EE5434 Machine Learning Homework 4 Report

Digit Classification based on ANN

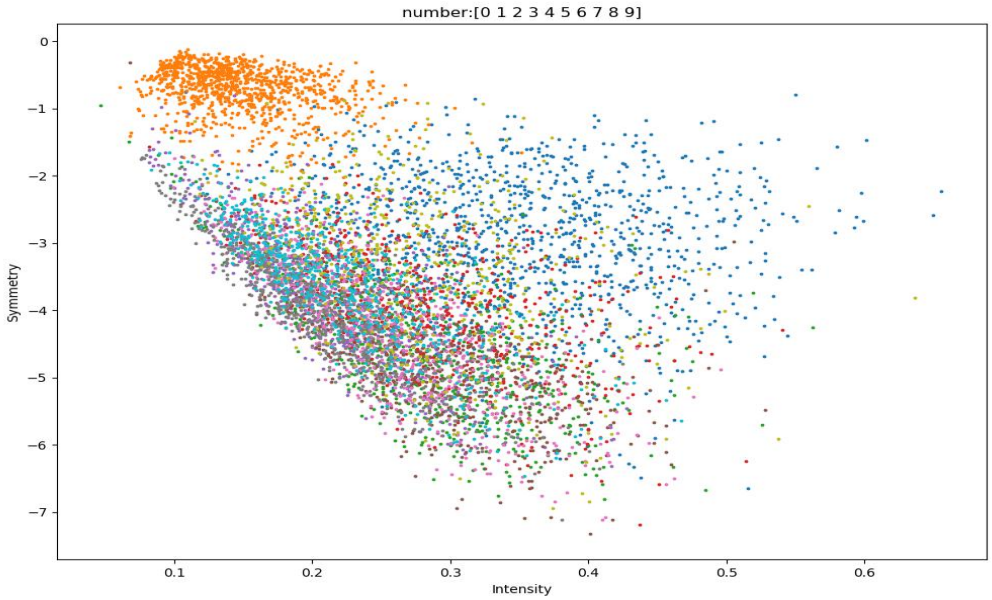
Name: ZHOU Ziyang

Student number: 55921210

Department : Electronic Engineering

Task 1:

Plot scatters of two features for 10 labels:

