CSE574 Introduction to Machine Learning
Programming Assignment 2
Handwritten Digits Classification

Haowei Zhou / haoweizh@buffalo.edu

• Abstract

In this project, I build a neural network with one hidden layer to classify the handwritten digits number and face with glasses. And I also compare the accuracy of classification with deep neural network and convolutional neural network, the influence of the number of hidden layers and hidden units.

• Feature Selection

In handwritten digits classification, we observe that some pixels in four corners of all images are all 0, so it is not the feature that we need. So I just add the value of each pixel, if the summary of this pixel is 0, we know that this pixel is not feature so I eliminate it from input.

• Choose the hyper-parameter

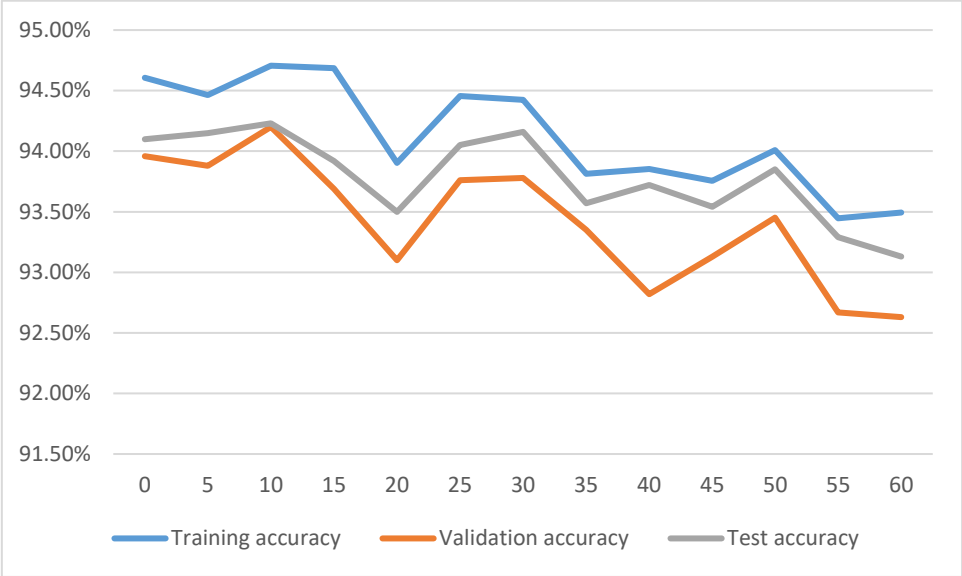
For handwritten digits classification, in order to avoid overfitting problem, we should choose appropriate hyper-parameter to magnitude of parameters in Neural Network. If we build the model without regularization, the Neural Network will be overfitting to training data set but get low accuracy to testing data set. If we build model with too much regularization, the Neural Network cannot fit the training data set and testing data set well.

We get the relationship of lambda and the accuracy of training set, validation set and test set as follow table:

Lambda	Training accuracy	Validation accuracy	Test accuracy
0	94.61%	93.96%	94.10%
5	94.46%	93.88%	94.15%
10	94.71%	94.20%	94.23%
15	94.69%	93.69%	93.92%

20	93.90%	93.10%	93.50%
25	94.46%	93.76%	94.05%
30	94.42%	93.78%	94.16%
35	93.81%	93.35%	93.57%
40	93.85%	92.82%	93.72%
45	93.76%	93.13%	93.54%
50	94.01%	93.45%	93.85%
55	93.45%	92.67%	93.29%
60	93.49%	92.63%	93.13%

And the line char looks more intuitive:

