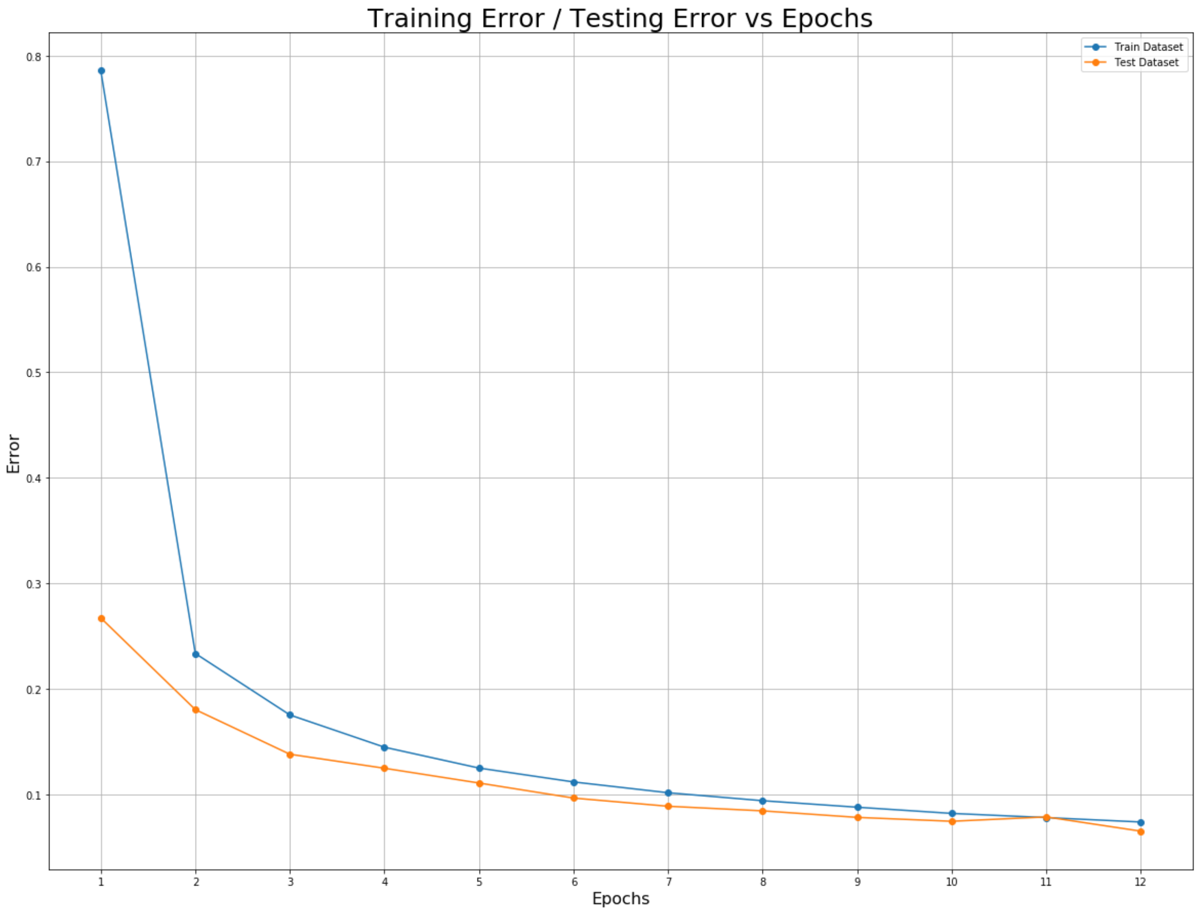

Fig 2: Plot showing Training and Testing errors vs Epoch for the Baseline Model

Testing Dataset Error - 0.06526645176736638

Testing Dataset Accuracy - 0.9781

Modified Model – 1

This uses a 5x5 kernel for both the first and second convolution layer. The number of feature maps in this model is 6 for the first and 16 for the second convolution layer.

