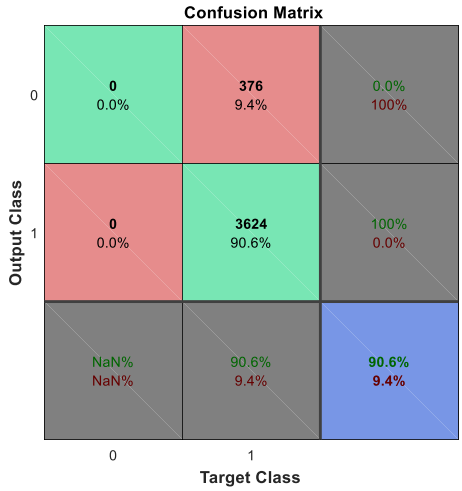
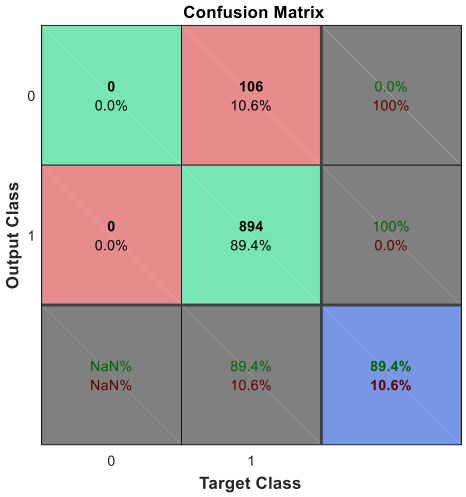
HW-3
Ximing Song
M08048930

Problem-1

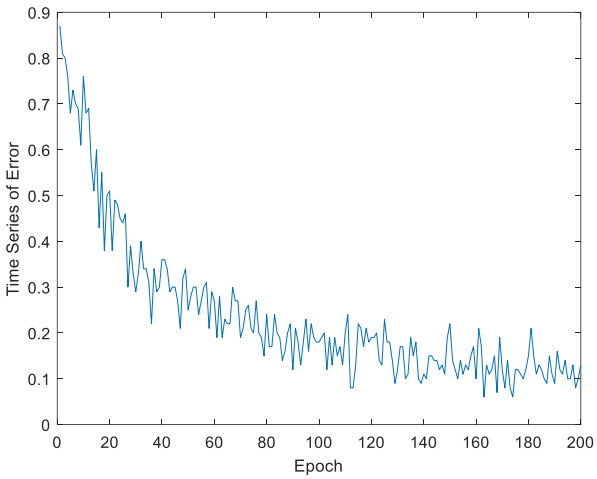
- System Description
  Number of Hidden Layers: 144
  Learning Rate: eta_o=0.01; eta_h=0.01
  Rule for choosing initial weights: random numbers from -0.5 to 0.5
  Criteria : Stop at certain epochs
- Results



Confusion Table of Training Set



Confusion Table of Testing Set



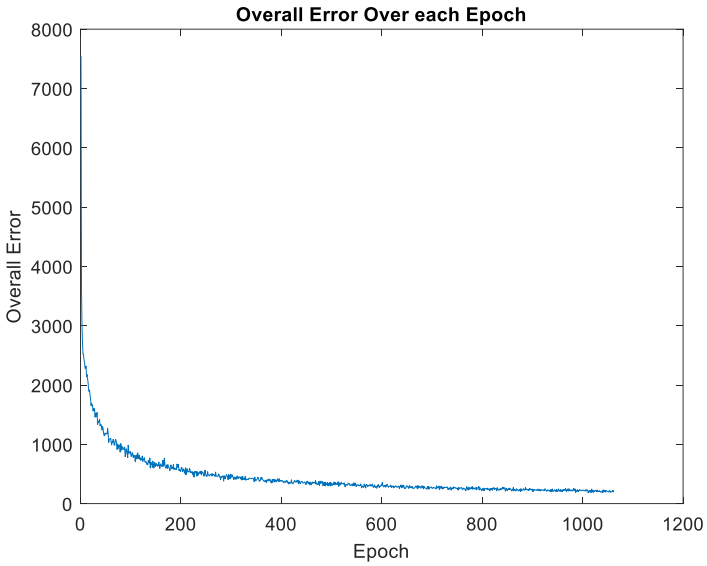
Time Series of Error

I have to say that I tried my best to find the way to plot the 10*10 confusion matrix with the data that I have already processed, but I failed to find the way. You may run the program and compare the variable of :
'Label_4000' with 'label_training_final', and compare 'Label_1000' with 'label_testing_final'. Then you will find out that the neuron network I trained got most of the correct value of each input.