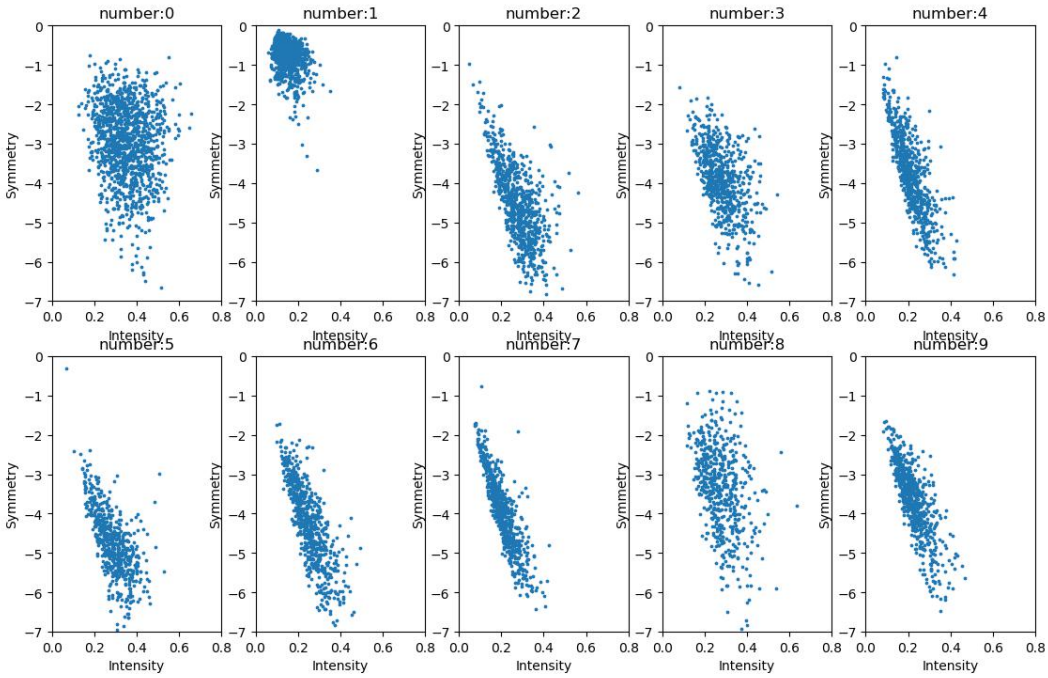


Analysis:

I plot all data points of 10 digits on a two-dimensional plane. The X-axis is intensity value and the Y-axis is symmetry value. Data points of 10 digits are represented in different colors.

We can see from this plot, data points of 10 digits mix together on a two-dimensional plane. It is almost impossible to draw one or more lines to classify digits with 10 labels if we use only two features.

Plot data points of 10 digits separately:



Analysis:

I plot data points of 10 digits on 2-dimensional plane separately as the figure above. The X-axis is intensity value and the Y-axis is symmetry value.

We can see from these figure that, only data points of digit 1 are different with other data points obviously. The distributions of digit 2 to 9 are very similar in shape. And the distribution of the digit 0 covers almost the entire two-dimensional plane. So that it is almost impossible to classify them by using only 2 features.

Conclusion:

Above all, if we have only two features, intensity and symmetry, data points of 10 digits mix together. So it is almost impossible to classify all digits with 10 labels. But according to figures above, it is still possible to distinguish digit 1 with other digits. If our target is to classify all digits, more features are necessary.

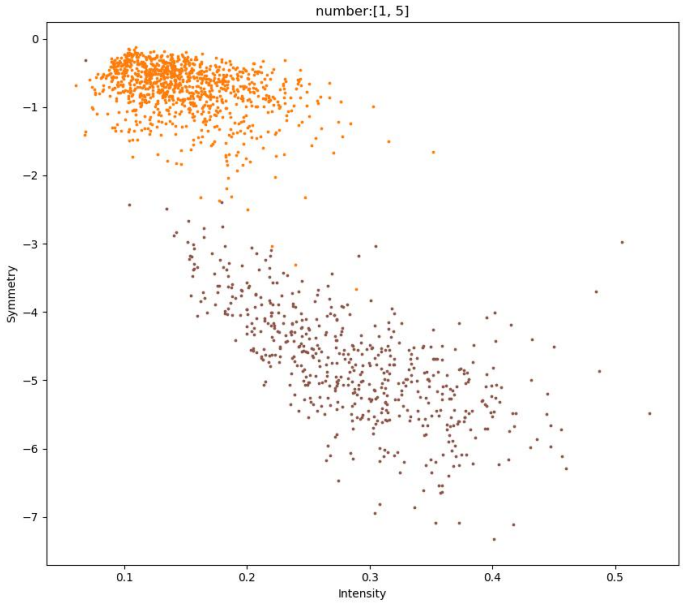
Task 2:

Task goal:

Classify digit 1 and digit 5 using two features data based on ANN. 3-fold cross-verification.

Method:

