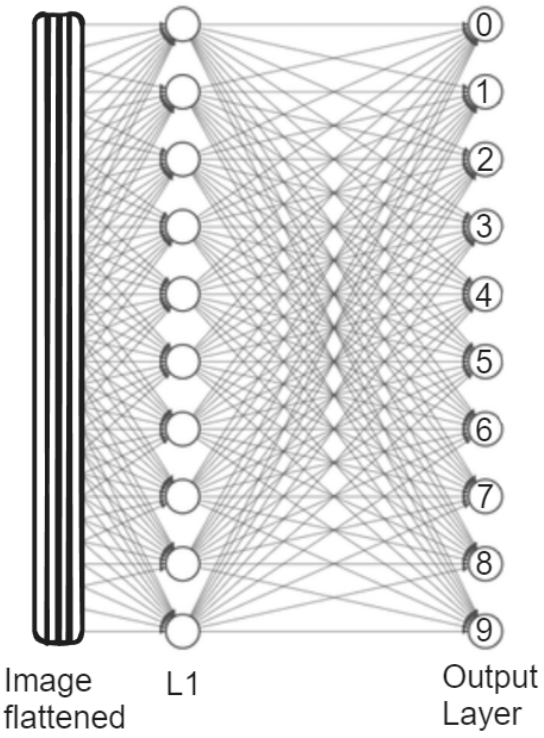
Neuron Understanding



This is a map of the simplest neural network created for Mnist dataset

This Neural Network flattens the 28 × 28 image into a long array to be sent into the network

It has 3 layers -

- Input layer (784 neurons)
- Layer 1 (10 neurons)
- Output Layer (10 neurons) - (Probabilities)

This paper targets to understand - In the condition of large data compression does a neuron depict significant understandable variables

(Spoiler - yes)

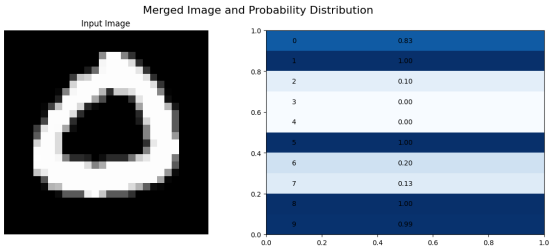
First of all this neural network is trained and gets up to 93% accuracy in test dataset

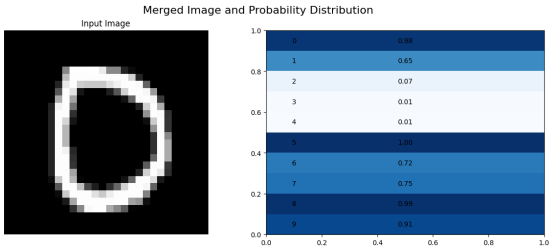
Secondly the entire code is uploaded

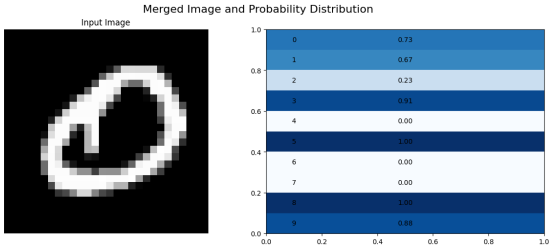
The main target to be understood is the activation of layer 1 (or the middle layer)

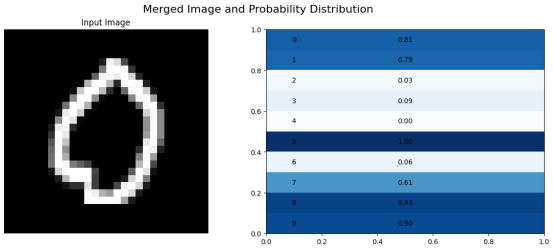
On the following examples, these were the activations of L1 layer neurons -

