CSE574 Introduction to Machine Learning
Programming Assignment 2
Handwritten Digits Classication

Due Date: April 18th 2018

yuxiang liu <yliu268@buffalo.edu>
shaoming xu <shaoming@buffalo.edu>

Explanation with supporting figures of how to choose the hyper-parameter for Neural Network 26 points

The training speed and accuracy of neural network are depended on several factors such as Regularization term lambda, number of Hidden layers, number of units in each layer, and number of iteration. In this part of assignment, we mainly study the relationships among time, accuracy, Lambda, and number of hidden units given the condition that there are only 1 hidden layer and 50 iterations. Data is from the MNIST dataset with 60000 examples and 10000 test cases. In the tests, the lambda value varied from 0 to 100, and the hidden unit number varied from 240.

part 1 Lambda and Accuracy

First we study the relationships between Lambda and accuracy. Here, we study it with different hidden units varied from 5, 15, 35, and 55. We can see the trends from figure 1.1 to 1.4.

From the figures, we can see the lambda is not as bigger as better. From figure 1.1 to 1.4 we can see the tendency of accuracy rate starts to drop after the Lambda value 45. And in table 1.1, we can see the highest test accuracy of different hidden layer obviously not lies on highest lambda value.

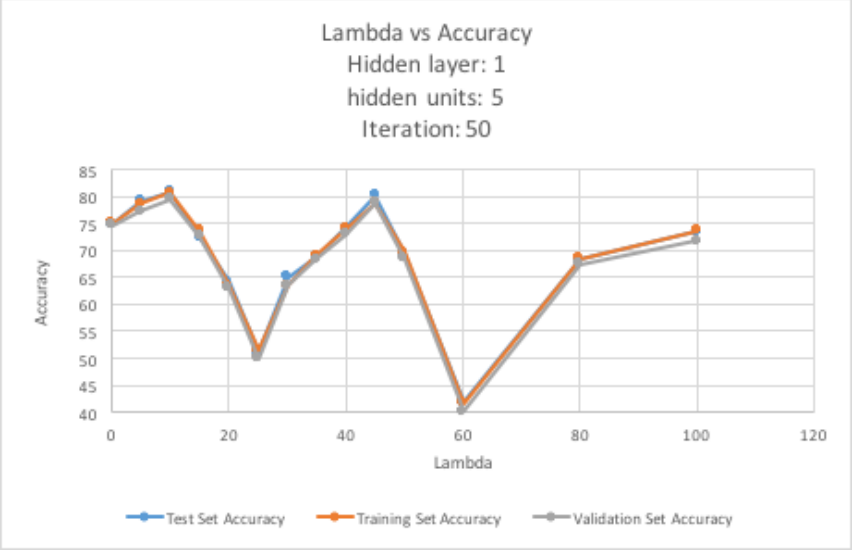


figure 1.1

hidden layer	lambda	time(second)	Training set Accuracy	Validation Set Accuracy	Test Set Accuracy
5	10	35.02328253	80.702	79.37	80.72
15	10	44.90295792	92.986	92.07	92.89
35	30	59.92752028	94.406	93.75	94.3
55	30	74.8428185	94.72	94	94.51

Table 1.1

The figure1.1 shows the lambda value varies from 0 to 100 in the case that hidden units number is 5. We can see with lower hidden units number, the accuracy tends to be lower too. And the lambda seems can influence the accuracy a lot. See figure 1.1, the line fluctuates a lot.

