MSBD5002 Assignment 3 Report

The following report will be divided into two major parts.

1. Binary Classification
2. Multi Class Classification

And for each major part, I will introduce 4 different sets of parameters and result. The reasons that there are 4 different sets is because 4 sets are enough for observation. For both classifier, cross entropy loss, and stochastic gradient descent optimizer are being used (As described in specification).

**Binary Classification**

The following 4 set of parameters have been applied.

1. Learning Rate 0.1, Batch Size 1, Epoch 50
2. Learning Rate 0.1, Batch Size 100, Epoch 1000
3. Learning Rate 0.00001, Batch Size 100, Epoch 1000
4. Learning Rate 0.0001, Batch Size 100, Epoch 2000

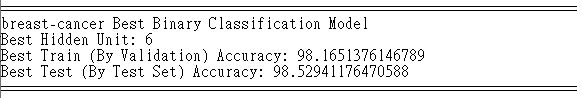
In the following report, I will simply use LR for learning rate, B for batch size, and E for numbers of epoch. For the best model, if the accuracy is the same, smaller hidden unit will be considered.

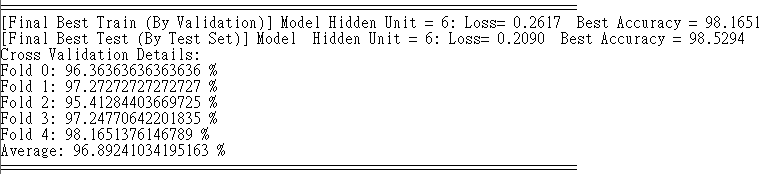
**LR 0.1, B 1, E 50**

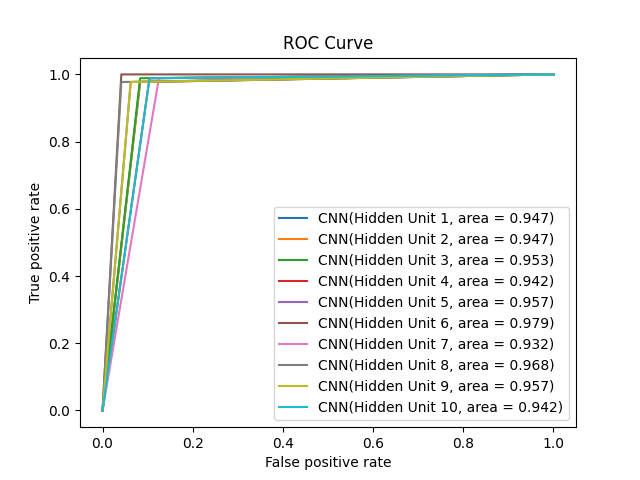
For the first set, it mainly for program correctness testing and debugging. Therefore, the data is just for reference.

breast-cancer dataset (LR 0.1, B 1, E 50)

Program output, we found the best is hidden unit 6 with 98.5% accuracy.

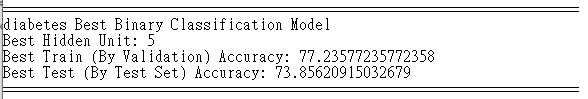


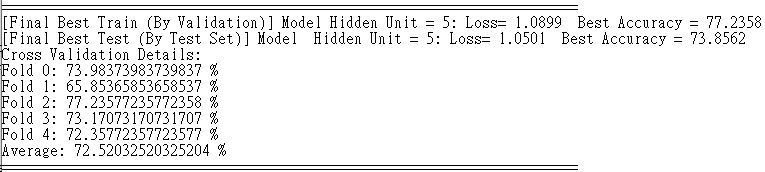


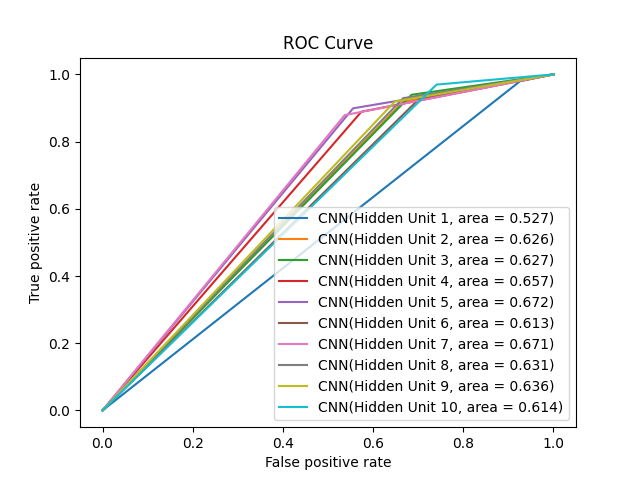


Diabetes dataset (LR 0.1, B 1, E 50)

Program output, we found the best is hidden unit 5 with 73.85% accuracy.