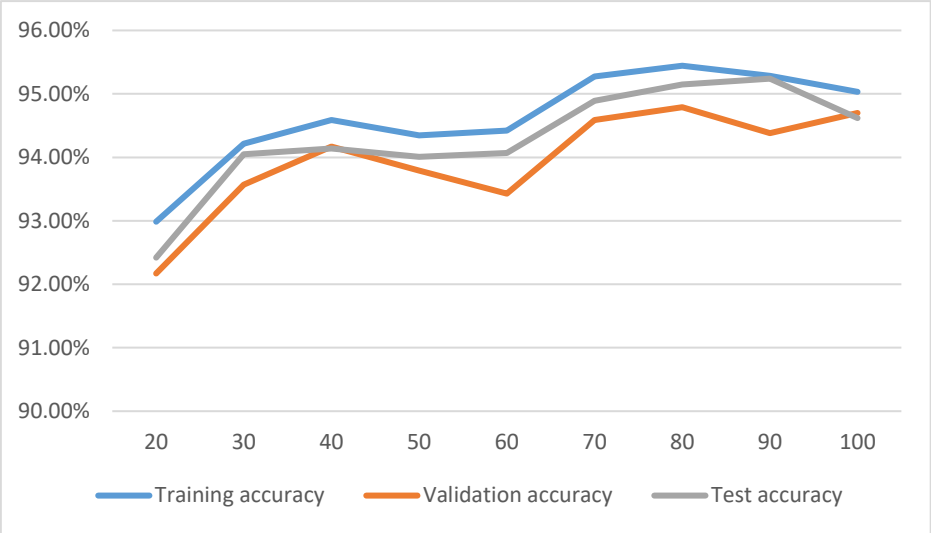


So we can get different accuracy of training set, validation set and test set when we set different lambda, when lambda equals to 10, we can get the highest accuracy of validation data, which is 94.20%.

Next I use lambda equals to 10 to get appropriate hidden units, and I get the accuracy for different hidden units as follow table:

Hidden units	Training accuracy	Validation accuracy	Test accuracy
20	92.99%	92.17%	92.42%
30	94.21%	93.57%	94.05%
40	94.59%	94.17%	94.14%
50	94.35%	93.79%	94.01%
60	94.42%	93.43%	94.07%
70	95.27%	94.59%	94.89%
80	95.44%	94.79%	95.15%
90	95.28%	94.38%	95.24%
100	95.03%	94.70%	94.62%

The line chart is below:

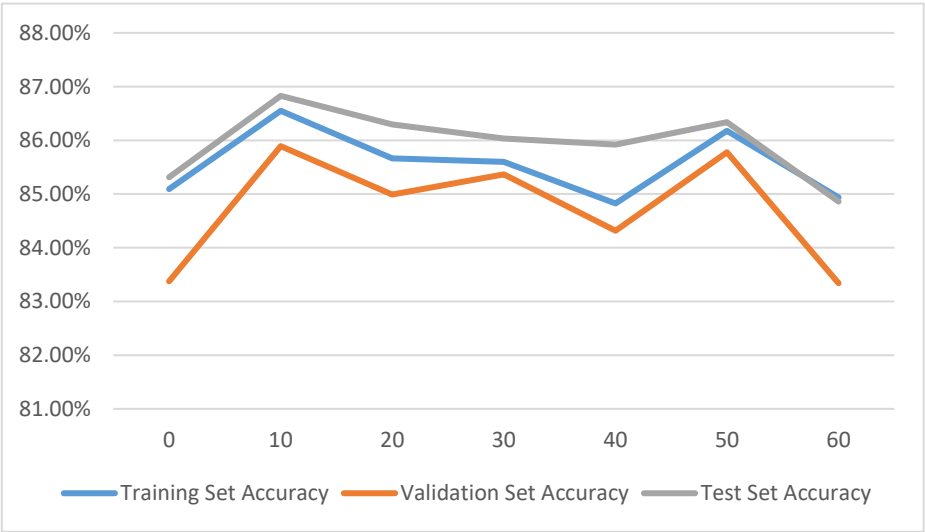


So we can conclude that when lambda equals to 10, hidden units equals to 80, we can get the highest accuracy which is 94.79%.

For face classification, the relationship of lambda and the accuracy of training set, validation set and test set as follow table:

Lambda	Training Set Accuracy	Validation Set Accuracy	Test Set Accuracy
0	85.09%	83.38%	85.31%
10	86.55%	85.89%	86.83%
20	85.66%	84.99%	86.30%
30	85.60%	85.37%	86.03%
40	84.82%	84.32%	85.92%
50	86.18%	85.78%	86.34%
60	84.94%	83.34%	84.86%

The line chart is below:



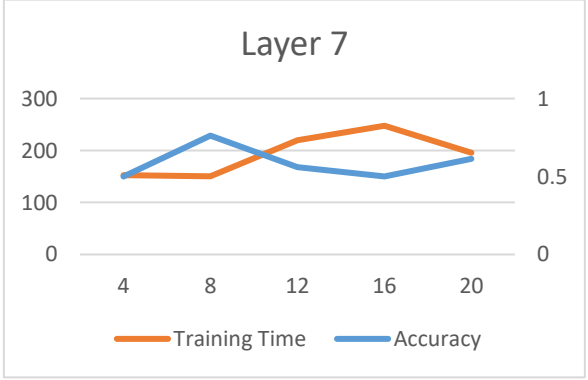
We can know when lambda equals to 10, we can get the highest accuracy of testing data, which is 85.89%.

