The program structure

Dataloader.py:

- Converts the mnist data into tables

Main.py:

- At the moment it just takes a number from the test data and tests it against the train data using **KNN.py**

KNN.py:

GetErrorPrecentage():

- Returns the percentage of numbers the algorithm fails to recognise

CreatePixelBoolCoordinateNumber()

- Converts an image table into a list of colored pixel coordinates

CreatePixelBoolCordinateTable():

- Uses **CreatePixelBoolCoordinateNumber()** toconvert multiple data entries into lists of colored pixel coordinates.

GetNeighbouringCoordinateRange():

- Returns a list of surrounding coordinates of a coordinate

GetClosestNeighbour():

- Returns the distance of the closest colored coordinate to a given coordinate. Either with the range of **GetNeighbouringCoordinateRange()** or the whole pixel coordinate image.

DistanceToNearestNeighbour():

- Controls the search radius of **GetClosestNeighbour()** to optimize runtime returns the distance to the nearest colored coordinate of a given coordinate

CompareNumberWithBoolCordinates():

- Checks the distance between colored pixels between a given image and images from the test data using **DistanceToNearestNeighbour()**. Then it returns a list containing the index of the training data and the distance, the list is sorted so the smallest distances comes first.

GetTheMajorityNeighbourNumber():

- Gets the closest numbers from the train data using a list form **CompareNumberWithBoolCordinates()**. Then it chooses the number that the majority of the closest numbers are.

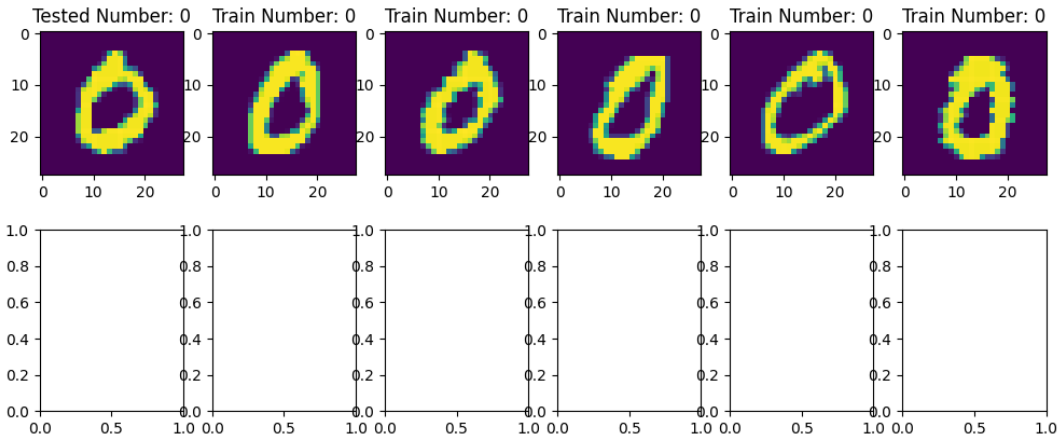
Time Complexity

- Converting to bool tables takes O(n*784) time
- Number recognition worst case takes $O(n*784^2)$

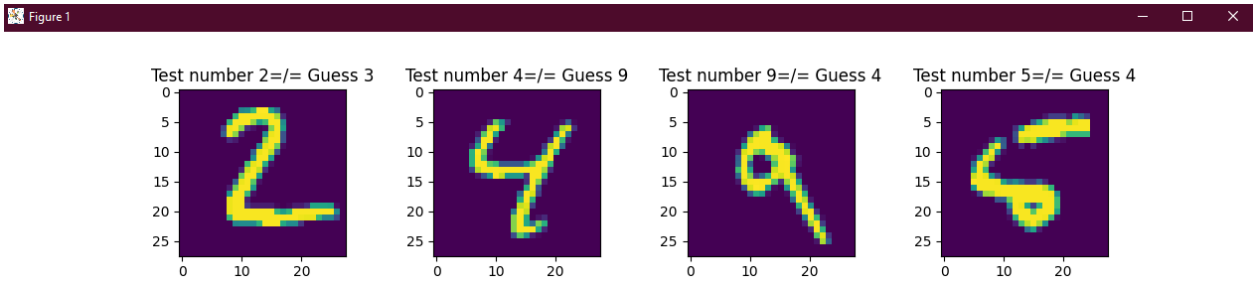
The Programs Functionality

Right now the program can recognize handwritten digits from with an accuracy of (91% succeed,due to me not wanting to spend half a day waiting for the result, i tested only 10 numbers against a training set of 10000 entries wich makes the recognition weaker)

The program output looks like this:



Error percentage test:



Things missing

- The ability to draw numbers yourself
- Optimization can always be done

References:

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=576361

Advice from Hannu, which I'm very thankful for.