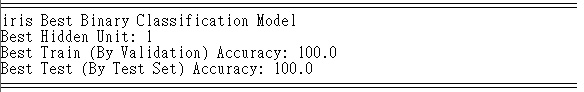


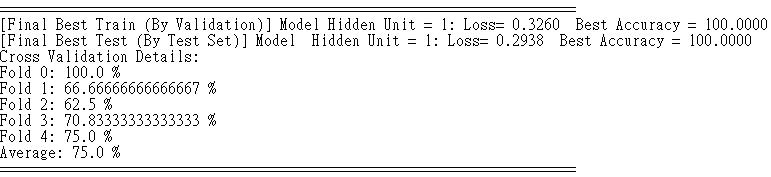


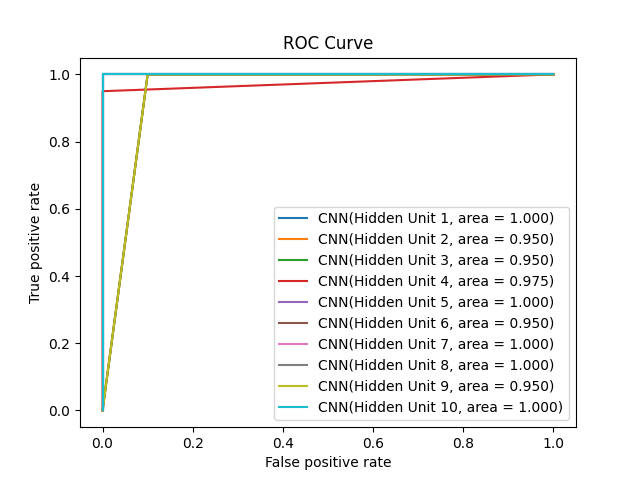


iris dataset (LR 0.1, B 1, E 50)

Program output, we found the best is hidden unit 1 with 100% accuracy.

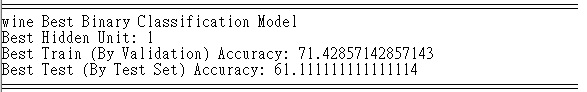


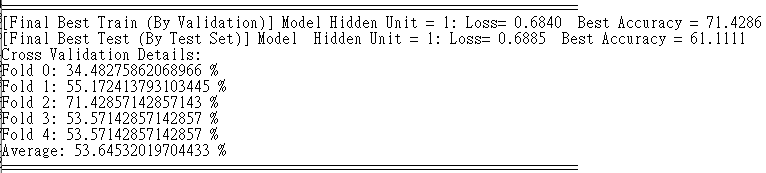


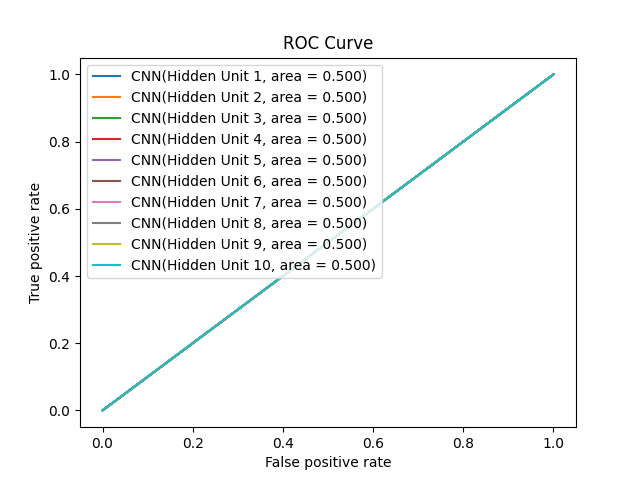


Wine Dataset (LR 0.1, B 1, E 50)

Program output, we found the best is hidden unit 1 with 61.1% accuracy.





