Marshall Lanning

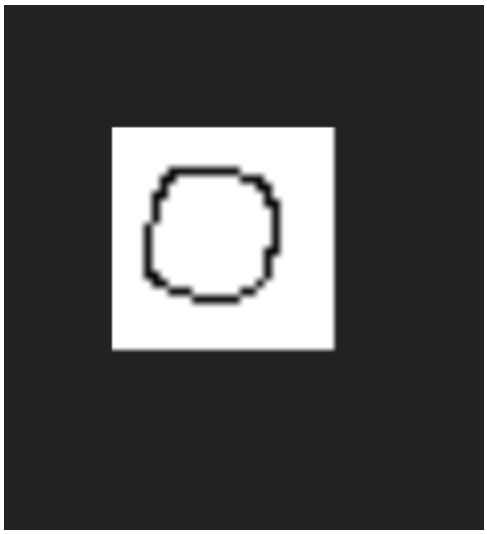
CECS 625 Parallel Programming Extra-cedit Project
October 29, 2019 (50 points)

Due: November 7 (Wed) Demo your extra-credit project in class.

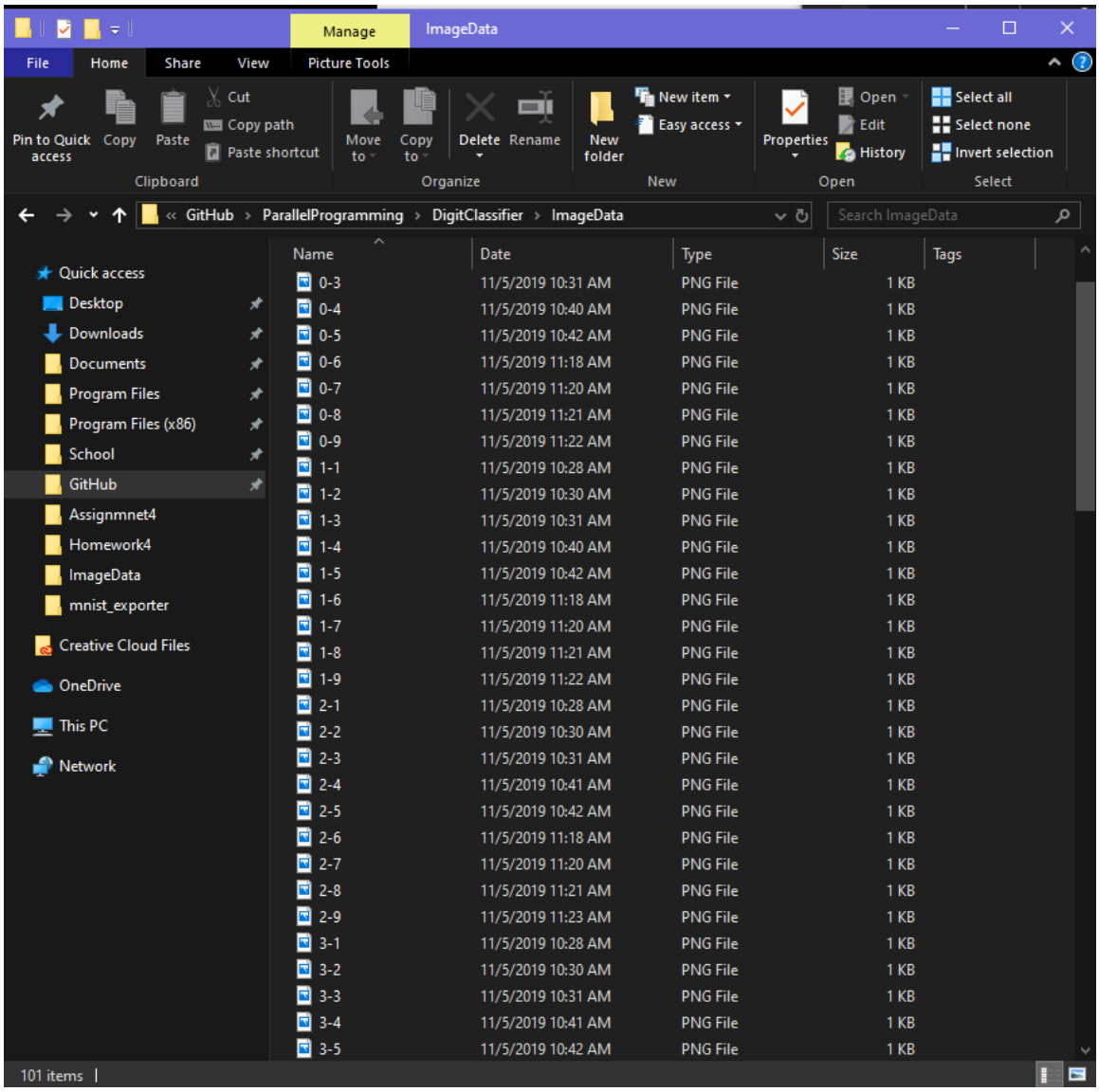
Based on the materials you learned from Section 6.3 and 6.5, write a C++ project to implement a written digit classifier as described below:

