**Introduction to Machine Learning (CSE 474/574)**

**Programming Assignment 2**

**Group Members:**

* Kaustuv Majumder ([kaustuvm@buffalo.edu](mailto:kaustuvm@buffalo.edu)) (50360369)
* Venkat Ashish Nagabhirava ([nagabhir@buffalo.edu](mailto:nagabhir@buffalo.edu)) (50388929)
* Ieshaan Sharma ([ieshaans@buffalo.edu](mailto:ieshaans@buffalo.edu)) (50367364)

**Introduction:**

A multilayer neural network will be implemented to classify handwritten digits on the MNIST Dataset. The same network will then be used analyze a face dataset (celebA) and compare the performance of the neural network against a deep neural network and a convolutional neural network using the TensorFlow Library.

**Datasets:**

1. MNIST Dataset:

The MNIST dataset [1] consists of a training set of 60000 examples and test set of 10000 examples. All digits have been size-normalized and centered in a fixed image of 28X28 size. In original dataset, each pixel in the image is represented by an integer between 0 and 255, where 0 is black, 255 is white and anything between represents different shade of gray.

1. CelebFaces Attributes Dataset (CelebA):

CelebFaces Attributes Dataset (CelebA) [3] is a large-scale face attributes dataset with more than 200K celebrity images. CelebA has large diversities, large quantities, and rich annotations, including:

* 10,177 number of identities,
* 202,599 number of face images, and
* 5 landmark locations, 40 binary attributes annotations per image.

**Feature Selection:**

In the dataset there are many features on which the values are exactly the same for all data points in the training set. With those feature the model cannot gain any more information about the difference (or variation) between data points. So, we can ignore those features and choose the ones that are useful.

**Number of important features = 717.**

**Hyper-Parameter Selection:**

We varied the regularization coefficient, 𝝺 and no of hidden units, m to choose the optimal values of the hyper-parameters.

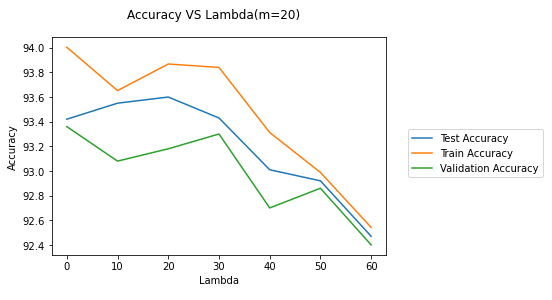
m has been taken as [4,8,12,16,20], and 𝝺 has been varied from 0 to 60 in steps of 10. The accuracies of test, train and validation data have been calculated with all the possible combinations of m and 𝝺. Below table shows illustrates the findings.



This table has been sorted in descending order of test accuracy.

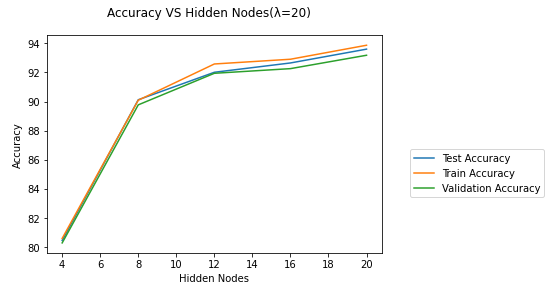
**Graphical Representations:**

1. The following graph captures the variation of accuracy with different 𝝺 values and a fixed m (=20)



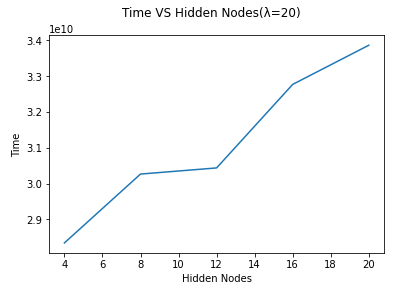
This graph shows that maximum test accuracy is achieved at 𝝺 = 20

1. The following graph captures the variation of accuracy with different m values and a fixed 𝝺 (=20)



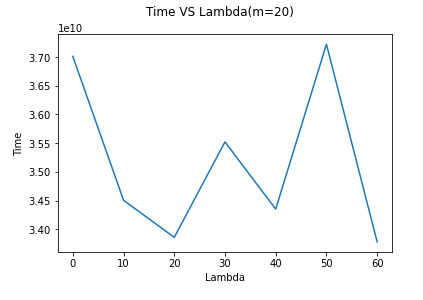
This graph shows that accuracy increases with increasing numbers of hidden nodes. There is a huge improvement when m increases from 4 to 8. After m=12, the rate of improvement slows down, indicating that increasing m beyond a certain limit will not make any significant improvement in accuracy.