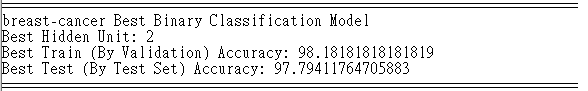


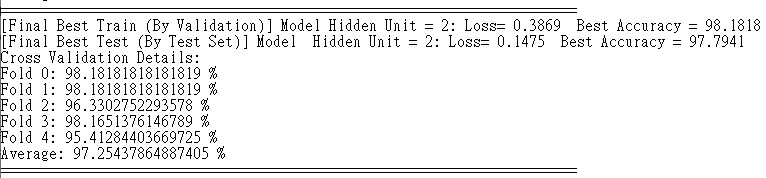
**LR 0.1, B 100, E 1000**

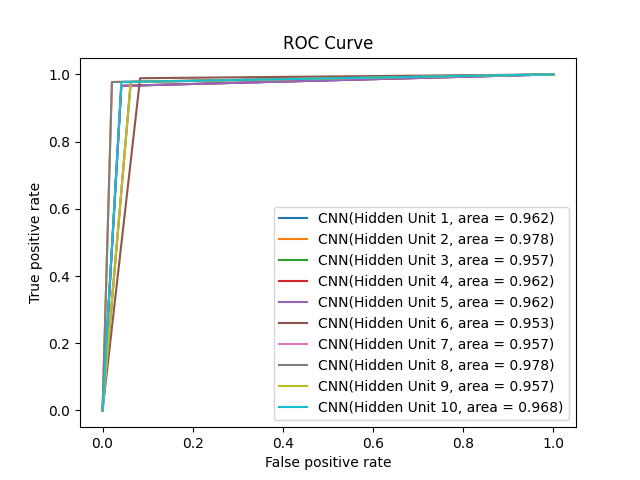
For this parameters, as we notice there is still dataset like diabetes set not finish training (with high loss and low accuracy), only epoch updated to see what is the difference.

breast-cancer dataset (LR 0.1, B 100, E 1000)

Program output, we found the best is hidden unit 2 with 98.5% accuracy.

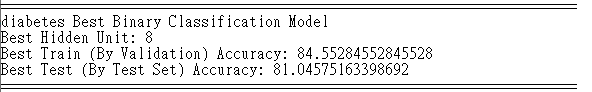


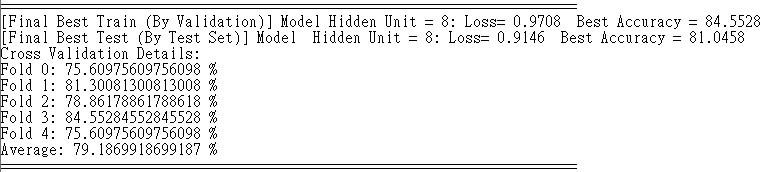


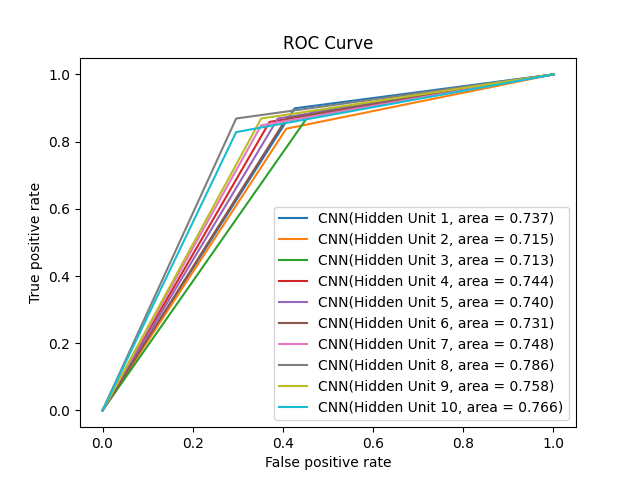


Diabetes dataset (LR 0.1, B 100, E 1000)

Program output, we found the best is hidden unit 8 with 81.55% accuracy.





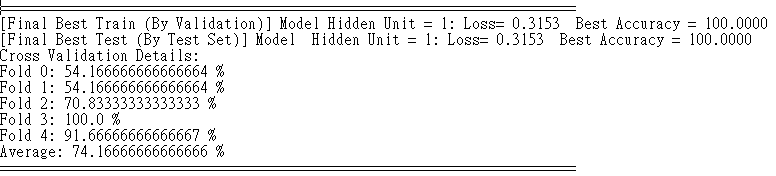


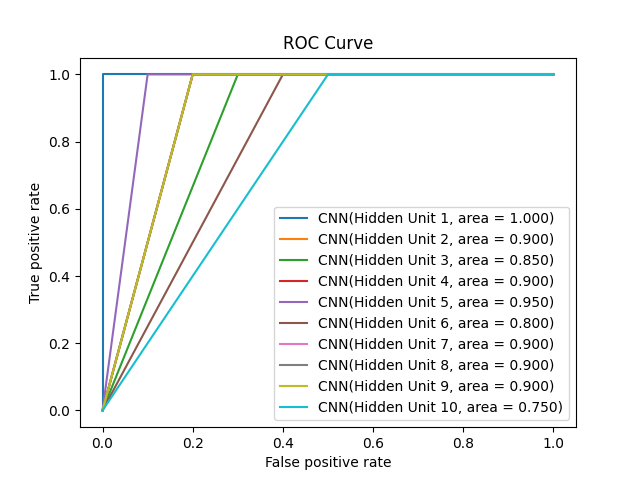
iris dataset (LR 0.1, B 100, E 1000)

Program output, we found the best is hidden unit 1 with 100% accuracy.