Input: An 28x28 float vector (i.e. an 28X28 grey-level image representin a written digit from 0-9)

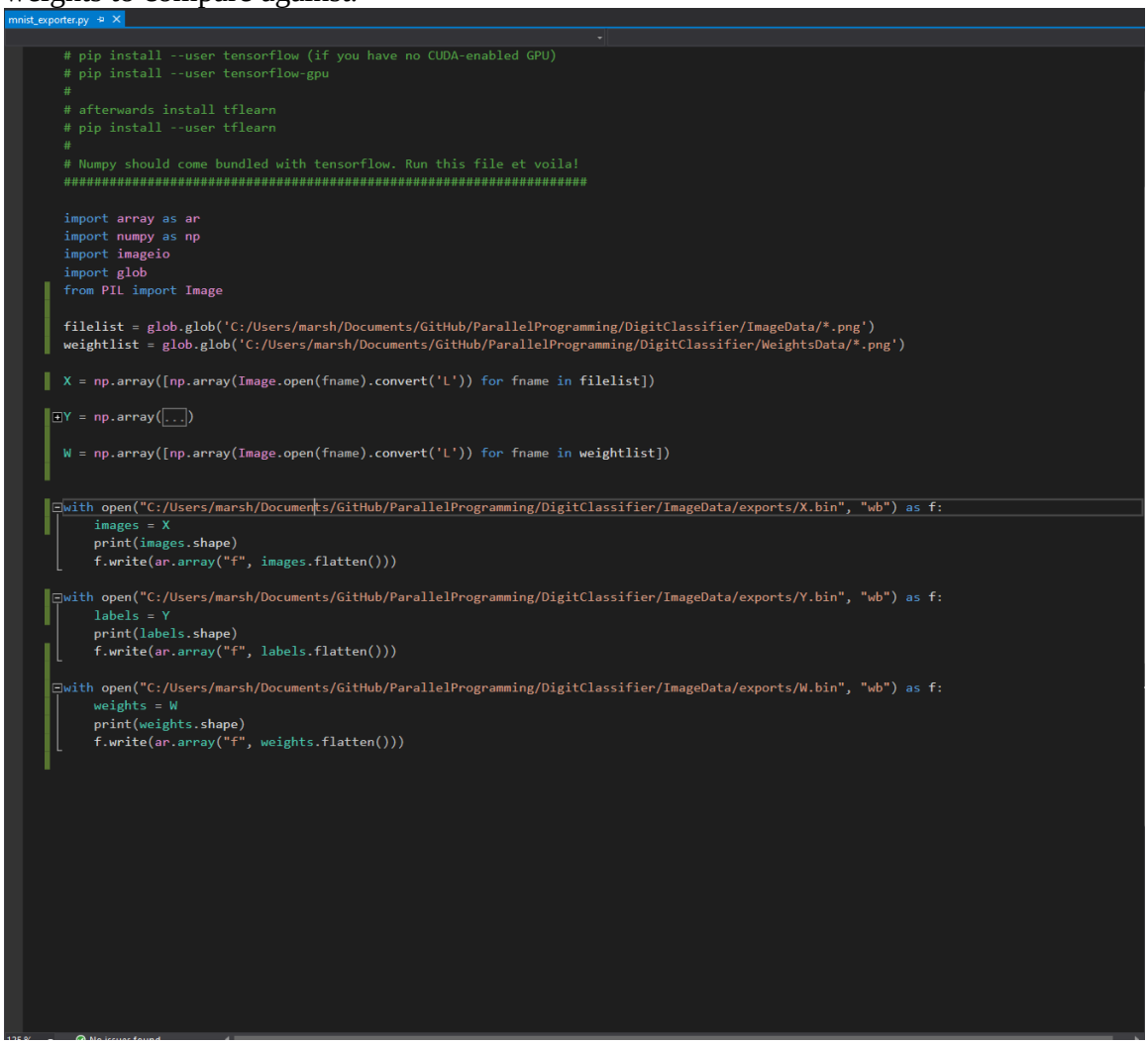
I used the Windows Paint application to write out 10 of each number in a 28X28 pixel fram so 100 numbers total. An example of 0 is below:



```
Input Data:
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1
1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

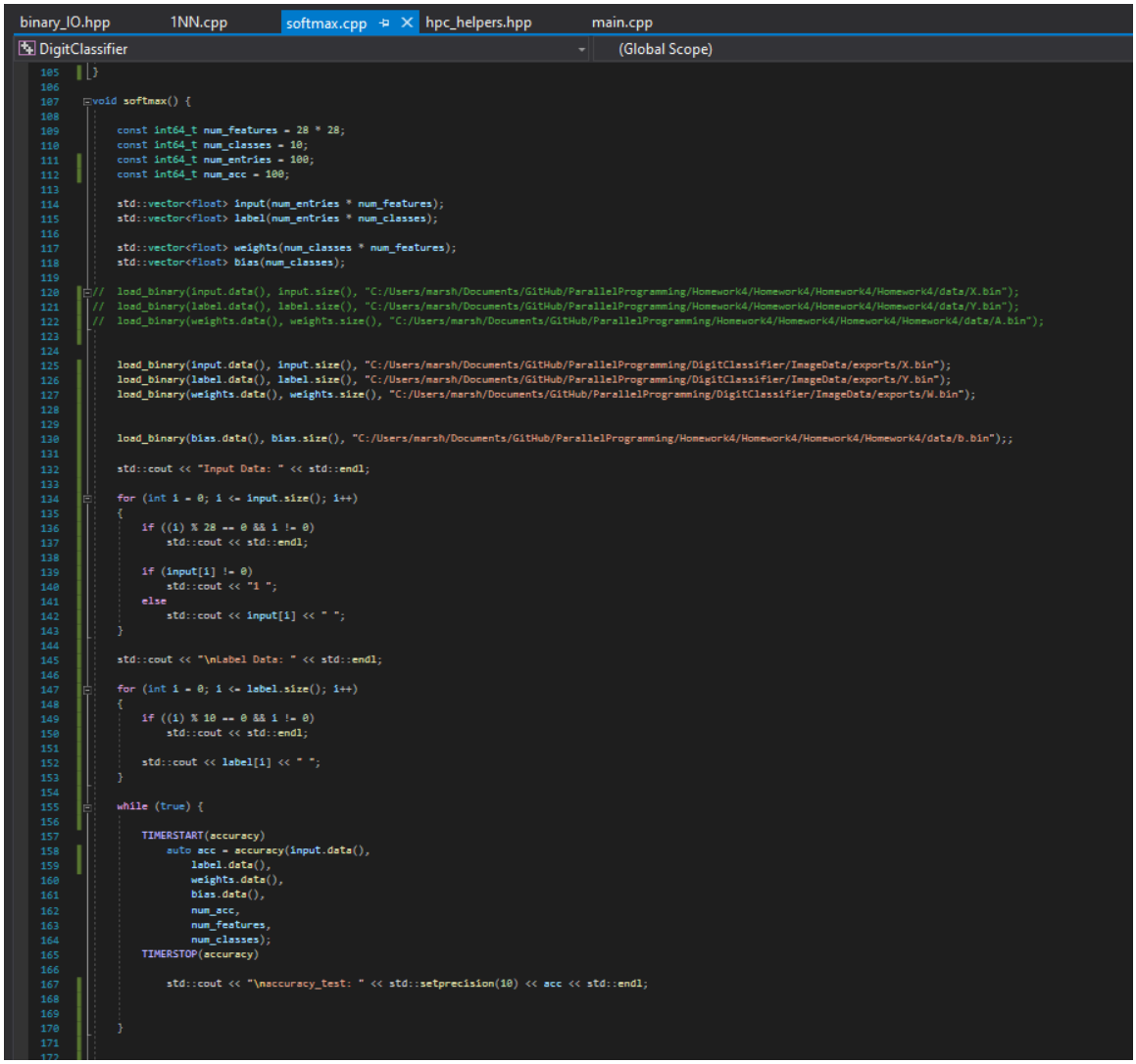

Manage
ImageData
File
Home
Share
View
Picture Tools
Pin to Quick access
Copy
Paste
Cut
Copy path
Paste shortcut
Clipboard
Move to
Copy to
Delete
Rename
Organize
New folder