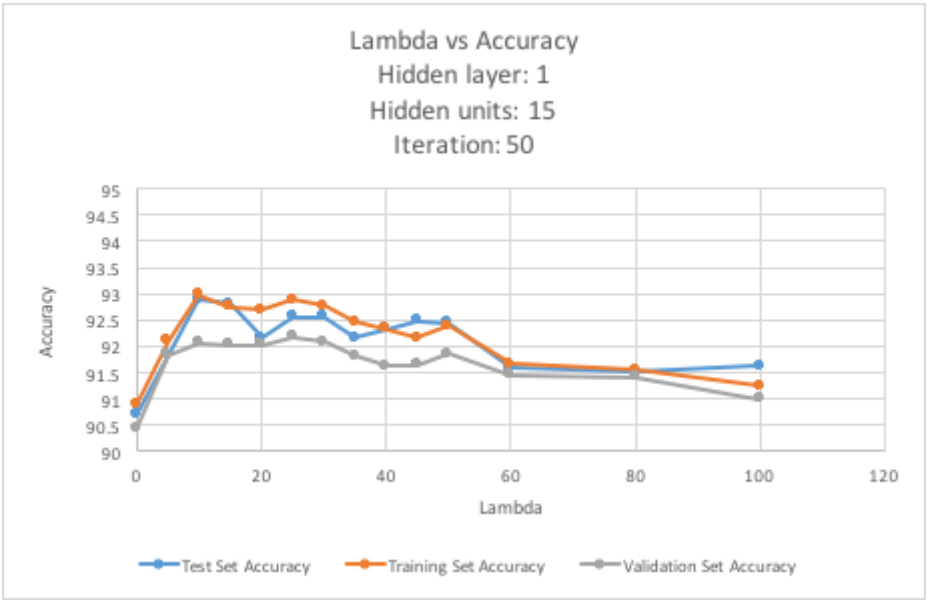


figure **1.2**

The figure1.2 shows the lambda value varies from 0 to 100 in the case that hidden units number is 15. We can see the accuracy is much higher and stable than that in figure1.1. The lambda values ranged from 5 to 50 has higher accuracy than others. And the accuracy already starts grudually dropping after lambda 10.

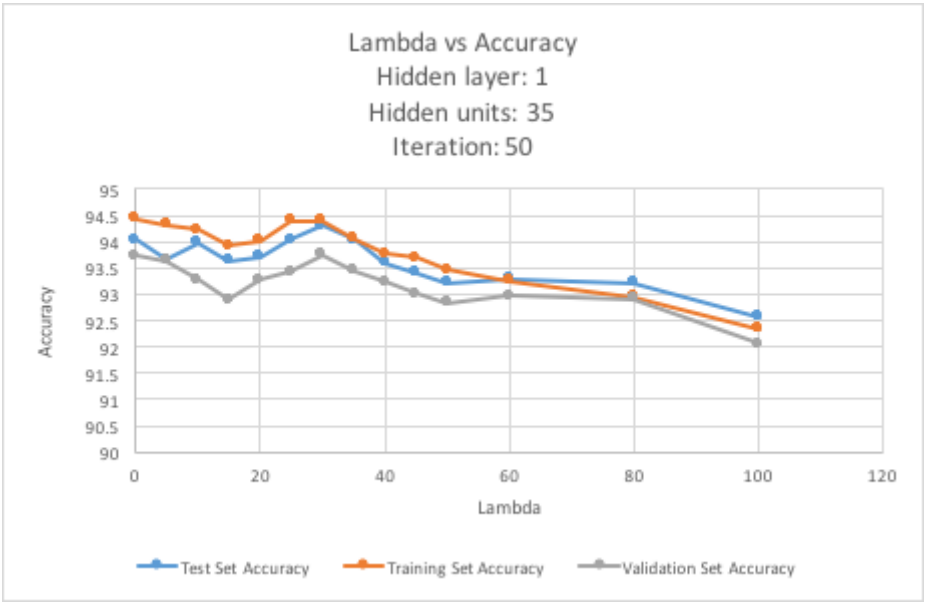
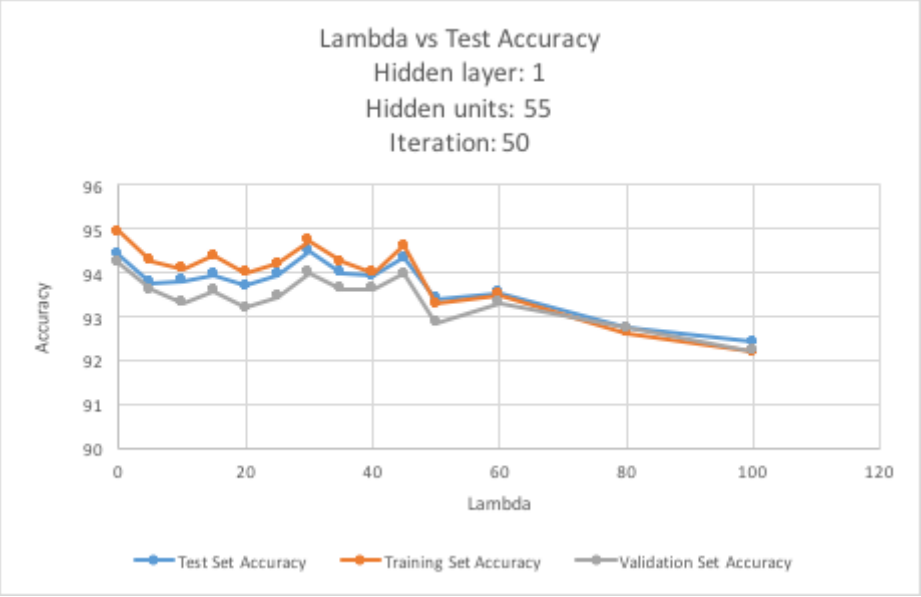