一張含有 文字 的圖片

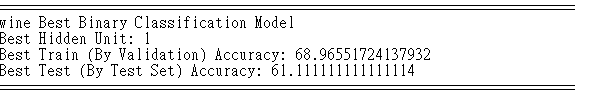
自動產生的描述

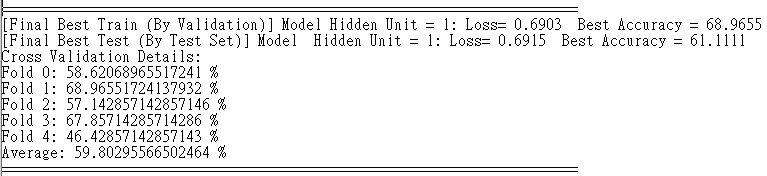


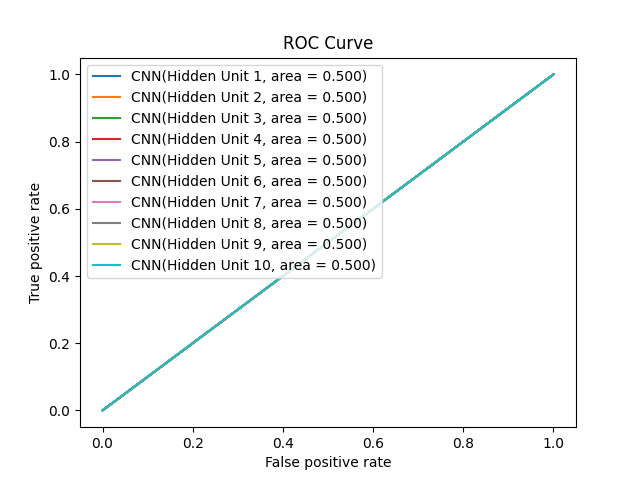


Wine Dataset (LR 0.1, B 100, E 1000)

Program output, we found the best is hidden unit 1 with 61.1% accuracy.





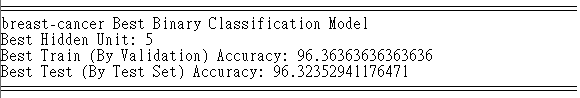


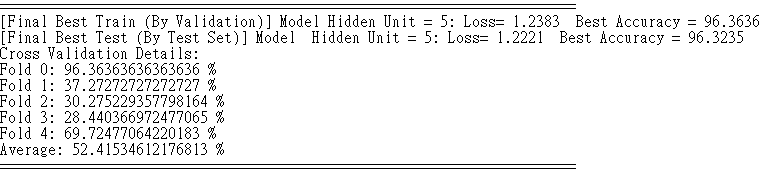
**LR 0.00001, B 100, E 1000**

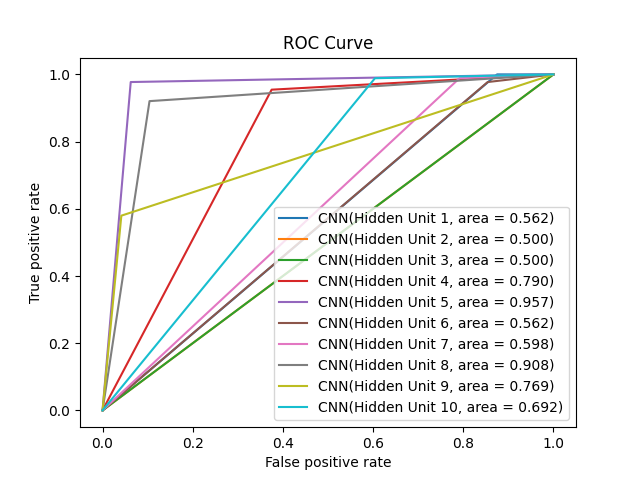
For this parameter, as we notice the result in iris dataset and wine dataset do not actually makes different with number of epochs. And therefore, learning rate is reduced to see if we can learn a better model for these two datasets.

breast-cancer dataset (LR 0.00001, B 100, E 1000)

Program output, we found the best is hidden unit 5 with 96.32% accuracy.