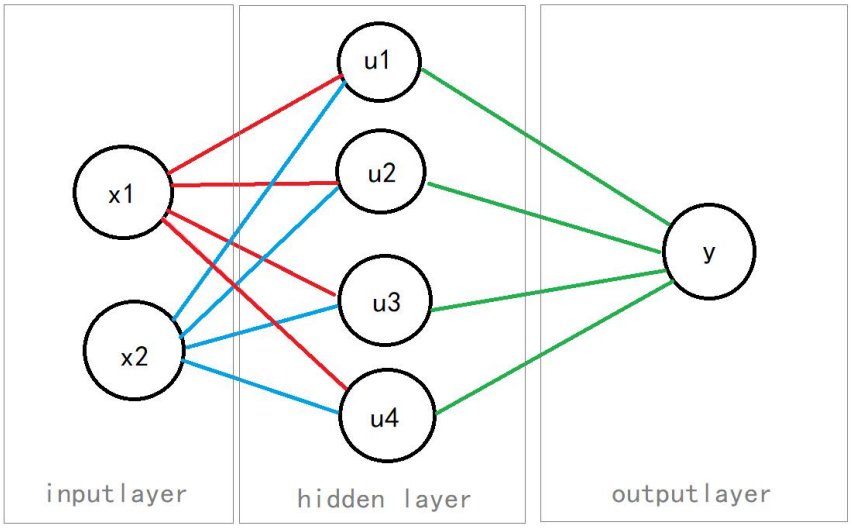
1. Plot data points on two-dimensional plane. The X-axis is intensity value and the Y-axis is symmetry value.



2. Build ANN model structure:

From the figure above, we can know the boundary between the two kinds of data point is very clear in two-dimensional plane. So it is not difficult to classify digit 1 and digit 5.

Here I use a single hidden layer ANN to classify digit 1 and digit 5. There are a number of units in the hidden layer. According to the input and output data, there are 2 units in input layer(two features) and 1 unit in output layer,which is the label of digit(1 or 5). The ANN structure is as following: (here I use hidden layer with 4 units to draw this figure)



3. Train ANN model with a number of hidden layer units. Use 3-fold cross-verification to minimize the errors.

Here I try 2-7 units of hidden layer to find the best model. After I got the trained ANN model, I use the whole train data to examine the prediction accuracy of the model. Here is the examining accuracy for different hidden layer units from 2 to 7:

```
ANN Model: single hidden layer with 2 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9904030710172744
fold 2 score :  0.9980769230769231
fold 3 score :  0.9942307692307693

train data accuracy with 2 units : 99.35939 %

------------------------------------------------------------
ANN Model: single hidden layer with 3 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9904030710172744
fold 2 score :  0.9980769230769231
fold 3 score :  0.9942307692307693

train data accuracy with 3 units : 99.42345 %

------------------------------------------------------------
ANN Model: single hidden layer with 4 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9923224568138196
fold 2 score :  0.9961538461538462
fold 3 score :  0.9942307692307693

train data accuracy with 4 units : 99.55157 %

------------------------------------------------------------
```

```
------------------------------------------------------------
ANN Model: single hidden layer with 5 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9961612284069098
fold 2 score :  0.9961538461538462
fold 3 score :  0.9942307692307693

train data accuracy with 5 units : 99.55157 %

------------------------------------------------------------
ANN Model: single hidden layer with 6 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9942418426103646
fold 2 score :  0.9980769230769231
fold 3 score :  0.9942307692307693

train data accuracy with 6 units : 99.61563 %

------------------------------------------------------------
ANN Model: single hidden layer with 7 units:

we use 3-fold crose-validation to get three model:
fold 1 score :  0.9961612284069098
fold 2 score :  0.9961538461538462
fold 3 score :  0.9942307692307693

train data accuracy with 7 units : 99.55157 %

------------------------------------------------------------
```

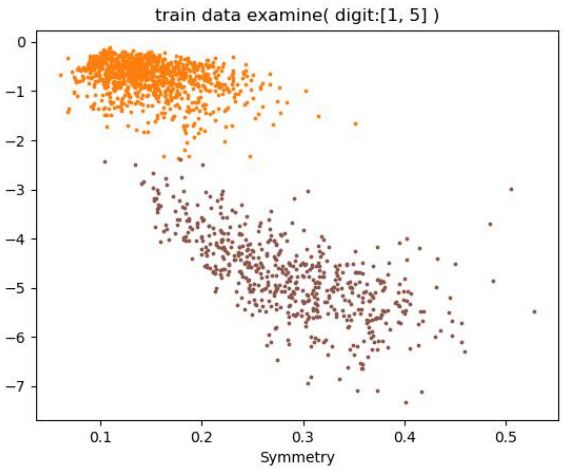
From these examining accuracy, we can know that, examining accuracy keep stable when the number of hidden layer units is larger than or equal to 4. But model with 6 units have the best performance.