- Compare with deep neural network

I test deep neural network with different number of layers hidden units, because too many layers and hidden units will cause program very slow, I test 3, 5 and 7 layers with 4, 8, 12, 16, 20 hidden units, and the accuracy and training time result is shown below.

Depth	Hidden units	Accuracy	Training Time
3	4	0.834974	145.2698967
3	8	0.704391	139.4176445
3	12	0.789175	160.6371744
3	16	0.79788	164.2309635
3	20	0.79296	180.2088568
5	4	0.817941	139.9697139
5	8	0.571915	145.4286532
5	12	0.526495	163.3620083
5	16	0.660106	171.7236762
5	20	0.619985	187.5398865
7	4	0.5	152.5897088
7	8	0.762301	150.4012358
7	12	0.560939	220.0347095
7	16	0.500757	247.6873846
7	20	0.613172	196.0465746

The line chart for layer 3, 5 and 7 is below:



We can see when the number of layers increases, the average accuracy is decreasing because of overfitting, but the training time is increasing. So we can know it is false that the more layers, the higher accuracy. Also, when the number of units increases, the accuracy fluctuates, when neural network has 3 layers, we get highest accuracy when hidden units is 4, when neural network has 5 layers, we get highest accuracy when hidden units is 4, when neural network has 7 layers, we get highest accuracy when hidden units is 8.

- Compare with convolutional neural network

The convolutional neural network's testing result is below:

```
Accuracy on Test-Set: 10.2% (1023 / 10000)
Optimization Iteration:      1, Training Accuracy:  10.9%
Time usage: 0:00:00
Accuracy on Test-Set: 10.8% (1082 / 10000)
Time usage: 0:00:07
Accuracy on Test-Set: 58.6% (5862 / 10000)
Optimization Iteration:    101, Training Accuracy:  64.1%
Optimization Iteration:    201, Training Accuracy:  70.3%
Optimization Iteration:    301, Training Accuracy:  73.4%
Optimization Iteration:    401, Training Accuracy:  82.8%
Optimization Iteration:    501, Training Accuracy:  87.5%
Optimization Iteration:    601, Training Accuracy:  93.8%
Optimization Iteration:    701, Training Accuracy:  92.2%
Optimization Iteration:    801, Training Accuracy:  92.2%
Optimization Iteration:    901, Training Accuracy:  90.6%
Time usage: 0:01:05
Accuracy on Test-Set: 92.3% (9235 / 10000)
Optimization Iteration:   1001, Training Accuracy:  92.2%
Optimization Iteration:   1101, Training Accuracy:  93.8%
Optimization Iteration:   1201, Training Accuracy:  98.4%
Optimization Iteration:   1301, Training Accuracy:  93.8%
Optimization Iteration:   1401, Training Accuracy:  98.4%
Optimization Iteration:   1501, Training Accuracy:  95.3%
Optimization Iteration:   1601, Training Accuracy: 100.0%
Optimization Iteration:   1701, Training Accuracy:  95.3%
Optimization Iteration:   1801, Training Accuracy:  95.3%
Optimization Iteration:   1901, Training Accuracy:  89.1%
Optimization Iteration:   2001, Training Accuracy:  98.4%
Optimization Iteration:   2101, Training Accuracy:  98.4%
Optimization Iteration:   2201, Training Accuracy: 100.0%
Optimization Iteration:   2301, Training Accuracy:  98.4%
Optimization Iteration:   2401, Training Accuracy:  92.2%
Optimization Iteration:   2501, Training Accuracy:  93.8%
```