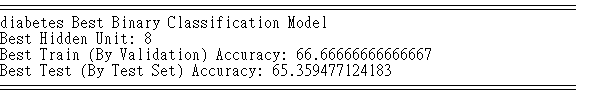


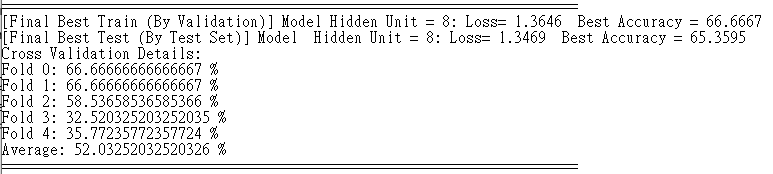


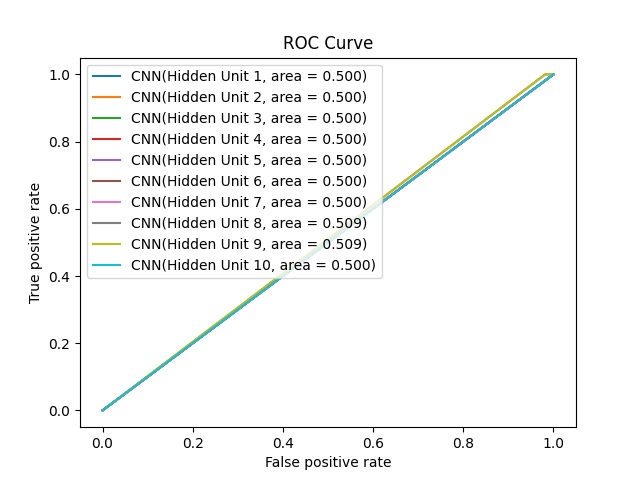


Diabetes dataset (LR 0.00001, B 100, E 1000)

Program output, we found the best is hidden unit 8 with 65.35% accuracy.

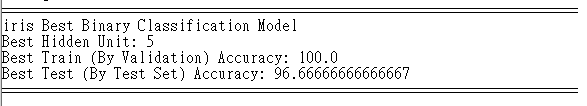






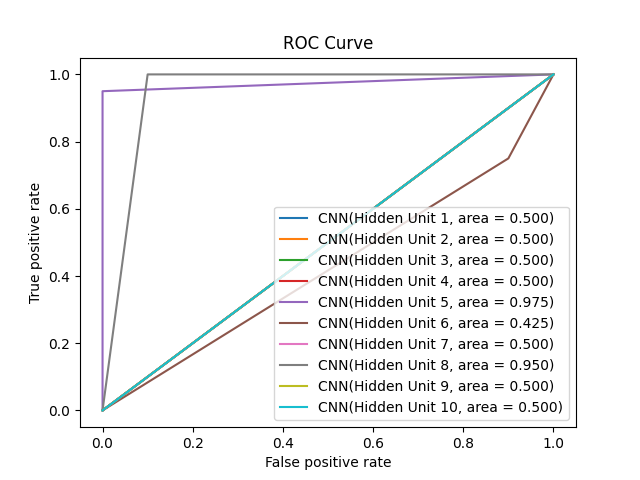
iris dataset (LR 0.00001, B 100, E 1000)

Program output, we found the best is hidden unit 5 with 96.6% accuracy.



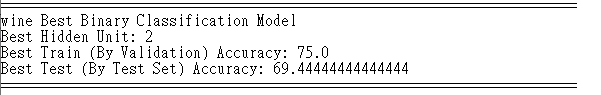
一張含有 文字 的圖片

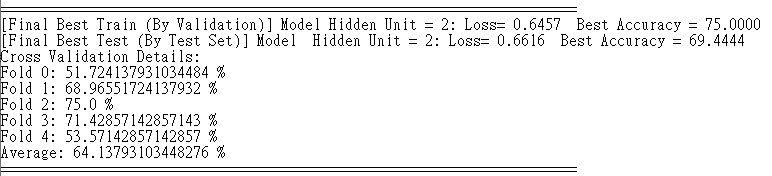
自動產生的描述

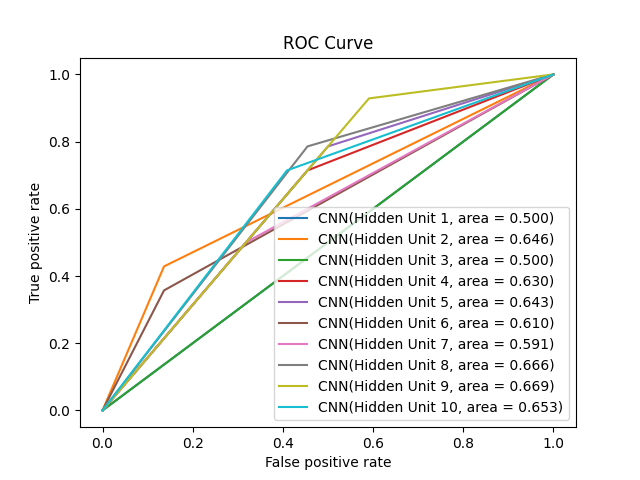


Wine Dataset (LR 0.00001, B 100, E 1000)

Program output, we found the best is hidden unit 2 with 69.4% accuracy.