4. Model examining.

Save the model parameters I got above. Then I use this model to classify the original train data. The model parameters , examining accuracy and classify result is as figures below:

```
whole train data accuracy =  99.61563 %
whole train data error =  0.00384

sturcture of the hidden layer:
( 6 units in 1 hidden layer)
```

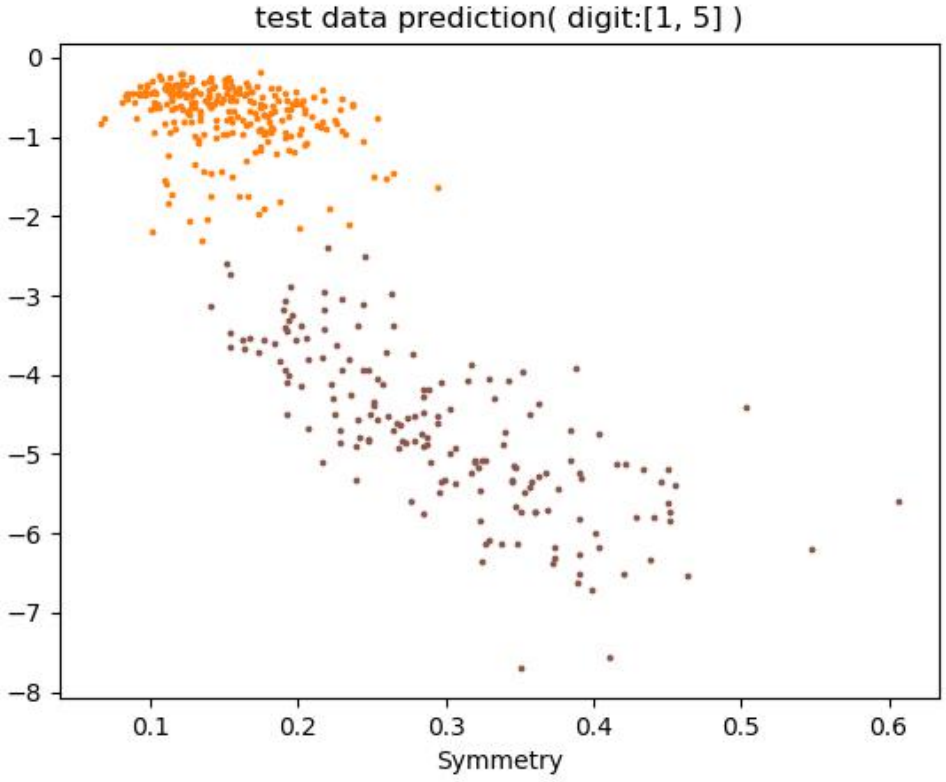


5. Use test data to test ANN model with 6 hidden layer units.

Here are classification accuracy , results and plot:

```
test data prediction accuracy = 98.113208 %

Randomly select 10 data to display and compare the recognition results
original digit:
     1   5   5   1   1   5   1   5   1   1
classification results:
     1   5   5   1   1   5   1   5   1   1
```



Task 3

Target: Classify digit 1 and digit 5 using raw data with 256 grayscale features based on ANN.

Method:

1. Build ANN model with 2 hidden layers.

There are 256 units in input layer (256 grayscale features), 1 unit in output layer (output label) and two hidden layers. Here I train two ANN model with different hidden layer structures. One has 6 units in the first hidden layer and 2 units in the second hidden layer. Another has 3 units in the first hidden layer and 2 units in the second hidden layer.

Use 3-fold cross-verification to get the best model.