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.76	0.24	0.19	0.16	0.13	0.12	0.10	0.10	0.09	0.08	0.08	0.07

Table 3: Epochs and the corresponding training error for the Modified Model – 1

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.27	0.20	0.18	0.14	0.13	0.10	0.09	0.09	0.08	0.08	0.07	0.06

Table 4: Epochs and the corresponding testing error for the Modified Model – 1

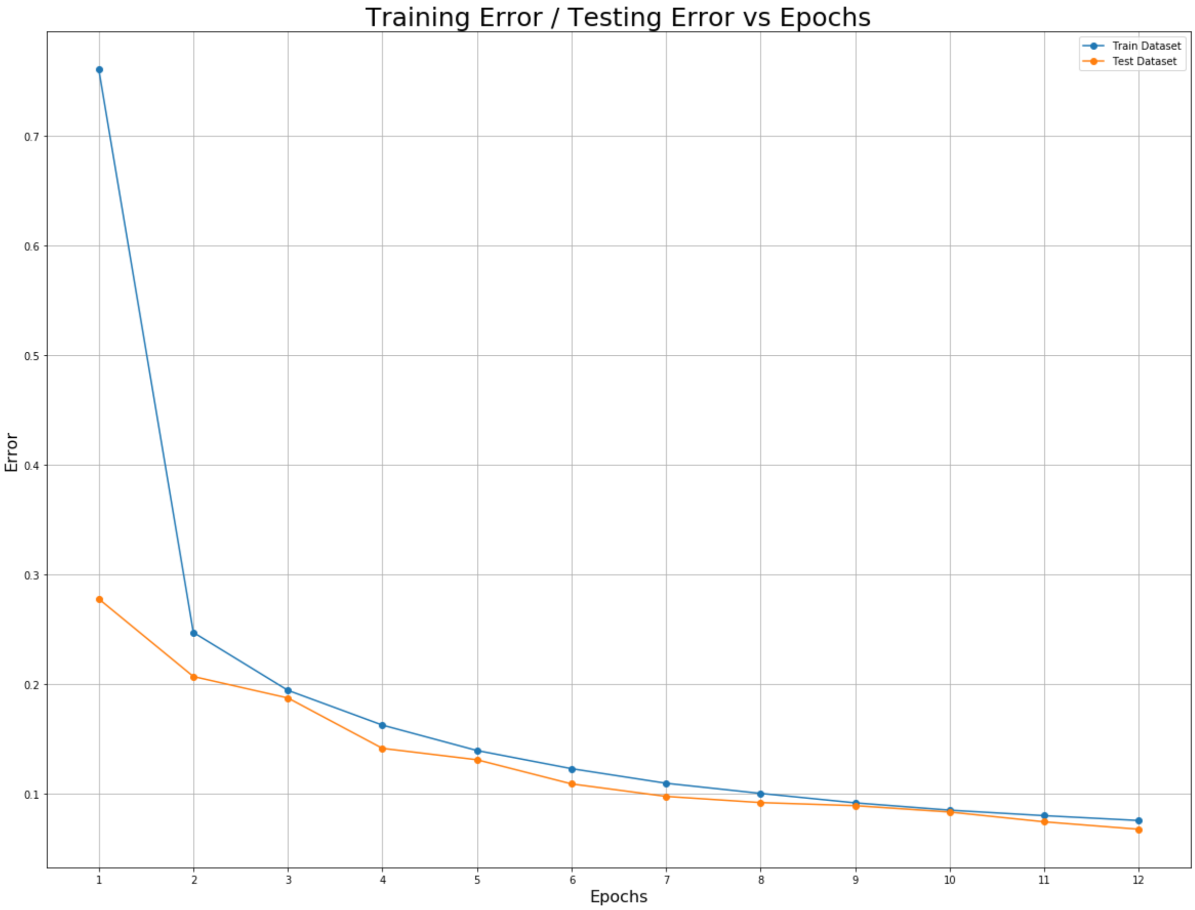

Fig 3: Plot showing Training and Testing errors vs Epoch for the Modified Model -1

Testing Dataset Error - 0.0679886656104587

Testing Dataset Accuracy - 0.9785

Modified Model – 2

This uses a 3x3 kernel for both the first and second convolution layer. The number of feature maps in this model is 32 for the first and 64 for the second convolution layer.