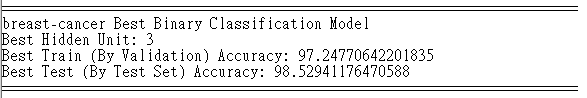


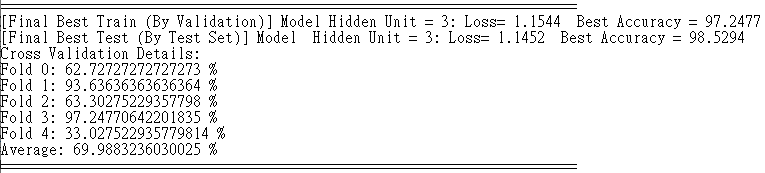
**LR 0.0001, B 100, E 2000**

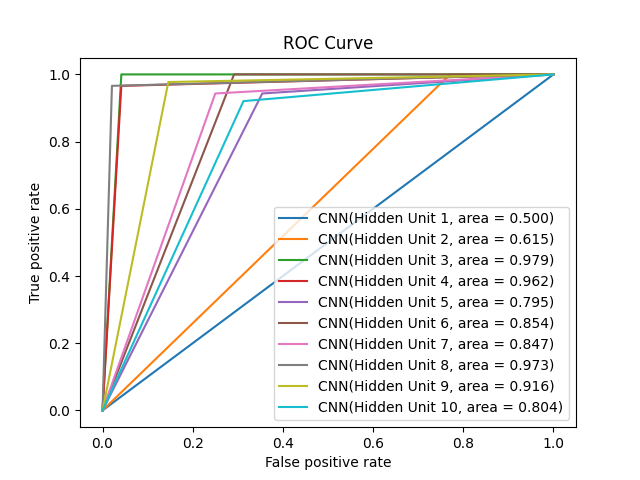
For this parameter, as we notice the result of most dataset being worse after reducing learning rate, therefore we increase learning rate a bit for better learning. For previous parameters, there is finally a different in wine dataset. Therefore, for previous approach and assumption should be right, and try with a larger epoch and learning rate in order to have better coverage.

breast-cancer dataset (LR 0.0001, B 100, E 2000)

Program output, we found the best is hidden unit 3 with 98.5% accuracy.

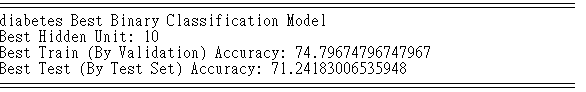


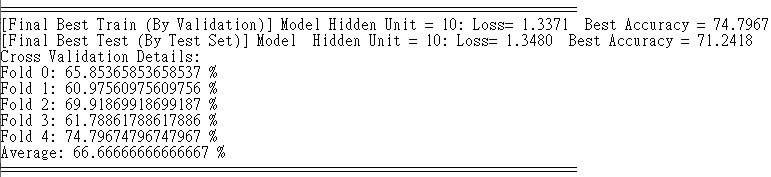


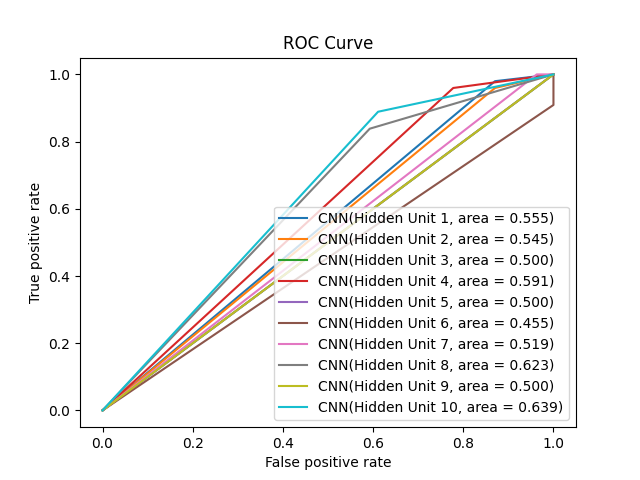


Diabetes dataset (LR 0.0001, B 100, E 2000)

Program output, we found the best is hidden unit 10 with 71.24% accuracy.