Here are train data examining results as below:

```
ANN Model: 2 hidden layer with 6 units and 2 units respectively:

we use 3-fold crose-validation to get three model:
model 1 score :  0.9942418426103646
model 2 score :  0.9980769230769231
model 3 score :  0.9980769230769231
model saved successfully!

ANN Model: 2 hidden layers with 6 units and 2 units:
whole train data examining accuracy =  99.871877 %

----------------------------------------------------------------
ANN Model: 2 hidden layer with 3 units and 2 units respectively:

we use 3-fold crose-validation to get three model:
model 1 score :  0.9980806142034548
model 2 score :  0.9961538461538462
model 3 score :  0.9961538461538462
model saved successfully!

ANN Model: 2 hidden layers with 3 units and 2 units:
whole train data examining accuracy =  100.0 %

----------------------------------------------------------------
```

And here is test data prediction accuracy as following:

```
ANN Model: 2 hidden layer with 6 units and 2 units:

test data prediction accuracy = 98.349057 %

----------------------------------------------------------------

ANN Model: 2 hidden layer with 3 units and 2 units:

test data prediction accuracy = 99.056604 %
```

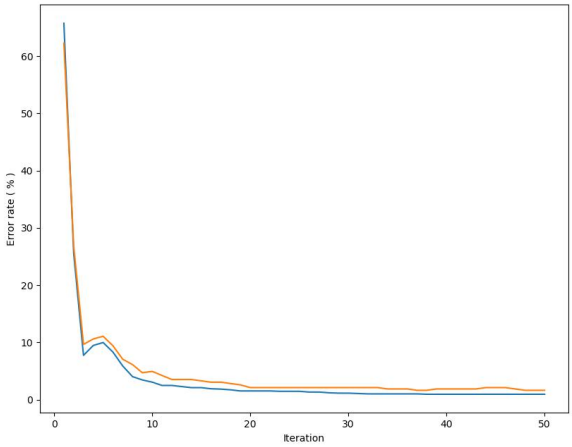
From the results above, we can know ANN model with structure [256,3,2,1] has a better performance in 2 digit classification(1 and 5) than that with structure [256,6,2,1]. The results of training data verification of two are similar. However, the ANN model with the structure of

[256,3,2,1] has better classification accuracy to test raw data.

In this case, we can know that more units of hidden layer, which means more complex of ANN model structure, is not always better. Too complex the model may lead to over-fitting.

2. plot the change of the in-sample error and test-set error for each iteration

In the plot below, the blue line is in-sample error and the red line is test-set error. Here I use classification error rate as Y-axis and the number of iteration as X-axis.



As can be seen from the figure, when the number of iterations is from 3 to 5, the error rate increases obviously. Later, as the number of iterations increased, the error rate of the in-sample data and the test-set data decreased. Finally, when the number of iterations exceeded 30, the error rate of in-sample data remained almost unchanged, but the error rate of test-set fluctuated up and down without convergence. It means, as the increasing of iterations, over-fitting would happen, which could cause the increase of test-set error rate. Over-training would lead to over-fitting.

Task 4

Task goal:

Apply neural network for classification for all 10 digits, using the raw features as input.

Method:

1. Build ANN model with 3 hidden layers.

There are 256 units in input layer (256 grayscale features), 1 unit in output layer (output label) and two hidden layers. Here I train two ANN model with different hidden layer structures. One has 60 units in the first hidden layer, 30 units in the second hidden layer and 10 units in the third hidden layer. The structure is as [256, 60, 30, 10, 1].

Use 3-fold cross-verification to get the best model.