figure 1.3

The figure1.3 shows the lambda value varies from 0 to 100 in the case that hidden units number is 35. We can see the accuracy is even higher and stable than that in figure1.2. The accuracy rate is high and stable from 0 to 30 and starts gradually dropping after 30.



1.4

The figure1.4 shows the lambda value varies from 0 to 100 in the case that hidden units number is 55. We can see the accuracy is highest all figures . The accuracy rate is high and stable from 0 to 45 and starts gradually dropping after 45.

part 2 hidden units and Accuracy & time

In the figure 1.5, we can see the accuracy increases quickly when hidden units number increases from 0 to 25. But after that, the accuracy keeps stable even though the number of hidden units still keeps increasing. And in the figure 1.5, we can see the training time has a positive linear correlation with the number of hidden. So I think any lambda in the range from 25 to 55 can be a good choice in this case.

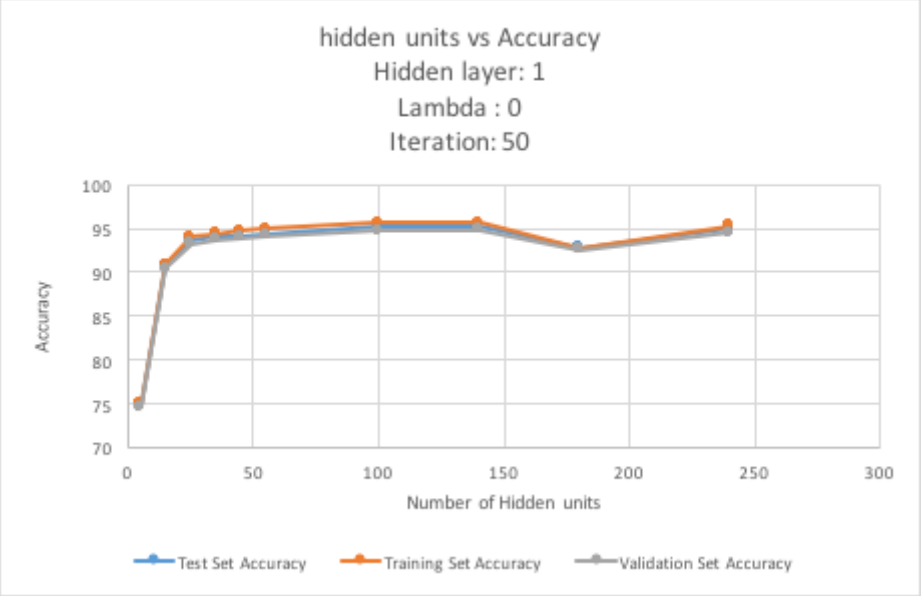


figure 1.5

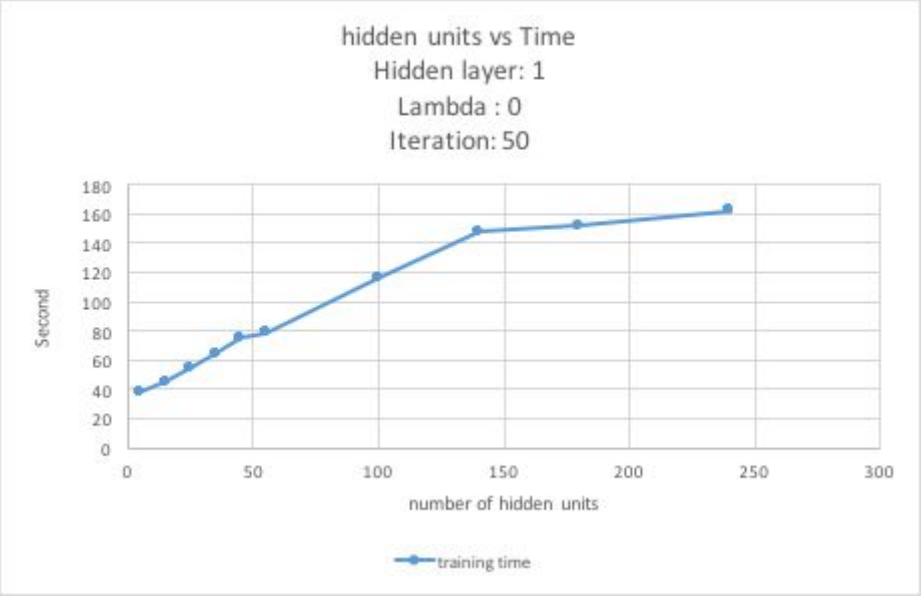


figure 1.6

Accuracy of classification method on the handwritten digits test data: 8 points

Please see Table1.1

Accuracy of classication method on the CelebA data set: 8 points

Given 1 hidden layer with 256 hidden units, lambda value to be 10, and iteration number to be 50.

1. Training set Accuracy:83.29383886255924%
2. Validation set Accuracy:82.17636022514071%