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.65	0.17	0.12	0.09	0.07	0.07	0.06	0.05	0.05	0.04	0.04	0.04

Table 5: Epochs and the corresponding training error for the Modified Model – 2

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.19	0.13	0.10	0.07	0.06	0.06	0.05	0.04	0.05	0.04	0.04	0.04

Table 6: Epochs and the corresponding testing error for the Modified Model – 2

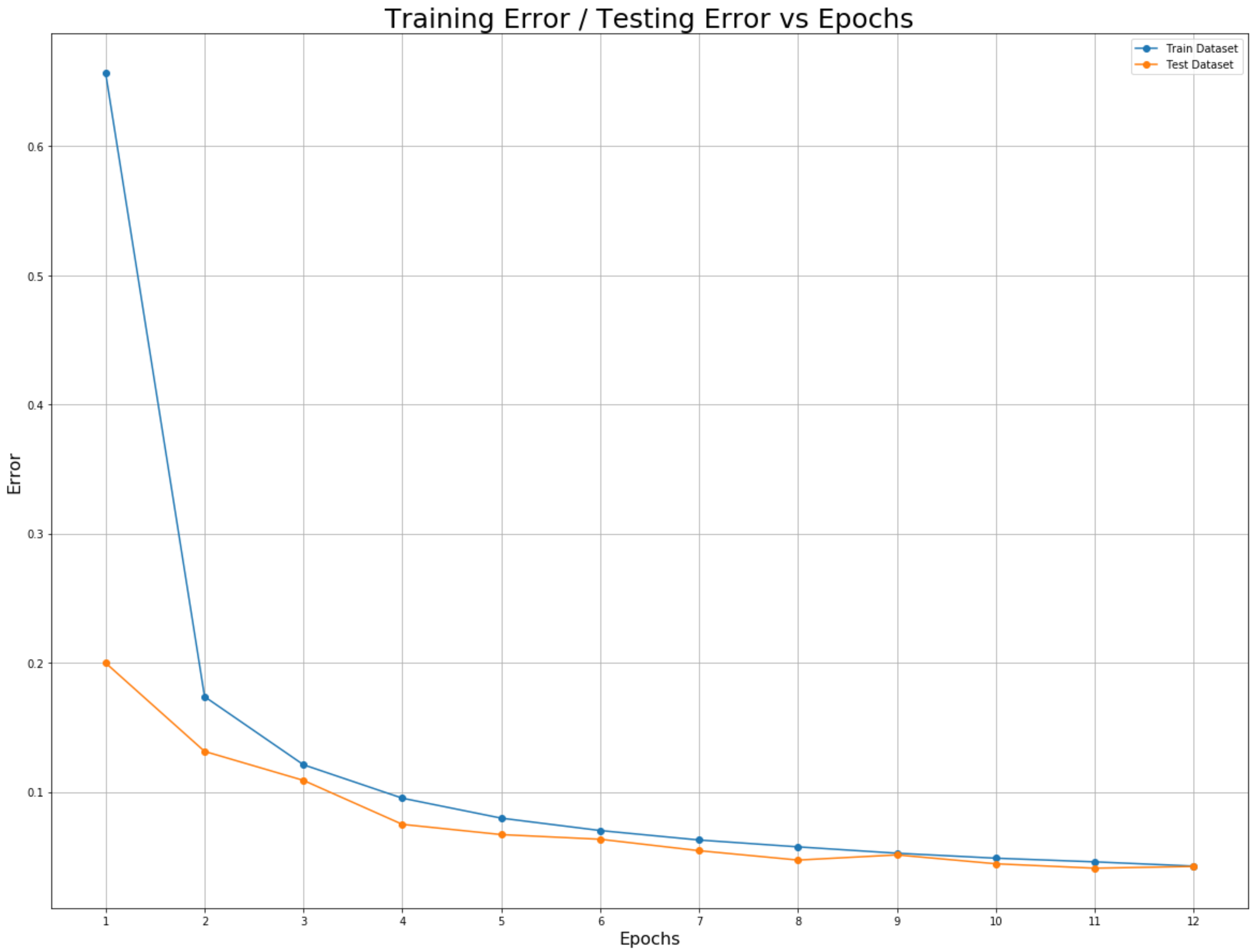

Fig 4: Plot showing Training and Testing errors vs Epoch for the Modified Model -2

Testing Dataset Error - 0.04239557061758824

Testing Dataset Accuracy - 0.9867

Modified Model – 3

This uses a 5x5 kernel for both the first and second convolution layer. The number of feature maps in this model is 32 for the first and 64 for the second convolution layer.