1. The following graph captures the training time with different m values and a fixed 𝝺 (=20)



This graph indicates that training time of the model increases with increasing number of hidden nodes.

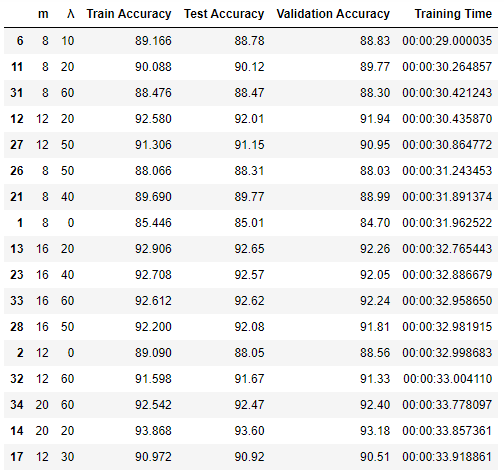
1. The following graph captures the training time with various values of 𝝺 and a fixed m(=20)



From this graph, no concrete conclusion can be drawn regarding the relation between 𝝺 and training time.

If training time is considered as an important factor, we need to find optimum values of 𝝺 and m considering training time along with accuracy.

The following table has been sorted in ascending order of training time. It only contains those data with accuracy > 85%



From the above table it can conclude that with 𝝺 = 10 and m = 8, we can still achieve good accuracy and maintain reasonable training time.

1. **Accuracy of classification method on the handwritten digits test data:**

The above findings can be summarized as follows:

Without time constraint:

* Regularization coefficient (𝝺) = 20
* Hidden Nodes (m) = 20
* Train Accuracy (%) = 93.87
* Validation Accuracy (%) = 93.18
* Test Accuracy (%) = 93.60

With time constraint:

* Regularization coefficient (𝝺) = 10
* Hidden Nodes (m) = 8
* Train Accuracy (%) = 89.17
* Validation Accuracy (%) = 88.83
* Test Accuracy (%) = 88.78

1. **Accuracy of classification method on the CelebA data set:**

* Regularization coefficient (𝝺) = 10
* Hidden Nodes (m) = 256
* Train Accuracy (%) = 84.66
* Validation Accuracy (%) = 83.19
* Test Accuracy (%) = 85.09
* Training Time (sec) = 99.663298

1. **Comparison of implemented neural network with a deep neural network:**

The deep neural network has been executed with various epoch and learning rates, and below is the finding:

|  |  |  |
| --- | --- | --- |
| **Epoch** | **Leaning Rate** | **Accuracy** |
| 100 | 0.0001 | 81.87 |
| 100 | 0.0003 | 82.5 |
| 200 | 0.0003 | 81.76 |
| 1000 | 0.0003 | 85.16 |

The deep neural network is then executed with a fixed epoch and learning rate (100, 0.0001), and with different hidden layers. The accuracies thus observed have been reported below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Epoch** | **Leaning Rate** | **Hidden Layer** | **Accuracy** | **Time** |
| 100 | 0.0001 | 2 | 82.43 |  |
| 100 | 0.0001 | 3 | 79.48 |  |
| 100 | 0.0001 | 5 | 77.81 |  |
| 100 | 0.0001 | 7 | 75.63 |  |

Comparing the finding with neural network implemented previously:

Neural Network:

* Accuracy (%) = 85.09
* Time =

Deep Neural Network:

* Accuracy (%) = 82.43
* Time =

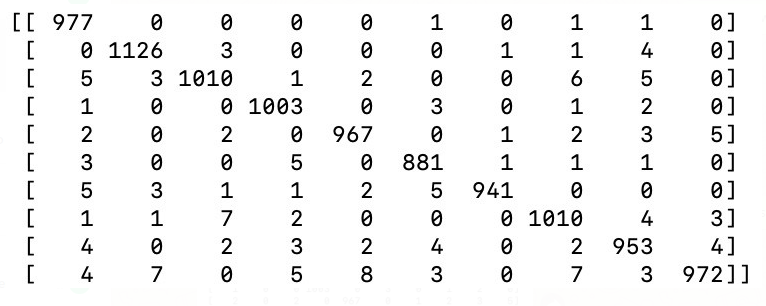
We can infer that highest accuracy is achieved with 2 hidden layers. From there on, the accuracy is decreasing. This can be attributed to the overfitting of the model.

1. **Results from convolutional neural network in terms of accuracy and training time:**

The CNN gives much better accuracy compared to single layer and multi-layer neural network.

* Accuracy (%) = 98.7% (9872 / 10000)
* Training Time = 0:11:44

Confusion matrix:



From the confusion matrix it is evident that most of the predictions are as per expectation.