3. Test set Accuracy:83.95155185465556%

Comparison of your neural network with a deep neural network (using TensorFlow) in terms of accuracy and training time: 8 points

In figure2.1, I find the accuracy has not positive correlation with the number of Hidden layer as we expected. Even more, in our case, the accuracy has approximately negative linear correlation with the number of Hidden layer.

But the the relation between training time and the number of hidden layer had approximately positive correlation as we expected. So in this case, we should keep using the model with only 1 hidden layer.

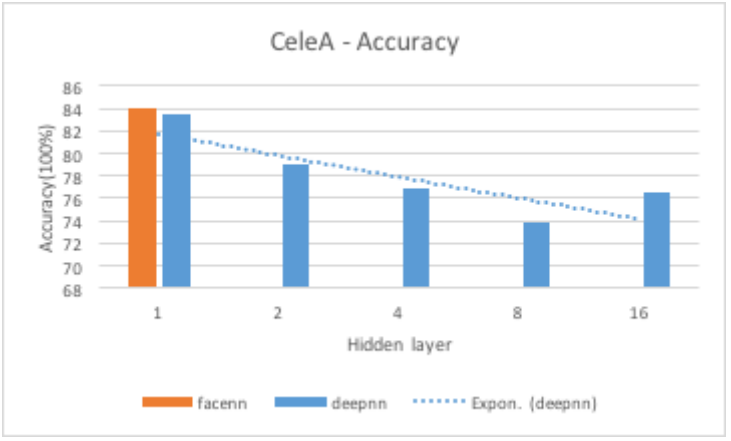


figure 2.1

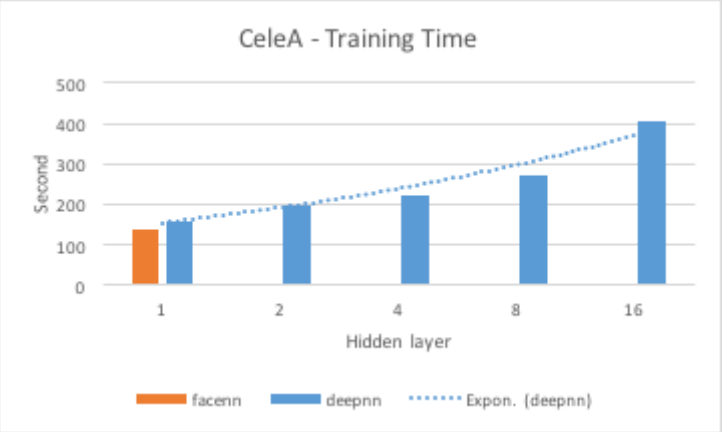


figure 2.2

Report the results from convolutional neural network in terms of accuracy and training time. 20 extra points

In this experiment we train the convolutional neural networks model on MNIST Dataset. The tensor in our CNN model flows through these layers in order: convolutional layer #1 → convolutional layer #2 → flat layer → fully connected layer(dense layer) #1 → fully connected layer(dense layer) #2. Then do the softmax on the tensor to get final predition tensor.

We aim to find the relationship between the accuracy and training time. Depending the assumption that the training time has a positive correlation with the iteration, we design this experiment based on the iterations range from 0 to 10000.