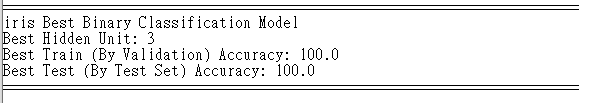


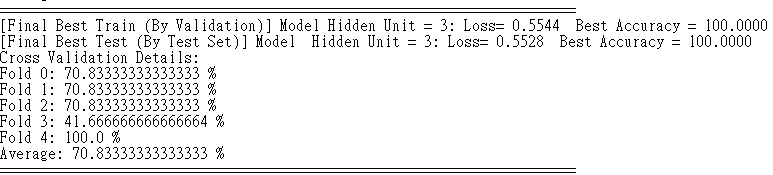


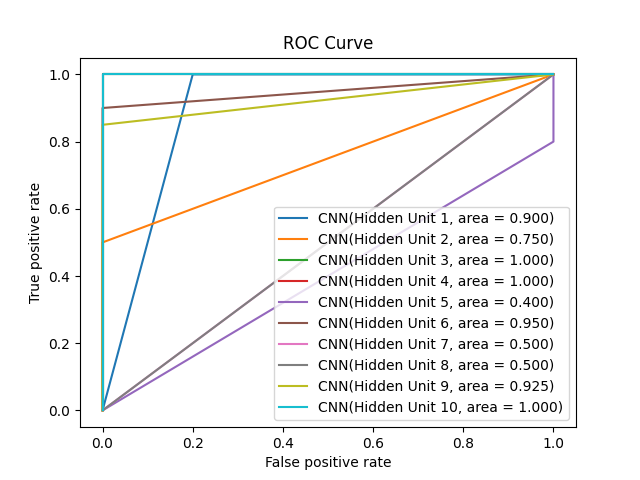


iris dataset (LR 0.0001, B 100, E 2000)

Program output, we found the best is hidden unit 3 with 100% accuracy.

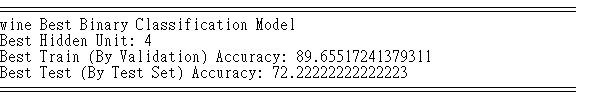


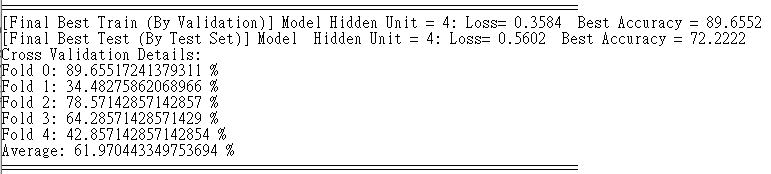


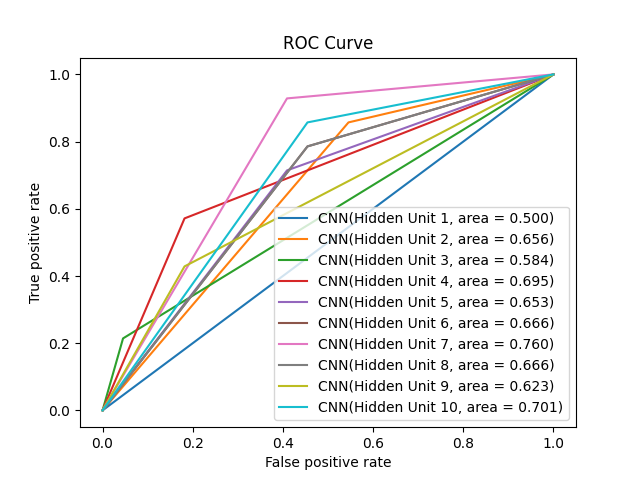


Wine Dataset (LR 0.0001, B 100, E 2000)

Program output, we found the best is hidden unit 4 with 72.2% accuracy.







**Multi-class Classification**

The following 4 set of parameters have been applied.

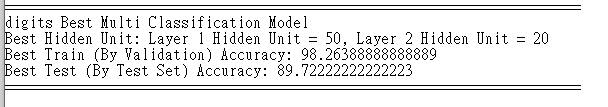
1. Learning Rate 0.1, Batch Size 1, Epoch 50
2. Learning Rate 0.1, Batch Size 100, Epoch 2000
3. Learning Rate 1, Batch Size 100, Epoch 100
4. Learning Rate 0.001, Batch Size 100, Epoch 2000

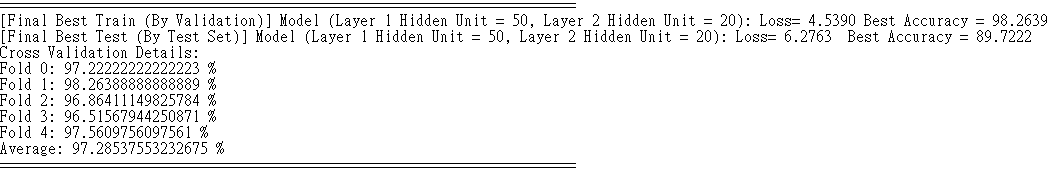
In the following report, I will simply use LR for learning rate, B for batch size, and E for numbers of epoch. For the best model, if the accuracy is the same, smaller hidden unit will be considered.

**LR 0.1, B 1, E 50**

For the first set, it mainly for program correctness testing and debugging. Therefore, the data is just for reference.

Program output, we found the best model with layer 1 with 50 hidden unit, and layer 2 with 20 hidden unit, with accuracy 89.72.