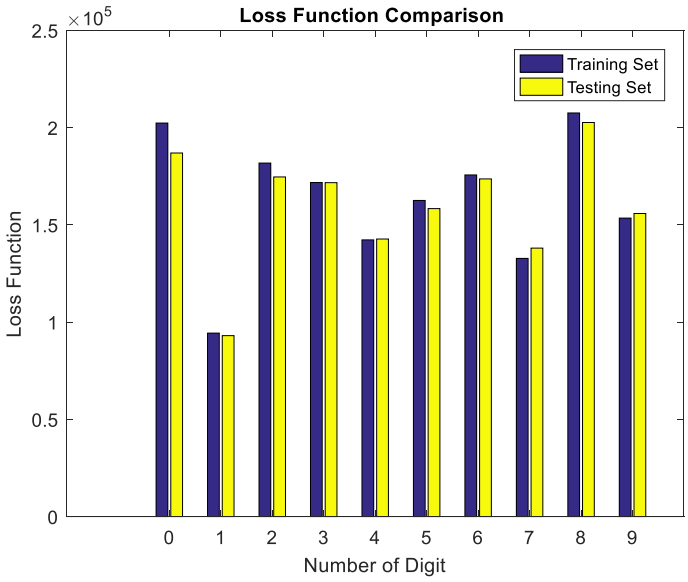
- Analysis of Results
  In my program, it used to take a long time for MATLAB on my PC to calculate the 200 epochs. The reason may be that I used too many iterated loops, making the calculation complicated exponentially. After I used the matrix calculation, the computation speed got much faster.
  The result of the program shows that as the number of epochs increase, the over all error decreases to a relatively low value. Although it still has not converged, the trend shows that it will converge if the number of epochs is large enough.

Problem-2

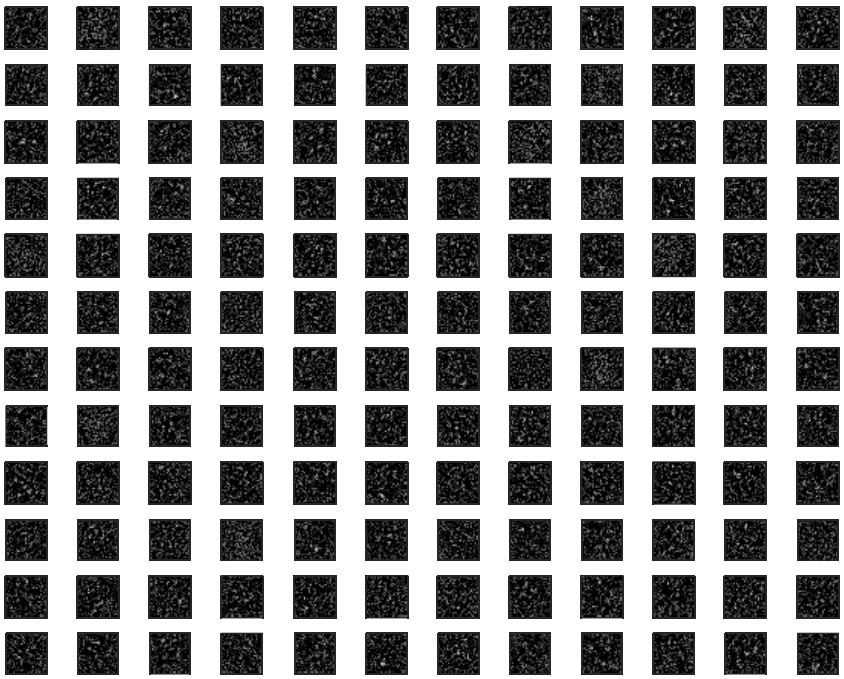
- System Description
  Learning Rate: Same as in Question-1, eta_o=eta_h=0.01
  Momentum: a=0.5
  Rule for choosing initial weights: random value from -0.5 to 0.5
  When to stop: I add a display function to observe the overall weight. Once the overall weight does not decrease so fast, I will stop the iteration.
  The number of hidden layers are still 144.
- Results



The Overall Error is shown in continuous line, so that then trend may be smoother.



- Features



Features of neurons

- Analysis of Results
  From my personal perspective, I think that the model that I trained is still not good enough, although the overall error decreases and converge to a certain small value as the number of epoch increases. If I can find a better momentum, a better learning rate, the model would be even greater than this.

  In the features of neurons given above, we can observe that there is some space in each 28*28 picture. This means that it has stronger contributions in the entire neural networks and thus the weight of that certain point is greater than others.

  From my point of view, I think the digit ‘8’ is the hardest one to reconstruct, since the error on the digit ‘8’ is the greatest among the others. The reason may be that it may influence the change of the weight in the entire neural network greater than the other digits do, which may cause greater overall error.