In figure 3.1, we prove our assumption is correct.

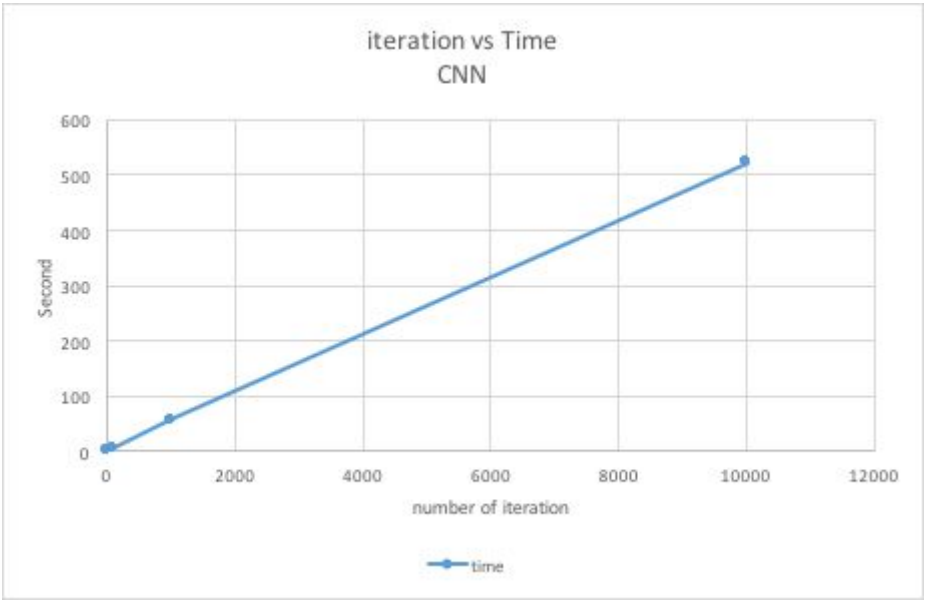


figure 3.1

In figure 3.2, we find there is a nearly logarithm relation between the accuracy and the number of iterations. In the iteration #1, the accuracy is 9.6% which is nearly the same as the chance 10% from randomly guessing a number in 10 numbers. Then after 100 iterations, the accuracy rapidly jumps to 70%. The the accuracy

increasing more and more slowly. In iteration #1000, the accuracy is already 93%. And after iteration #10000, the accuracy is 98.7.

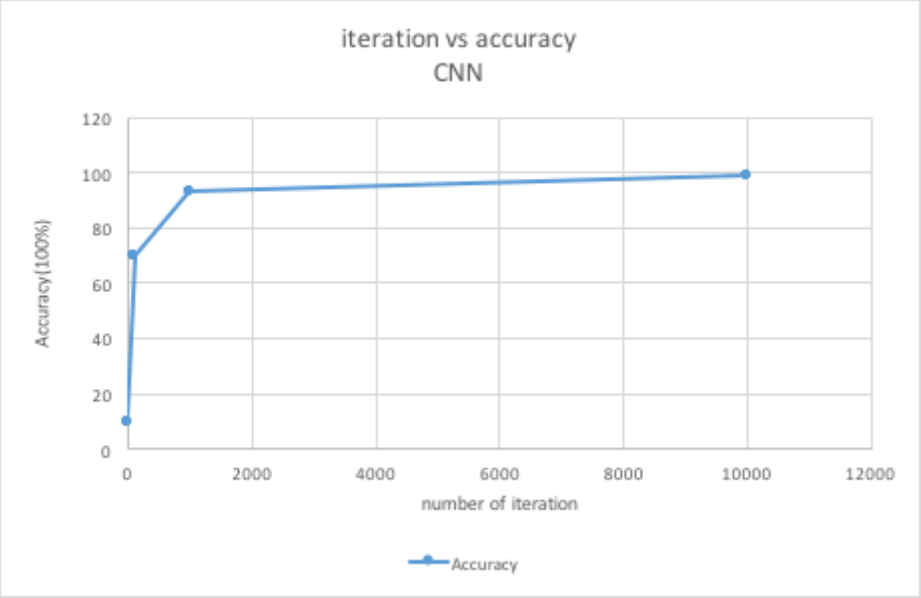


figure 3.2

iteration	Accuracy	time
0	9.3	0
1	9.6	0
100	70	6
1000	93	56
10000	98.7	523

table 3.1

To show our CNN model prediction accuracy on different digit numbers, we plot the confusion matrix and the errors example.