Here I use four different ANN structure:

[256, 30, 10, 1] ; [256, 60, 20, 1] ; [256, 60, 30,10, 1] ; [256, 120, 40, 10, 1]
The in-sample training results are as following:

```
ANN Model: hidden layer with (30, 10) structure:

we use 3-fold crose-validation to get three models:
fold 1 score :  0.9469354175236528
fold 2 score :  0.9444444444444444
fold 3 score :  0.9469135802469136
3-fold cross-validation mean error is: 0.053902

train data accuracy with hidden layer  (30, 10) :  97.256892 %

--------------------------------------------------------------------------
ANN Model: hidden layer with (60, 20) structure:

we use 3-fold crose-validation to get three models:
fold 1 score :  0.9539284245166598
fold 2 score :  0.939917695473251
fold 3 score :  0.9534979423868313
3-fold cross-validation mean error is: 0.050885

train data accuracy with hidden layer  (60, 20) :  98.038678 %

--------------------------------------------------------------------------
ANN Model: hidden layer with (60, 30, 10) structure:

we use 3-fold crose-validation to get three models:
fold 1 score :  0.9531057178116001
fold 2 score :  0.942798353909465
fold 3 score :  0.9465020576131687
3-fold cross-validation mean error is: 0.052531

train data accuracy with hidden layer  (60, 30, 10) :  98.258126 %

--------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------
ANN Model: hidden layer with (120, 40, 10) structure:

we use 3-fold crose-validation to get three models:
fold 1 score :  0.9584533113944879
fold 2 score :  0.948559670781893
fold 3 score :  0.9526748971193416
3-fold cross-validation mean error is: 0.046771

train data accuracy with hidden layer  (120, 40, 10) :  99.204499 %

--------------------------------------------------------------------------
```

```
model saved successfully!

final model structure: [ 256, (120, 40, 10) , 1 ]

whole train data accuracy =  99.2045 %
whole train data error =  0.00796
```

Then I use these four models to predict test data. Here is prediction accuracy:

```
test data prediction accuracy :

ANN structure: [ 256, (30, 10) , 1 ]
prediction accuracy =  91.629297 %


ANN structure: [ 256, (60, 20) , 1 ]
prediction accuracy =  91.529646 %


ANN structure: [ 256, (60, 30, 10) , 1 ]
prediction accuracy =  91.629297 %


ANN structure: [ 256, (120, 40, 10) , 1 ]
prediction accuracy =  92.227205 %
```

Analysis:

From these experiments, ANN model with more hidden layers and more neuron units has a better performance both in training data examining and test data prediction. I did not try ANN model with more complex structure because it spend too much time in training process.

Besides, ANN model with more complex structure also means more detailed learning to training data, which could also lead to over fitting.

Conclusion:

In task 1, data points with two features of 10 digits are plotted on a 2-dimensional plane in order to explore how to classification digits by using only two features, intensity value and symmetry value. But the result shows it is an almost impossible task.

In task 2, our target is to distinguish digit 1 and digit 5 using only 2 features, intensity value and symmetry value. I use ANN model with one hidden layer and try different number of neuron units. Experiment results show the best number of neuron units is 6, which can achieve recognition accuracy of 98.1132%. Therefore, digit 1 and 5 with 2 features can be distinguished by single hidden layer ANN model excellently.

In task 3, we use raw data of 256 greyscale features to distinguish digit 1 and digit 5. I try two ANN models with structure [256,6,3,1] and [256,3,2,1]. The experiment results show structure [256, 3, 2, 1] have a better classification performance. By measuring the error of each iteration, the reason that the classification performance of the structure[256,6,3,1] is not ideal may be that there is a certain degree of over fitting.

In task 4, I use the same method as task 3 to classify digit 0 to 9, but this time I make the neural network model more complex. I use four structures of [256, 30, 10, 1] , [256, 60, 20, 1] , [256, 60, 30, 10, 1] , [256, 120, 40, 10, 1] to do this. The training and test results shows, more complex of ANN model structure, better the classification accuracy, but worse the time cost. At present, the best ANN model structure is [256, 120, 40, 10, 1] which can achieve recognition accuracy of 92.2272%.

Above all, in the subject of digit recognition, digital data with only two features is difficult to recognize. Using the digit data with 256 greyscale features is more effective. ANN model with multiple hidden layers and a large number of neuron units has a good performance in digit classification, but the time cost of training is also huge. This project explored the influence of the number of hidden layers and neurons on the digit recognition accuracy. The best digital recognition model needs further experiments.