New item
Easy access
New
Properties
Open
Edit
History
Open
Select all
Select none
Invert selection
Select
« GitHub › ParallelProgramming › DigitClassifier › ImageData
Search ImageData
Quick access
Desktop
Downloads
Documents
Program Files
Program Files (x86)
School
GitHub
Assignmnet4
Homework4
ImageData
mnist_exporter
Creative Cloud Files
OneDrive
This PC
Network
Name
Date
Type
Size
Tags
0-3 11/5/2019 10:31 AM PNG File 1 KB
0-4 11/5/2019 10:40 AM PNG File 1 KB
0-5 11/5/2019 10:42 AM PNG File 1 KB
0-6 11/5/2019 11:18 AM PNG File 1 KB
0-7 11/5/2019 11:20 AM PNG File 1 KB
0-8 11/5/2019 11:21 AM PNG File 1 KB
0-9 11/5/2019 11:22 AM PNG File 1 KB
1-1 11/5/2019 10:28 AM PNG File 1 KB
1-2 11/5/2019 10:30 AM PNG File 1 KB
1-3 11/5/2019 10:31 AM PNG File 1 KB
1-4 11/5/2019 10:40 AM PNG File 1 KB
1-5 11/5/2019 10:42 AM PNG File 1 KB
1-6 11/5/2019 11:18 AM PNG File 1 KB
1-7 11/5/2019 11:20 AM PNG File 1 KB
1-8 11/5/2019 11:21 AM PNG File 1 KB
1-9 11/5/2019 11:22 AM PNG File 1 KB
2-1 11/5/2019 10:28 AM PNG File 1 KB
2-2 11/5/2019 10:30 AM PNG File 1 KB
2-3 11/5/2019 10:31 AM PNG File 1 KB
2-4 11/5/2019 10:41 AM PNG File 1 KB
2-5 11/5/2019 10:42 AM PNG File 1 KB
2-6 11/5/2019 11:18 AM PNG File 1 KB
2-7 11/5/2019 11:20 AM PNG File 1 KB