Optimization Iteration: 2601, Training Accuracy: 96.9%
Optimization Iteration: 2701, Training Accuracy: 95.3%
Optimization Iteration: 2801, Training Accuracy: 96.9%
Optimization Iteration: 2901, Training Accuracy: 95.3%
Optimization Iteration: 3001, Training Accuracy: 96.9%
Optimization Iteration: 3101, Training Accuracy: 96.9%
Optimization Iteration: 3201, Training Accuracy: 100.0%
Optimization Iteration: 3301, Training Accuracy: 98.4%
Optimization Iteration: 3401, Training Accuracy: 98.4%
Optimization Iteration: 3501, Training Accuracy: 100.0%
Optimization Iteration: 3601, Training Accuracy: 95.3%
Optimization Iteration: 3701, Training Accuracy: 96.9%
Optimization Iteration: 3801, Training Accuracy: 100.0%
Optimization Iteration: 3901, Training Accuracy: 95.3%
Optimization Iteration: 4001, Training Accuracy: 98.4%
Optimization Iteration: 4101, Training Accuracy: 98.4%
Optimization Iteration: 4201, Training Accuracy: 89.1%
Optimization Iteration: 4301, Training Accuracy: 100.0%
Optimization Iteration: 4401, Training Accuracy: 100.0%
Optimization Iteration: 4501, Training Accuracy: 93.8%
Optimization Iteration: 4601, Training Accuracy: 100.0%
Optimization Iteration: 4701, Training Accuracy: 96.9%
Optimization Iteration: 4801, Training Accuracy: 98.4%
Optimization Iteration: 4901, Training Accuracy: 98.4%
Optimization Iteration: 5001, Training Accuracy: 95.3%
Optimization Iteration: 5101, Training Accuracy: 92.2%
Optimization Iteration: 5201, Training Accuracy: 98.4%
Optimization Iteration: 5301, Training Accuracy: 100.0%
Optimization Iteration: 5401, Training Accuracy: 100.0%
Optimization Iteration: 5501, Training Accuracy: 96.9%
Optimization Iteration: 5601, Training Accuracy: 98.4%
Optimization Iteration: 5701, Training Accuracy: 98.4%
Optimization Iteration: 5801, Training Accuracy: 100.0%
Optimization Iteration: 5901, Training Accuracy: 96.9%
Optimization Iteration: 6001, Training Accuracy: 98.4%
Optimization Iteration: 6101, Training Accuracy: 100.0%
Optimization Iteration: 6201, Training Accuracy: 96.9%
Optimization Iteration: 6301, Training Accuracy: 98.4%
Optimization Iteration: 6401, Training Accuracy: 98.4%
Optimization Iteration: 6501, Training Accuracy: 95.3%
Optimization Iteration: 6601, Training Accuracy: 98.4%
Optimization Iteration: 6701, Training Accuracy: 100.0%
Optimization Iteration: 6801, Training Accuracy: 100.0%
Optimization Iteration: 6901, Training Accuracy: 100.0%

```
Optimization Iteration:   7001, Training Accuracy: 100.0%
Optimization Iteration:   7101, Training Accuracy: 100.0%
Optimization Iteration:   7201, Training Accuracy:  96.9%
Optimization Iteration:   7301, Training Accuracy:  96.9%
Optimization Iteration:   7401, Training Accuracy: 100.0%
Optimization Iteration:   7501, Training Accuracy:  96.9%
Optimization Iteration:   7601, Training Accuracy: 100.0%
Optimization Iteration:   7701, Training Accuracy: 100.0%
Optimization Iteration:   7801, Training Accuracy:  98.4%
Optimization Iteration:   7901, Training Accuracy:  96.9%
Optimization Iteration:   8001, Training Accuracy:  98.4%
Optimization Iteration:   8101, Training Accuracy: 100.0%
Optimization Iteration:   8201, Training Accuracy: 100.0%
Optimization Iteration:   8301, Training Accuracy: 100.0%
Optimization Iteration:   8401, Training Accuracy: 100.0%
Optimization Iteration:   8501, Training Accuracy: 100.0%
Optimization Iteration:   8601, Training Accuracy:  98.4%
Optimization Iteration:   8701, Training Accuracy: 100.0%
Optimization Iteration:   8801, Training Accuracy:  98.4%
Optimization Iteration:   8901, Training Accuracy: 100.0%
Optimization Iteration:   9001, Training Accuracy: 100.0%
Optimization Iteration:   9101, Training Accuracy: 100.0%
Optimization Iteration:   9201, Training Accuracy:  98.4%
Optimization Iteration:   9301, Training Accuracy: 100.0%
Optimization Iteration:   9401, Training Accuracy: 100.0%
Optimization Iteration:   9501, Training Accuracy:  96.9%
Optimization Iteration:   9601, Training Accuracy: 100.0%
Optimization Iteration:   9701, Training Accuracy:  98.4%
Optimization Iteration:   9801, Training Accuracy:  98.4%
Optimization Iteration:   9901, Training Accuracy:  98.4%
Time usage: 0:10:44
Accuracy on Test-Set: 98.6% (9864 / 10000)
```

We can see when optimization iteration increases, the training time is increasing as well, but at the same time, the training accuracy is increasing and becomes more stable close to 100%. So the CNN's performance is better than deep neural network in this project.

• Conclusion

Through comparing my neural network with deep neural network and convolutional neural network, we can know CNN has highest performance in image reorganization. Deep neural network's performance is not as good as my handwritten neural network because too many hidden layers cause overfitting, so it is false that the more layers, the better accuracy of classification.