0

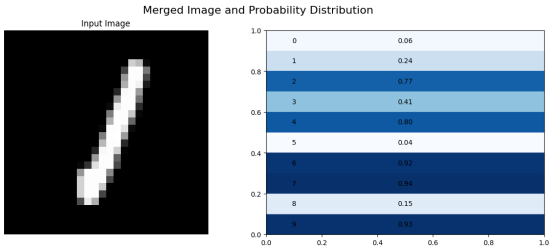


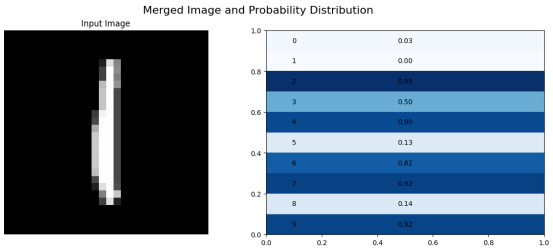


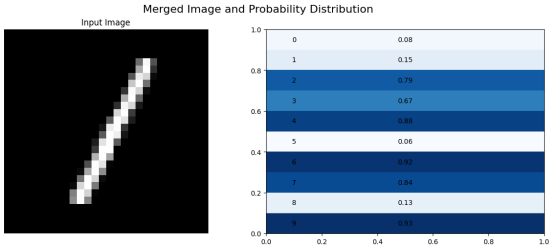


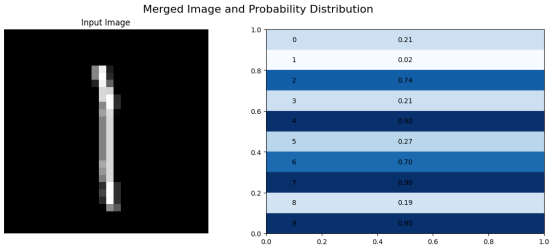


1

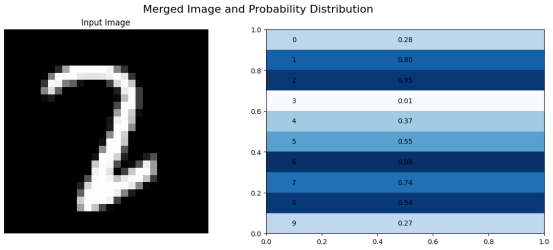


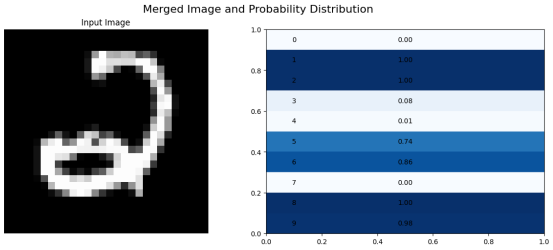


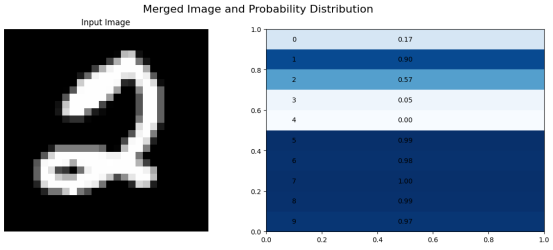

Merged Image and Probability Distribution
Input Image
0 0.08
1 0.15
2 0.79
3 0.67
4 0.88
5 0.06
6 0.92
7 0.84
8 0.13
9 0.93

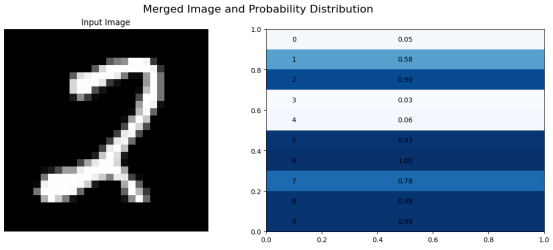

Merged Image and Probability Distribution
Input Image
0 0.21
1 0.02
2 0.74
3 0.21
4 0.90
5 0.27
6 0.70
7 0.50
8 0.19
9 0.90

2

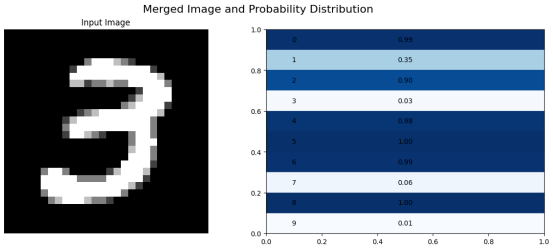

Merged Image and Probability Distribution
Input Image
0 0.28
1 0.80
2 0.95
3 0.01
4 0.37
5 0.55
6 0.98
7 0.74
8 0.94
9 0.27