2-8 11/5/2019 11:21 AM PNG File 1 KB
2-9 11/5/2019 11:23 AM PNG File 1 KB
3-1 11/5/2019 10:28 AM PNG File 1 KB
3-2 11/5/2019 10:30 AM PNG File 1 KB
3-3 11/5/2019 10:31 AM PNG File 1 KB
3-4 11/5/2019 10:41 AM PNG File 1 KB
3-5 11/5/2019 10:42 AM PNG File 1 KB
101 items

I used python like the mnist_exporter to export my data set correctly and make my own weights to compare against.



Output: The digit in the image (i.e. 0, 1, 2, 3, 4, 5, 6, 7, 8, or 9)

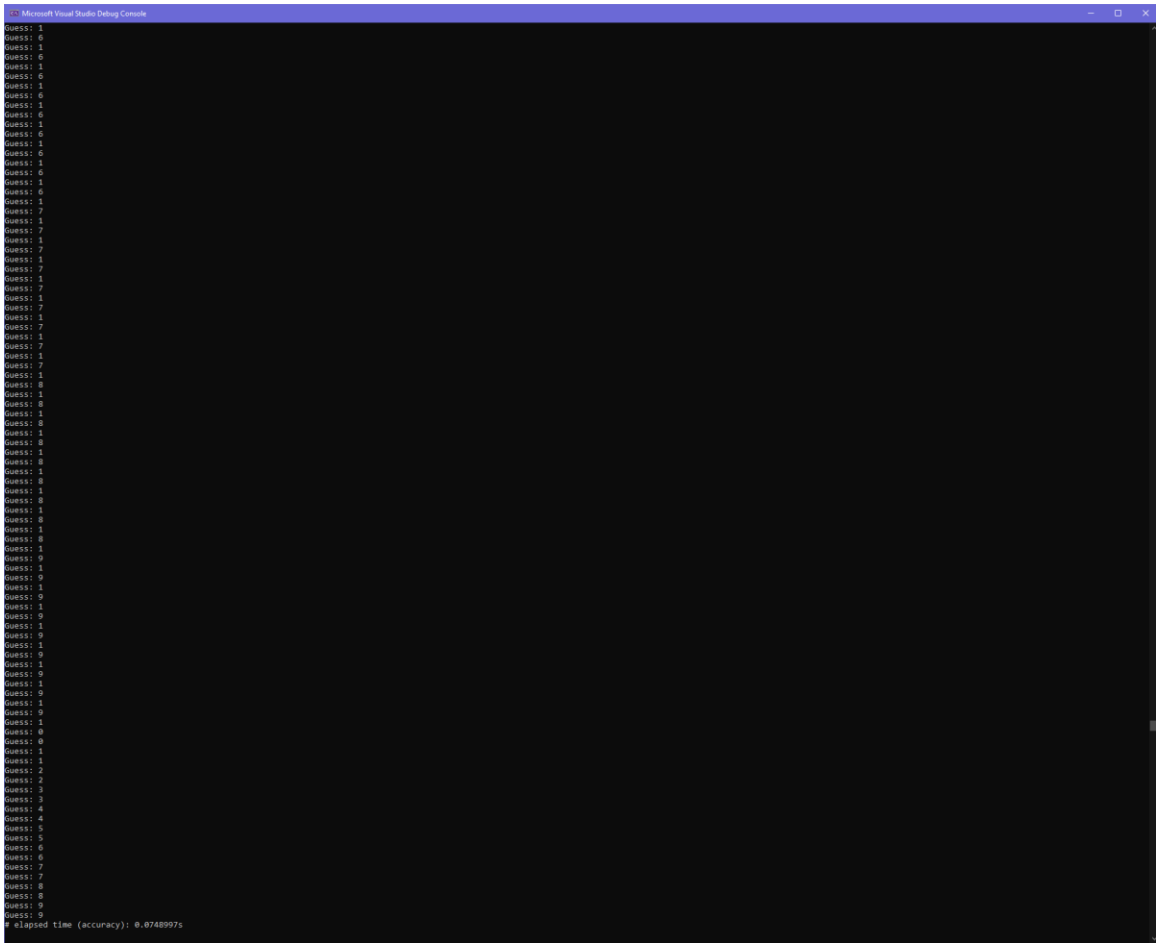
Softmax implementation:



Nearest Neighbor Implementation:

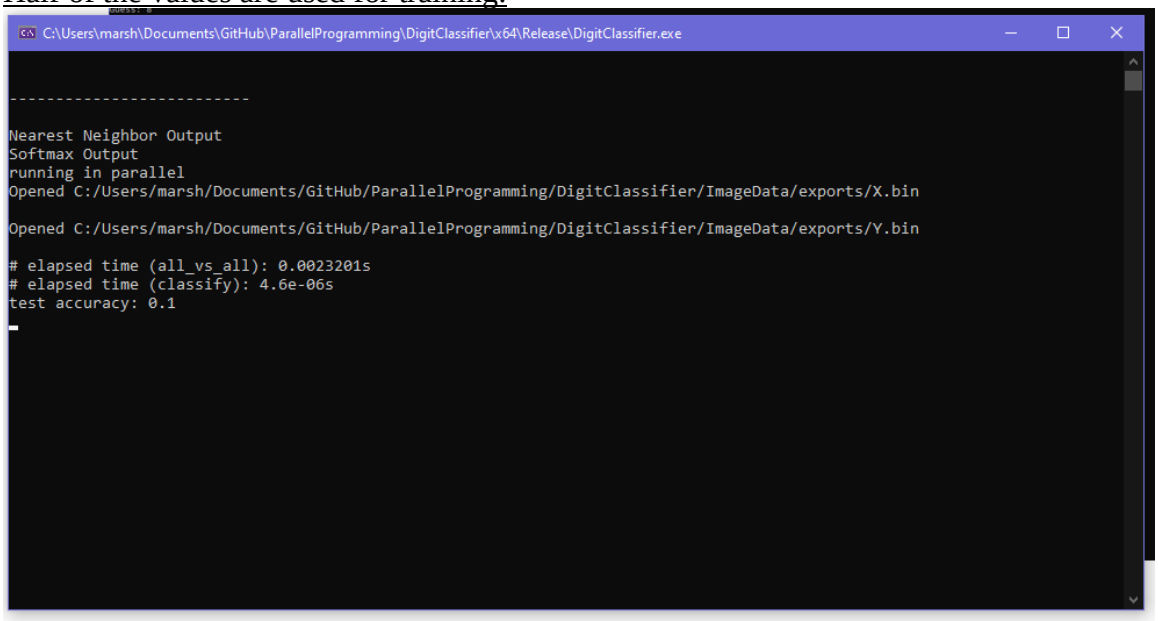
```
}