The figure 3.3 shows the digits been wrongly predicted by our model when the iteration number is 100 and accuracy is 70%. We can see the model performs well on most numbers other than 5. To figure out the 5 from 0 and 3 in figure 3.3 is easy task for human.

The figure 3.3 shows the confusion matrix. We find the model perform badly on number 5, 8 and 9.

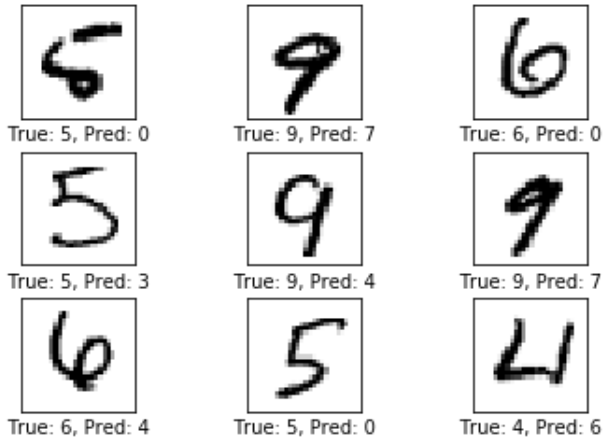


figure 3.3

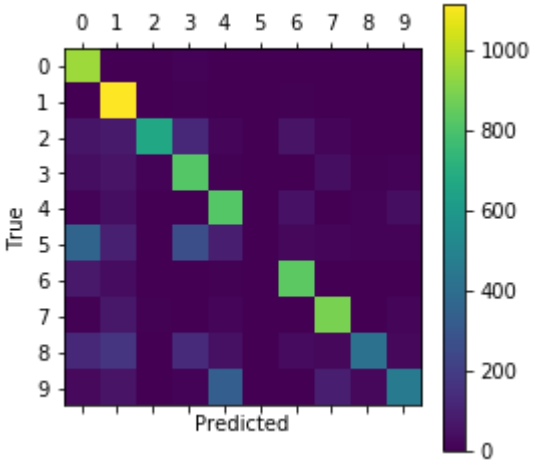


figure 3.4

The figure 3.5 and 3.6 shows the wrong digits predicted by our model and the confusion matrix when the iteration number is 10000 and accuracy is 98.7%

In figure 3.5, we can find most the digits been predicted wrongly are truly confusing even for human. And in figure 3.6 we can find the colors on the diagonal line are very bright which means most the predictions are correct.

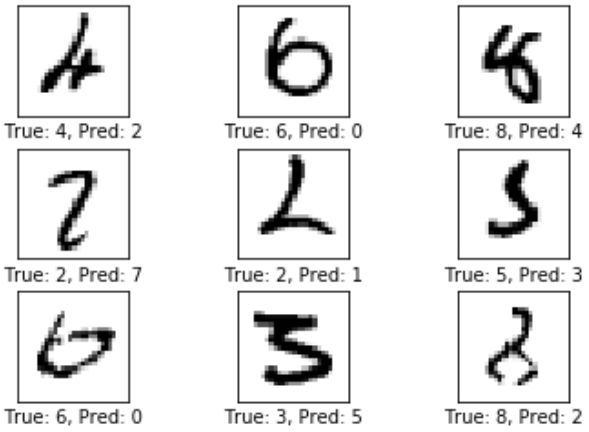


figure 3.5

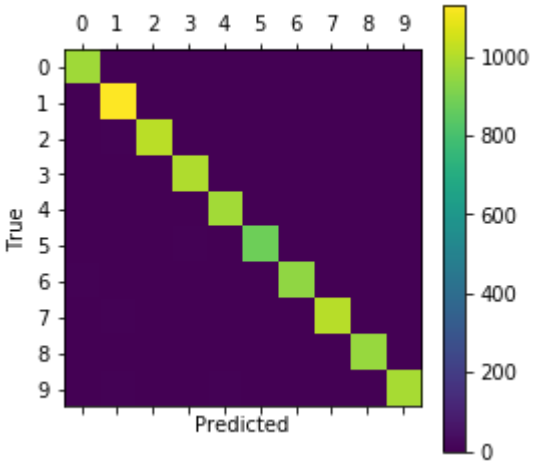


figure 3.6