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.55	0.14	0.10	0.08	0.06	0.06	0.05	0.04	0.04	0.04	0.03	0.03

Table 7: Epochs and the corresponding training error for the Modified Model – 3

Epoch	1	2	3	4	5	6	7	8	9	10	11	12
Error	0.18	0.10	0.07	0.06	0.05	0.04	0.05	0.04	0.04	0.03	0.03	0.03

Table 8: Epochs and the corresponding testing error for the Modified Model – 3

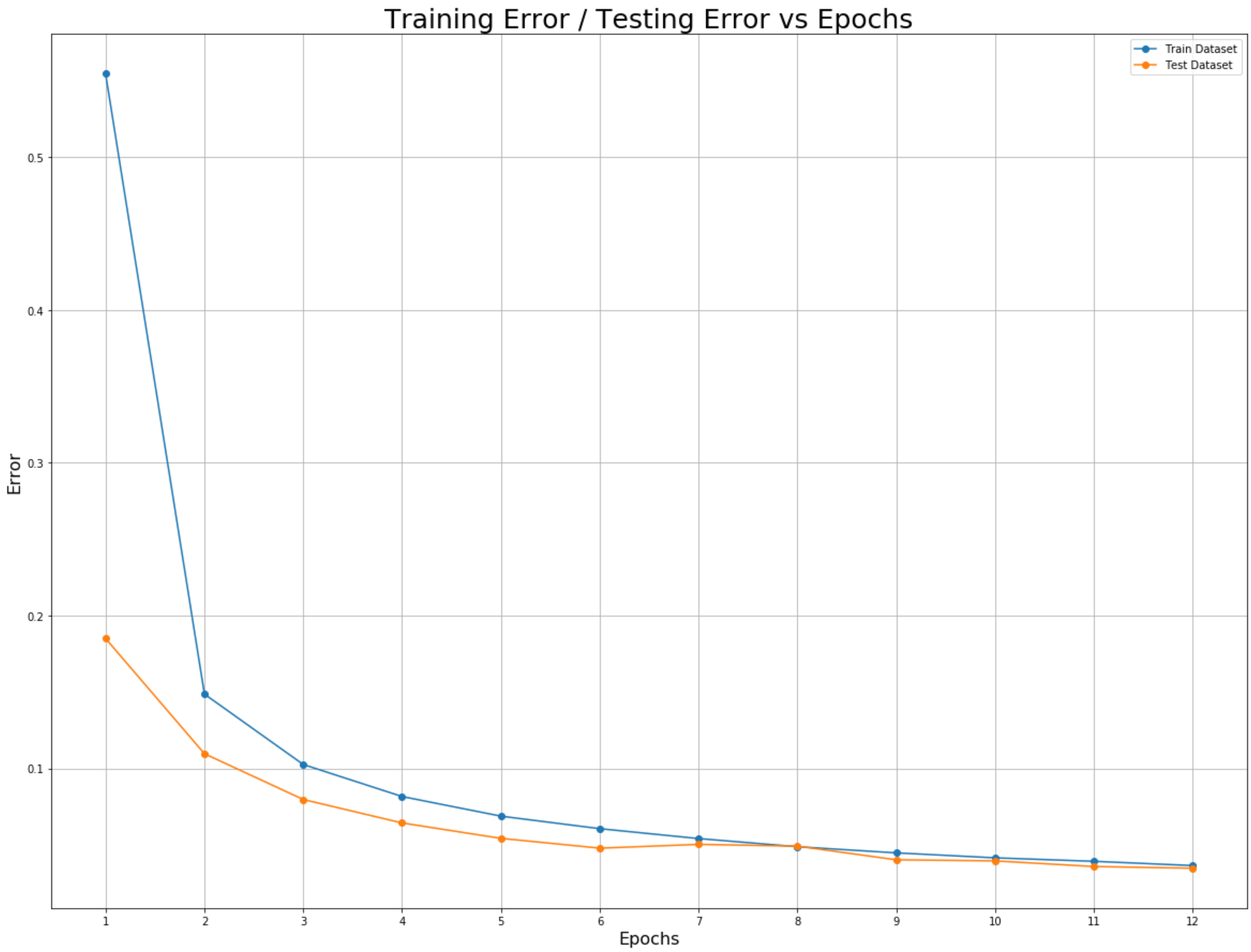

Fig 5: Plot showing Training and Testing errors vs Epoch for the Modified Model -3

Testing Dataset Error - 0.0347287143761059525

Testing Dataset Accuracy - 0.9887