void NearestNeighbor() {

    // run parallelized when any command line argument given
    const bool parallel = true;

    std::cout << "running "
        << (parallel ? "in parallel" : "sequentially")
        << std::endl;

    // the shape of the data matrices
    const int64_t num_features = 28 * 28;
    const int64_t num_classes = 10;
    const int64_t num_entries = 100;
    const int64_t num_train = 90;
    const int64_t num_test = num_entries - num_train;

    // memory for the data matrices and all-pair matrix
    std::vector<float> input(num_entries * num_features);
    std::vector<float> label(num_entries * num_classes);
    std::vector<float> delta(num_test * num_train);

    // get the images and labels from disk
    load_binary(input.data(), input.size(), "C:/Users/marsh/Documents/GitHub/ParallelProgramming/DigitClassifier/ImageData/exports/X.bin");
    load_binary(label.data(), label.size(), "C:/Users/marsh/Documents/GitHub/ParallelProgramming/DigitClassifier/ImageData/exports/Y.bin");

    TIMERSTART(all_vs_all)
        const int64_t inp_off = num_train * num_features;
    all_vs_all(input.data() + inp_off,
        input.data(),
        delta.data(),
        num_test, num_train,
        num_features, parallel);
    TIMERSTOP(all_vs_all)

        TIMERSTART(classify)
        const int64_t lbl_off = num_train * num_classes;
    auto acc = accuracy(label.data() + lbl_off,
        label.data(),
        delta.data(),
        num_test, num_train,
        num_classes, parallel);
    TIMERSTOP(classify)

        std::cout << "test accuracy: " << acc << std::endl;

    int n = 0;