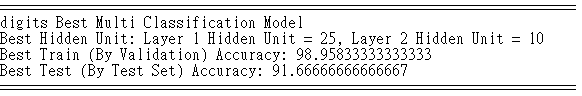


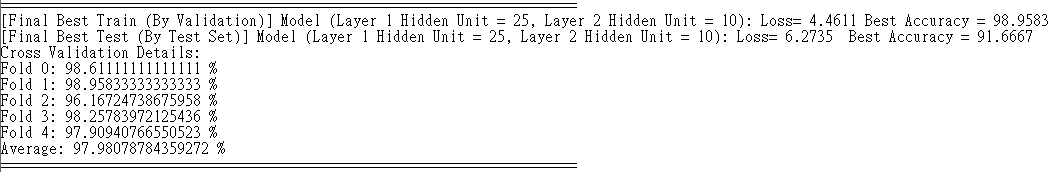
|  |  |  |  |
| --- | --- | --- | --- |
|  | 5 | 10 | 20 |
| 15 |  |  |  |
| 25 |  |  |  |
| 50 |  |  |  |

**LR 0.1, B 100, E 2000**

As I want to see if accuracy can be further improved, therefore epoch increase.

Program output, we found the best model with layer 1 with 25 hidden unit, and layer 2 with 10 hidden unit, with accuracy 91.6.



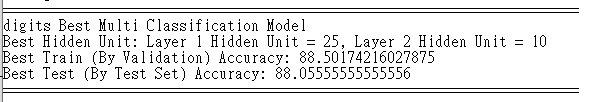


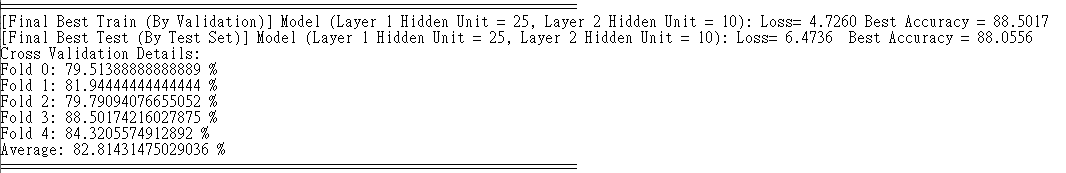
|  |  |  |  |
| --- | --- | --- | --- |
|  | 5 | 10 | 20 |
| 15 |  |  |  |
| 25 |  |  |  |
| 50 |  |  |  |

**LR 0.001, B 100, E 2000**

It is one example to see if decreasing learning rate, can it learn with huge epoch.

Program output, we found the best model with layer 1 with 25 hidden unit, and layer 2 with 10 hidden unit, with accuracy 88.055.



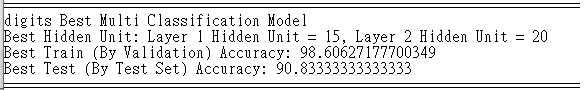


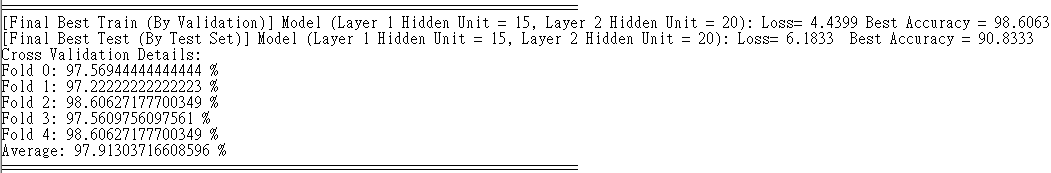
|  |  |  |  |
| --- | --- | --- | --- |
|  | 5 | 10 | 20 |
| 15 |  |  |  |
| 25 |  |  |  |
| 50 |  |  |  |

**LR 1, B 100, E 100**

As I want to see if can shorter training time achieve a similar result, therefore lr increase and e decrease.

Program output, we found the best model with layer 1 with 25 hidden unit, and layer 2 with 10 hidden unit, with accuracy 91.6.





|  |  |  |  |
| --- | --- | --- | --- |
|  | 5 | 10 | 20 |
| 15 |  |  |  |
| 25 |  |  |  |
| 50 |  |  |  |

For any details or original ouput, please refer to output.zip, which contains all output images and useful output text.

**Summary**

Smaller learning rate will lead to a slower learning. Large epoch will help to learn more in order to head to optimal learning. In binary classification, in final setup, learning rate 0.0001 and 2000 epoch have nearly arrived an optimal learning such that each dataset can maximize the accuracy. For multi-class classification, luckily with 0.1 learning rate and 2000 epoch, the accuracy is satisfactory. The approach in the next two parameters is to see how learning rate and epoch related such that to minimize training time, or further increasing the accuracy by learning rate.