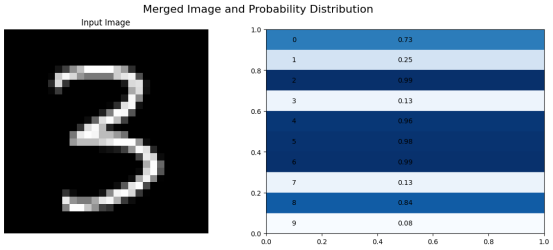

Merged Image and Probability Distribution
Input Image
0 0.00
1 1.00
2 1.00
3 0.08
4 0.01
5 0.74
6 0.86
7 0.00
8 1.00
9 0.98

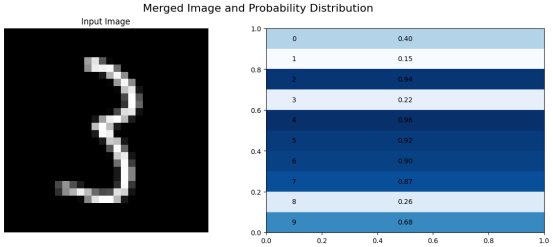

Merged Image and Probability Distribution
Input Image
0 0.17
1 0.90
2 0.57
3 0.05
4 0.00
5 0.99
6 0.98
7 1.00
8 0.99
9 0.97

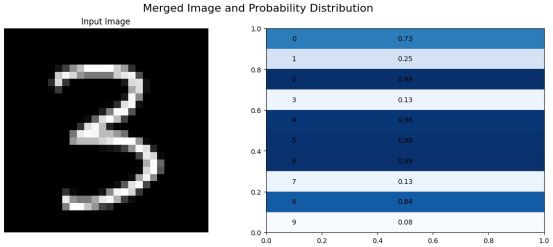

Merged Image and Probability Distribution
Input Image
0 0.05
1 0.58
2 0.90
3 0.03
4 0.06
5 0.97
6 1.00
7 0.78
8 0.98
9 0.98

3

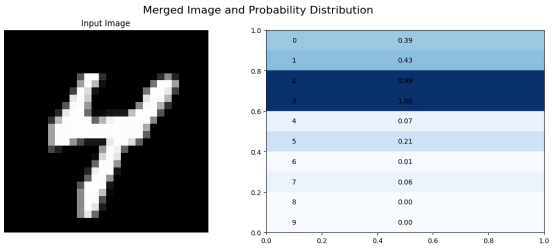

Merged Image and Probability Distribution
Input Image
0 0.99
1 0.35
2 0.90
3 0.03
4 0.98
5 1.00
6 0.99
7 0.06
8 1.00
9 0.01

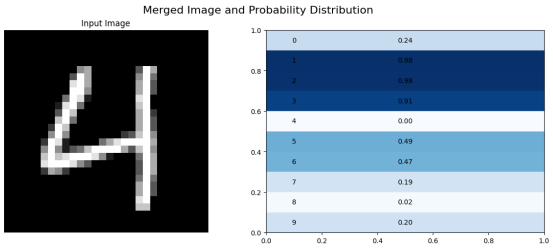

Merged Image and Probability Distribution
Input Image
0 0.73
1 0.25
2 0.99
3 0.13
4 0.96
5 0.98
6 0.99
7 0.13
8 0.84
9 0.08

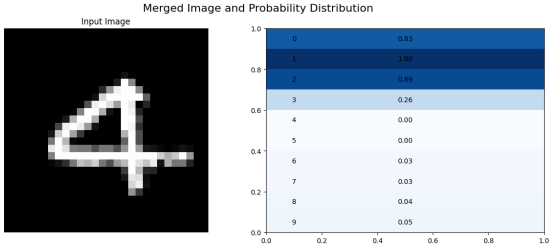

Merged Image and Probability Distribution
Input Image
0 0.40
1 0.15
2 0.94
3 0.22
4 0.96
5 0.92
6 0.90
7 0.87
8 0.26
9 0.68

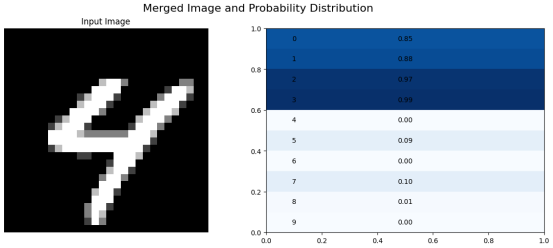

Merged Image and Probability Distribution
Input Image
0 0.73
1 0.25
2 0.99
3 0.13
4 0.96
5 0.98
6 0.99
7 0.13
8 0.84
9 0.08

4

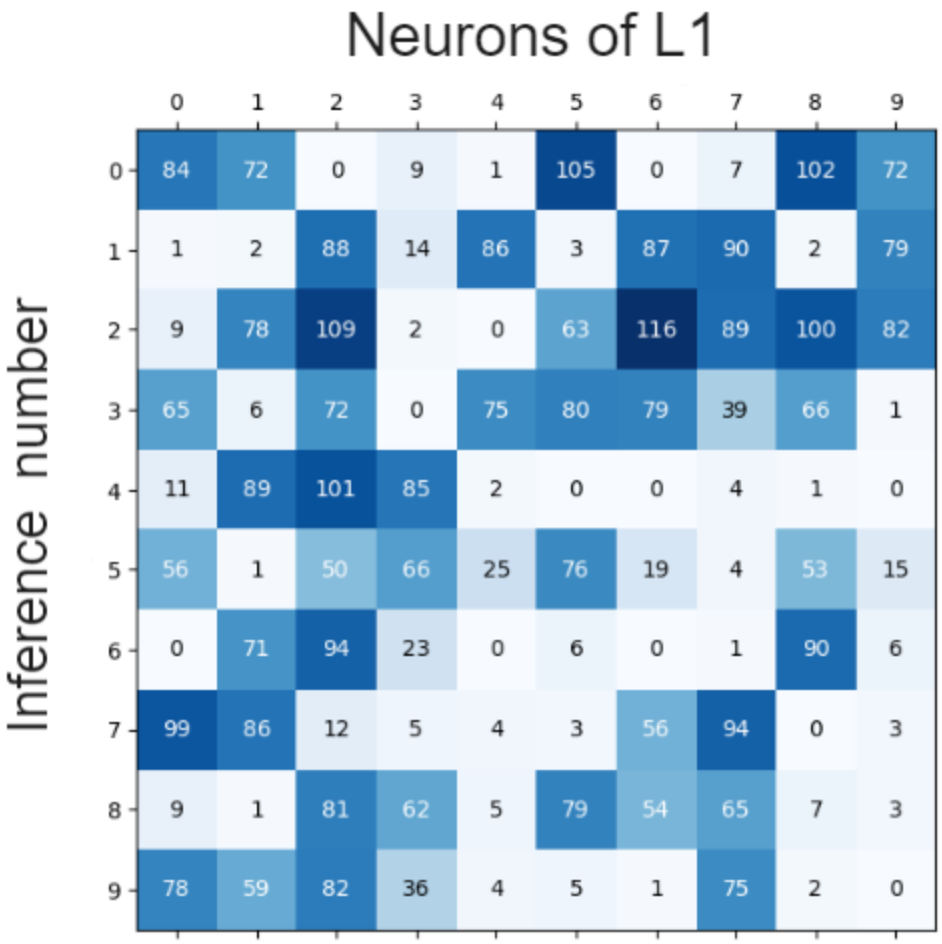

Merged Image and Probability Distribution
Input Image
0 0.39
1 0.43
2 0.99
3 1.00
4 0.07
5 0.21
6 0.01
7 0.06
8 0.00
9 0.00


Merged Image and Probability Distribution
Input Image
0 0.24
1 0.98
2 0.98
3 0.91
4 0.00
5 0.49
6 0.47
7 0.19
8 0.02
9 0.20


Merged Image and Probability Distribution
Input Image
0 0.83
1 1.00
2 0.89
3 0.26
4 0.00
5 0.00
6 0.03
7 0.03
8 0.04
9 0.05


Merged Image and Probability Distribution
Input Image
0 0.85
1 0.88
2 0.97
3 0.99
4 0.00
5 0.09
6 0.00
7 0.10
8 0.01
9 0.00

You Get the point, these were the distribution for the dominance of a specific neuron used in that number



The dominance of a specific neuron in recognizing a number indicates that each neuron in the layer has learned to recognize certain features. This finding supports the hypothesis that neurons in a neural network learn to identify and respond to specific patterns in the input data.