    std::cin >> n;
}
```

Output of Softmax guessing an iteration of my 100 numbers. The way the numbers are guessed seem somewhat accurate, the order that they are implemented in is 9 0's, 9 1's, 9 2's, etc so you see the guesses increasing in value as it moves through the dataset.

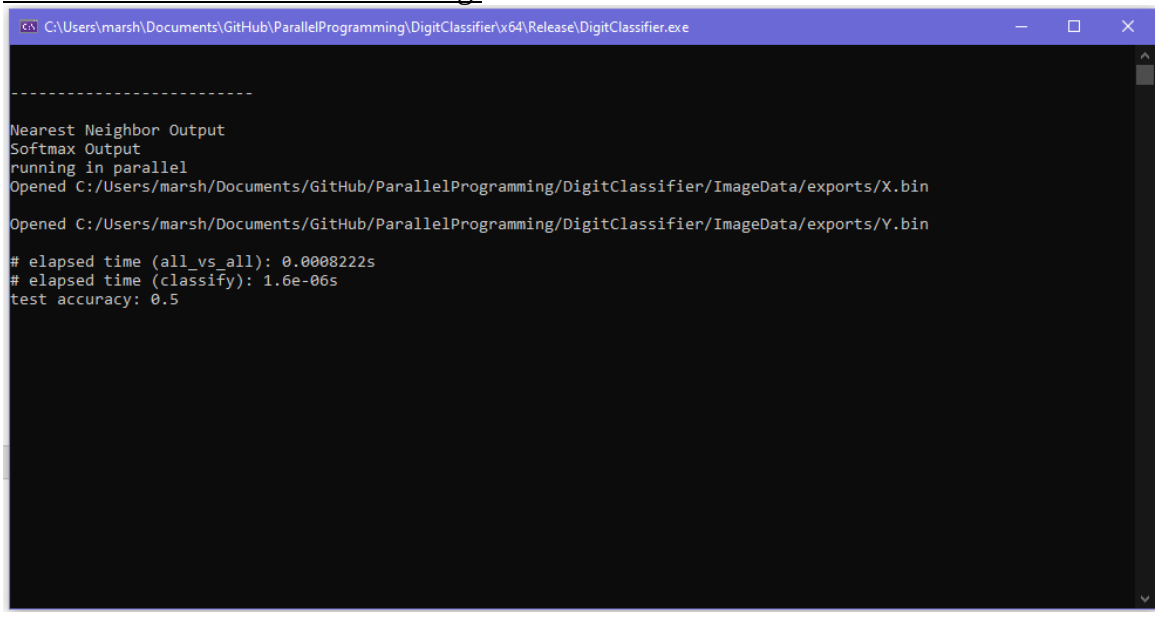


1NN Implementation:

Half of the values are used for training:



90 of the values are used for training:



Better accuracy when you train more!

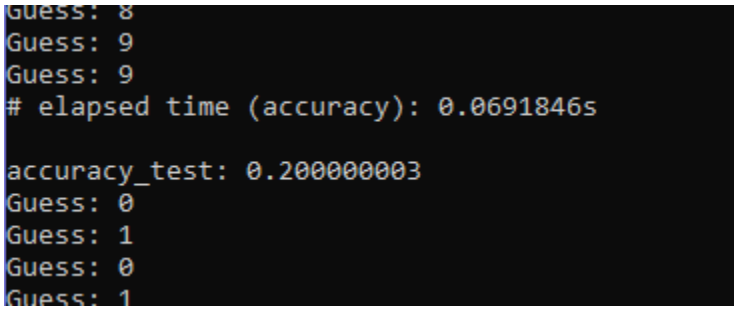
I believe my Softmax and 1NN implementations are not very accurate. This is because of my dataset. It is not consistent and a lot of room for ambiguity because the utility used with paint draws a line that covers a 1X1 pixel dimension. In other words, the written digits take up a very small amount of the solution space.

I also was not very careful about making sure they all looked a like and were in the same area. For example, look at my 1's:

Compare the accuracy and computing times of these two classifier algorithms.

My Softmax implementation can be better by improving the dataset. Espically the weights data set which is just example images of 0-9. If I made them more consistent and possibly even had more data to use I would see more accuracy. The computing times are really not much different and my dataset is not big enough to have significance. In the original dataset for Homework 4 the timing of Softmax was better than 1NN, Softmax can classify these images in about half the time.

Softmax:



1NN: