

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

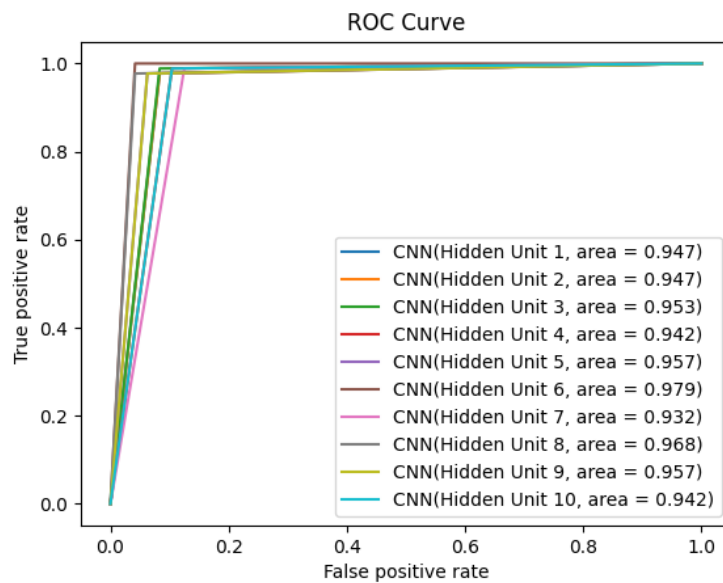
For the first set, it mainly for program correctness testing and debugging. Therefore, the data is just for reference.

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

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

```
breast-cancer Best Binary Classification Model
Best Hidden Unit: 6
Best Train (By Validation) Accuracy: 98.1651376146789
Best Test (By Test Set) Accuracy: 98.52941176470588
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 6: Loss= 0.2617 Best Accuracy = 98.1651
[Final Best Test (By Test Set)] Model Hidden Unit = 6: Loss= 0.2090 Best Accuracy = 98.5294
Cross Validation Details:
Fold 0: 96.36363636363636 %
Fold 1: 97.27272727272727 %
Fold 2: 95.41284403669725 %
Fold 3: 97.24770642201835 %
Fold 4: 98.1651376146789 %
Average: 96.89241034195163 %
```

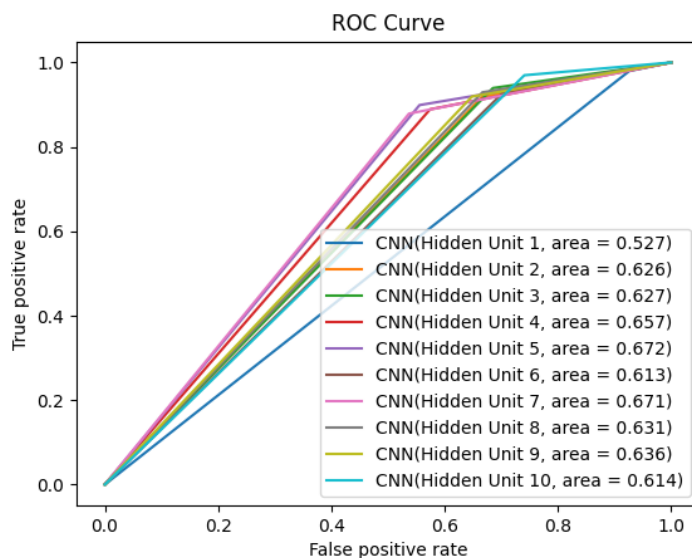


Diabetes dataset (LR 0.1, B 1, E 50)

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

```
diabetes Best Binary Classification Model
Best Hidden Unit: 5
Best Train (By Validation) Accuracy: 77.23577235772358
Best Test (By Test Set) Accuracy: 73.85620915032679
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 5: Loss= 1.0899 Best Accuracy = 77.2358
[Final Best Test (By Test Set)] Model Hidden Unit = 5: Loss= 1.0501 Best Accuracy = 73.8562
Cross Validation Details:
Fold 0: 73.98373983739837 %
Fold 1: 65.85365853658537 %
Fold 2: 77.23577235772358 %
Fold 3: 73.17073170731707 %
Fold 4: 72.35772357723577 %
Average: 72.52032520325204 %
```

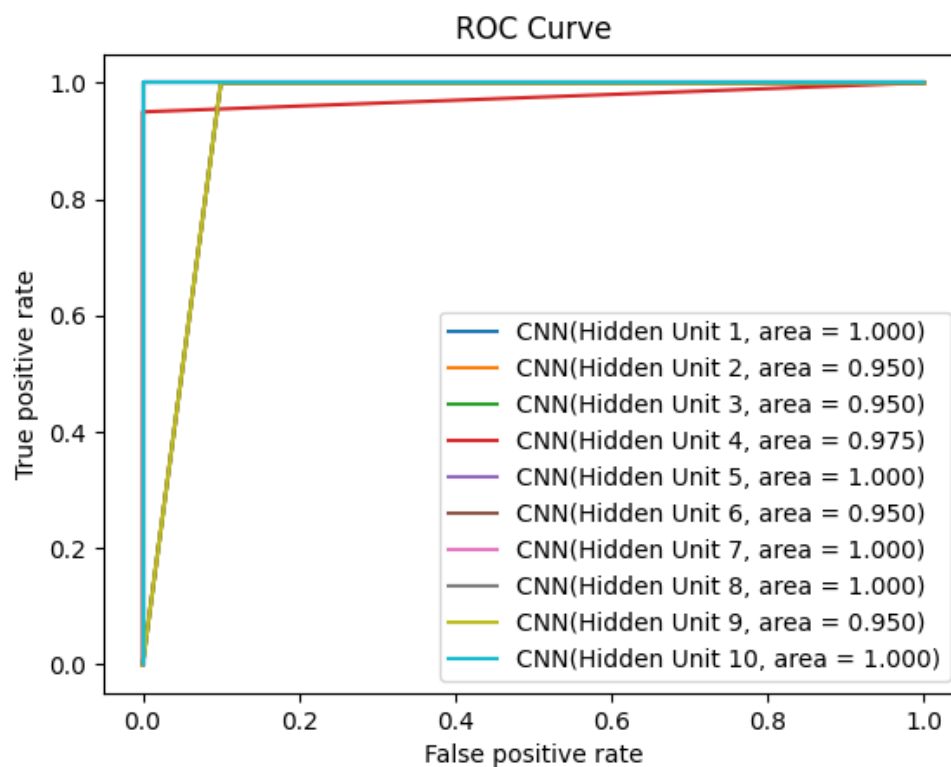


iris dataset (LR 0.1, B 1, E 50)

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

```
iris Best Binary Classification Model
Best Hidden Unit: 1
Best Train (By Validation) Accuracy: 100.0
Best Test (By Test Set) Accuracy: 100.0
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 1: Loss= 0.3260 Best Accuracy = 100.0000
[Final Best Test (By Test Set)] Model Hidden Unit = 1: Loss= 0.2938 Best Accuracy = 100.0000
Cross Validation Details:
Fold 0: 100.0 %
Fold 1: 66.66666666666667 %
Fold 2: 62.5 %
Fold 3: 70.83333333333333 %
Fold 4: 75.0 %
Average: 75.0 %
```

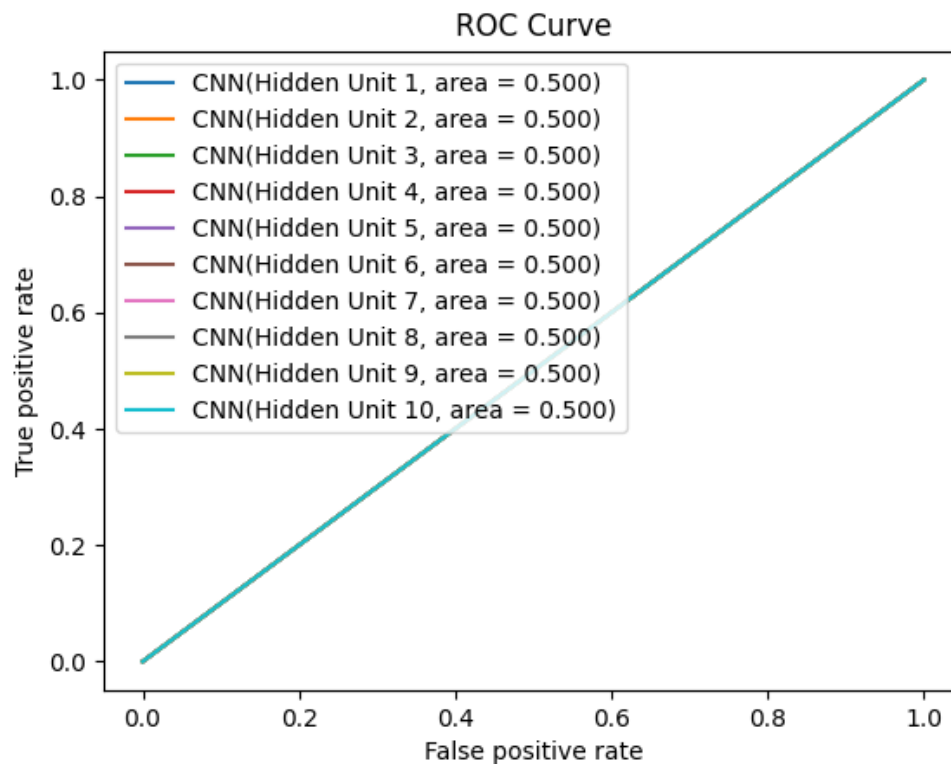


Wine Dataset (LR 0.1, B 1, E 50)

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

```
wine Best Binary Classification Model  
Best Hidden Unit: 1  
Best Train (By Validation) Accuracy: 71.42857142857143  
Best Test (By Test Set) Accuracy: 61.11111111111114
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 1: Loss= 0.6840 Best Accuracy = 71.4286  
[Final Best Test (By Test Set)] Model Hidden Unit = 1: Loss= 0.6885 Best Accuracy = 61.1111  
Cross Validation Details:  
Fold 0: 34.48275862068966 %  
Fold 1: 55.172413793103445 %  
Fold 2: 71.42857142857143 %  
Fold 3: 53.57142857142857 %  
Fold 4: 53.57142857142857 %  
Average: 53.64532019704433 %
```



LR 0.1, B 100, E 1000

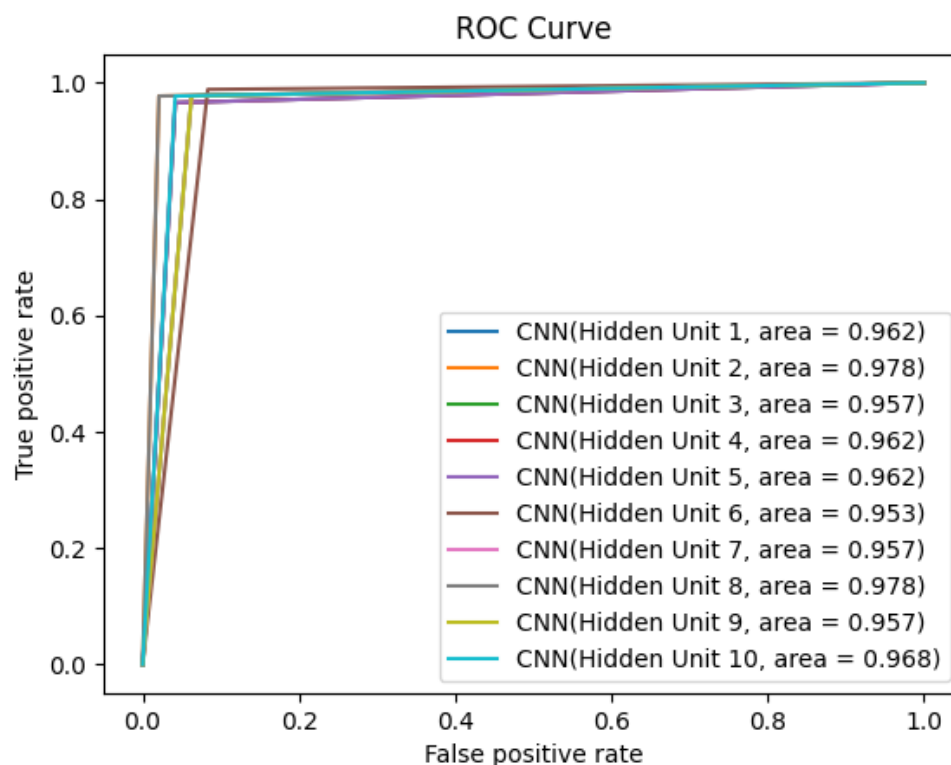
For this parameters, as we notice there is still dataset like diabetes set not finish training (with high loss and low accuracy), only epoch updated to see what is the difference.

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

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

```
breast-cancer Best Binary Classification Model
Best Hidden Unit: 2
Best Train (By Validation) Accuracy: 98.181818181819
Best Test (By Test Set) Accuracy: 97.79411764705883
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 2: Loss= 0.3869 Best Accuracy = 98.1818
[Final Best Test (By Test Set)] Model Hidden Unit = 2: Loss= 0.1475 Best Accuracy = 97.7941
Cross Validation Details:
Fold 0: 98.181818181819 %
Fold 1: 98.181818181819 %
Fold 2: 96.3302752293578 %
Fold 3: 98.1651376146789 %
Fold 4: 95.41284403669725 %
Average: 97.25437864887405 %
```

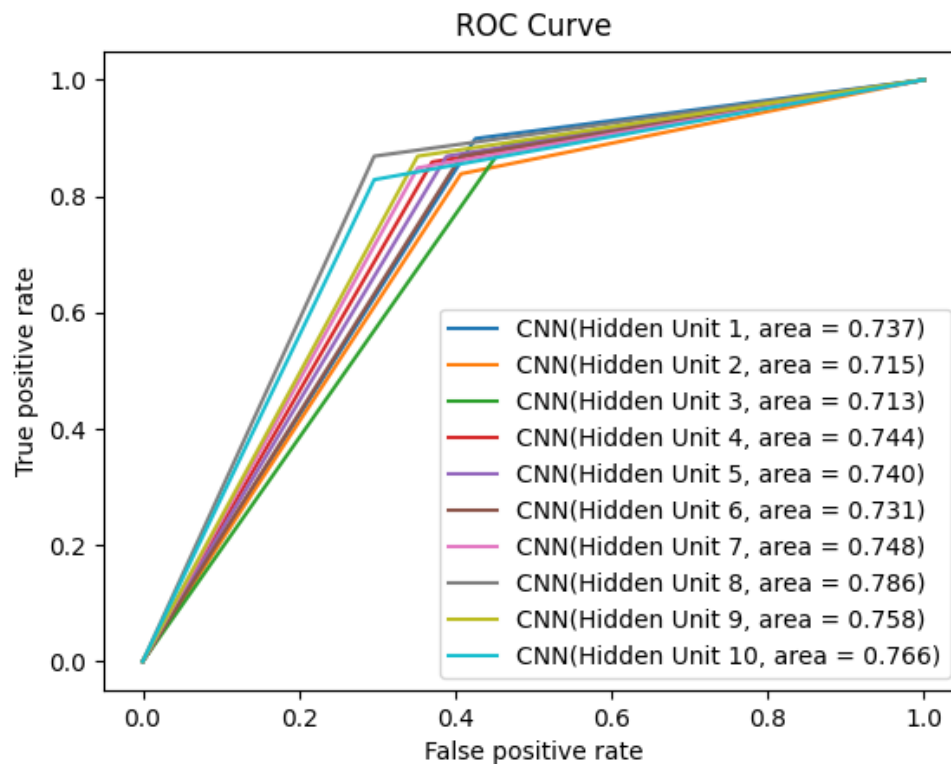


Diabetes dataset (LR 0.1, B 100, E 1000)

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

```
diabetes Best Binary Classification Model
Best Hidden Unit: 8
Best Train (By Validation) Accuracy: 84.55284552845528
Best Test (By Test Set) Accuracy: 81.04575163398692
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 8: Loss= 0.9708 Best Accuracy = 84.5528
[Final Best Test (By Test Set)] Model Hidden Unit = 8: Loss= 0.9146 Best Accuracy = 81.0458
Cross Validation Details:
Fold 0: 75.60975609756098 %
Fold 1: 81.30081300813008 %
Fold 2: 78.86178861788618 %
Fold 3: 84.55284552845528 %
Fold 4: 75.60975609756098 %
Average: 79.1869918699187 %
```

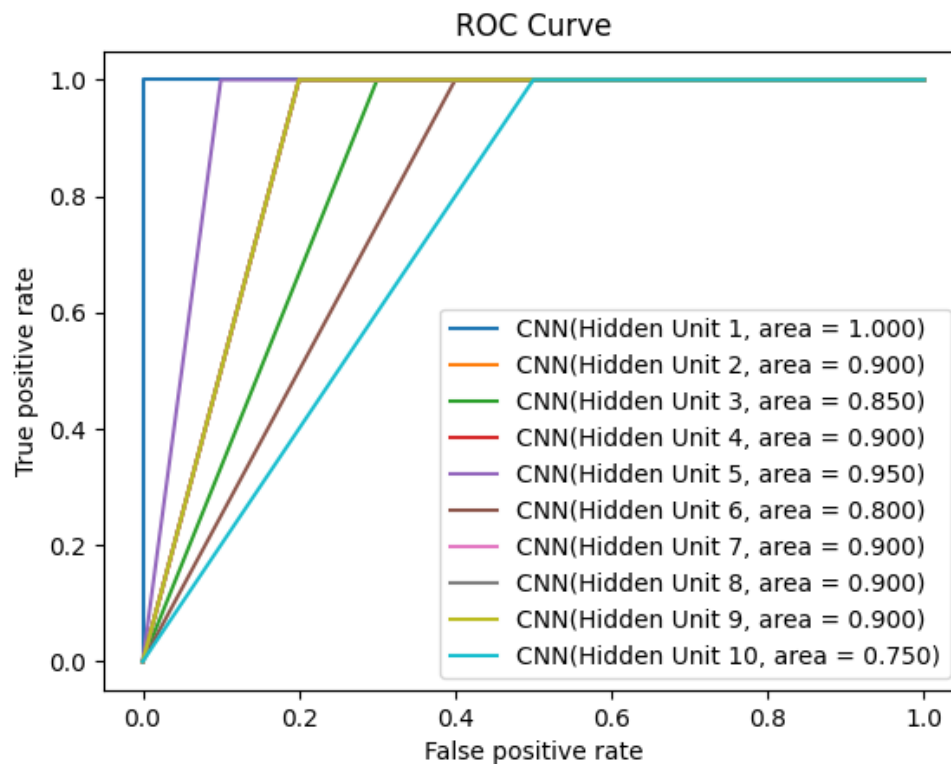


iris dataset (LR 0.1, B 100, E 1000)

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

```
iris Best Binary Classification Model
Best Hidden Unit: 1
Best Train (By Validation) Accuracy: 100.0
Best Test (By Test Set) Accuracy: 100.0
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 1: Loss= 0.3153 Best Accuracy = 100.0000
[Final Best Test (By Test Set)] Model Hidden Unit = 1: Loss= 0.3153 Best Accuracy = 100.0000
Cross Validation Details:
Fold 0: 54.166666666666664 %
Fold 1: 54.166666666666664 %
Fold 2: 70.83333333333333 %
Fold 3: 100.0 %
Fold 4: 91.66666666666667 %
Average: 74.16666666666666 %
```

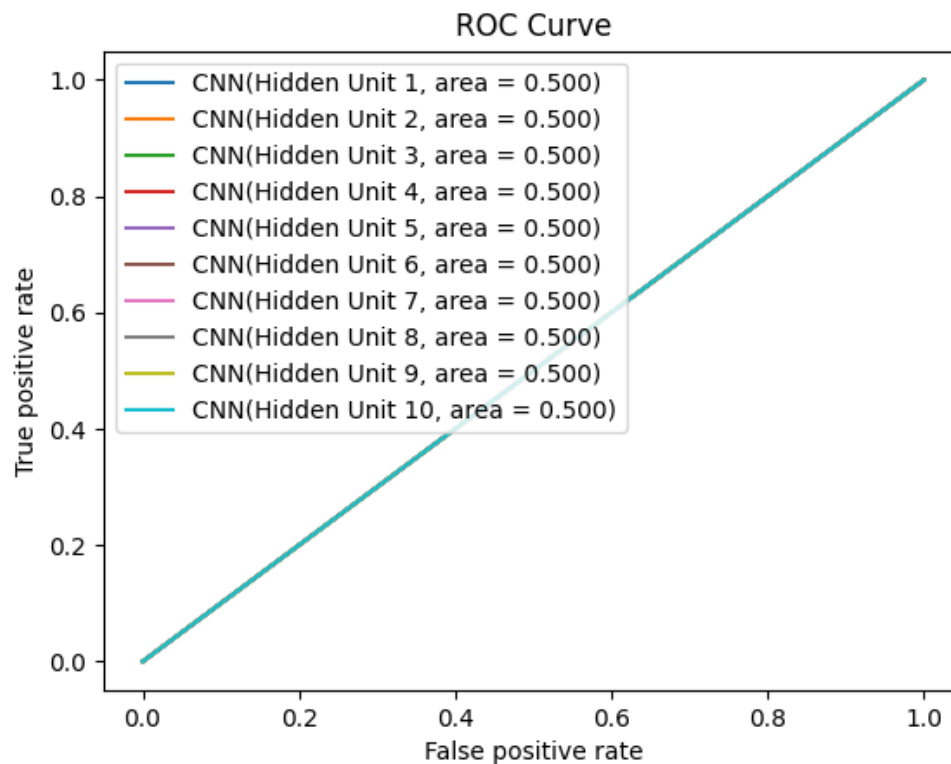


Wine Dataset (LR 0.1, B 100, E 1000)

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

```
wine Best Binary Classification Model
Best Hidden Unit: 1
Best Train (By Validation) Accuracy: 68.96551724137932
Best Test (By Test Set) Accuracy: 61.11111111111114
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 1: Loss= 0.6903 Best Accuracy = 68.9655
[Final Best Test (By Test Set)] Model Hidden Unit = 1: Loss= 0.6915 Best Accuracy = 61.1111
Cross Validation Details:
Fold 0: 58.62068965517241 %
Fold 1: 68.96551724137932 %
Fold 2: 57.142857142857146 %
Fold 3: 67.85714285714286 %
Fold 4: 46.42857142857143 %
Average: 59.80295566502464 %
```



LR 0.00001, B 100, E 1000

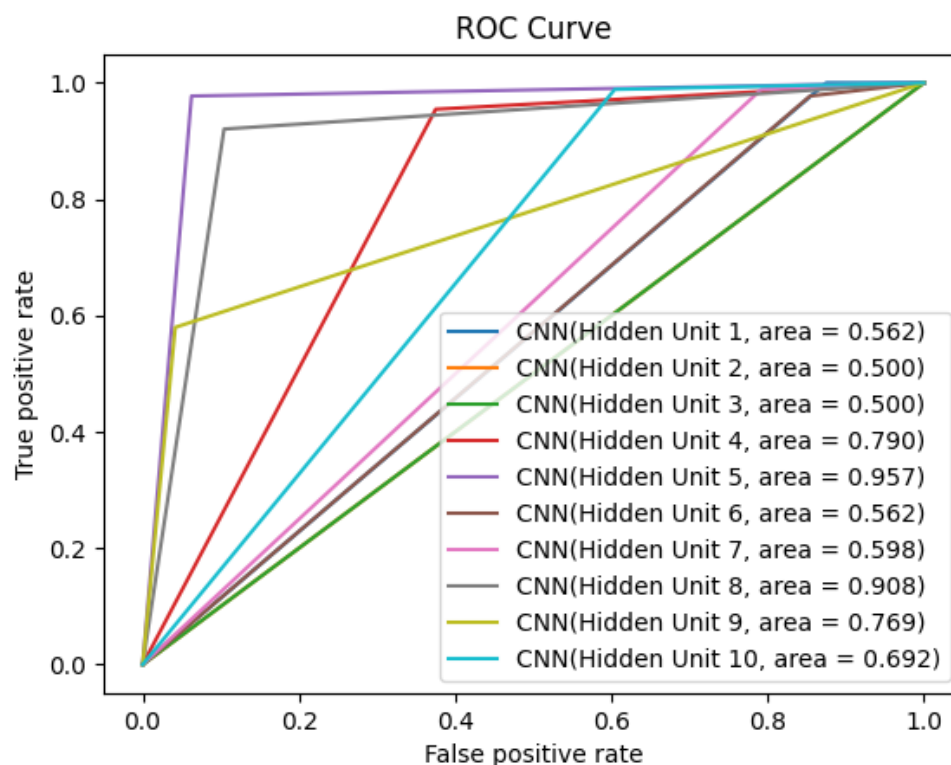
For this parameter, as we notice the result in iris dataset and wine dataset do not actually makes different with number of epochs. And therefore, learning rate is reduced to see if we can learn a better model for these two datasets.

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

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

```
breast-cancer Best Binary Classification Model
Best Hidden Unit: 5
Best Train (By Validation) Accuracy: 96.36363636363636
Best Test (By Test Set) Accuracy: 96.32352941176471
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 5: Loss= 1.2383 Best Accuracy = 96.3636
[Final Best Test (By Test Set)] Model Hidden Unit = 5: Loss= 1.2221 Best Accuracy = 96.3235
Cross Validation Details:
Fold 0: 96.36363636363636 %
Fold 1: 37.27272727272727 %
Fold 2: 30.275229357798164 %
Fold 3: 28.440366972477065 %
Fold 4: 69.72477064220183 %
Average: 52.41534612176813 %
```

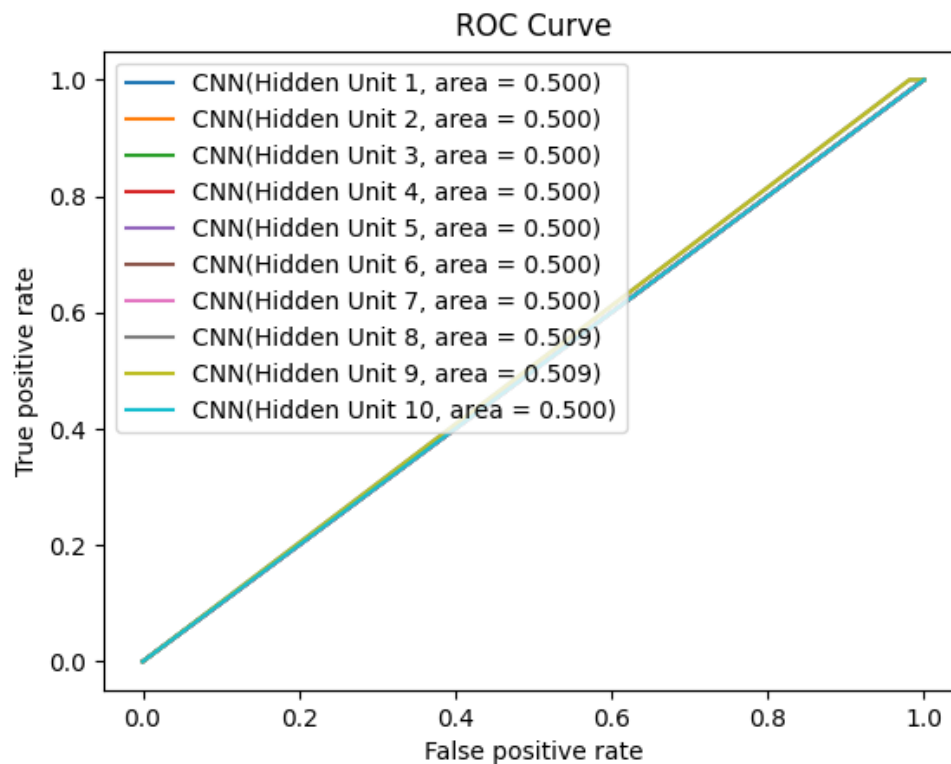


Diabetes dataset (LR 0.00001, B 100, E 1000)

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

```
diabetes Best Binary Classification Model
Best Hidden Unit: 8
Best Train (By Validation) Accuracy: 66.6666666666667
Best Test (By Test Set) Accuracy: 65.359477124183
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 8: Loss= 1.3646 Best Accuracy = 66.6667
[Final Best Test (By Test Set)] Model Hidden Unit = 8: Loss= 1.3469 Best Accuracy = 65.3595
Cross Validation Details:
Fold 0: 66.6666666666667 %
Fold 1: 66.6666666666667 %
Fold 2: 58.53658536585366 %
Fold 3: 32.520325203252035 %
Fold 4: 35.77235772357724 %
Average: 52.03252032520326 %
```

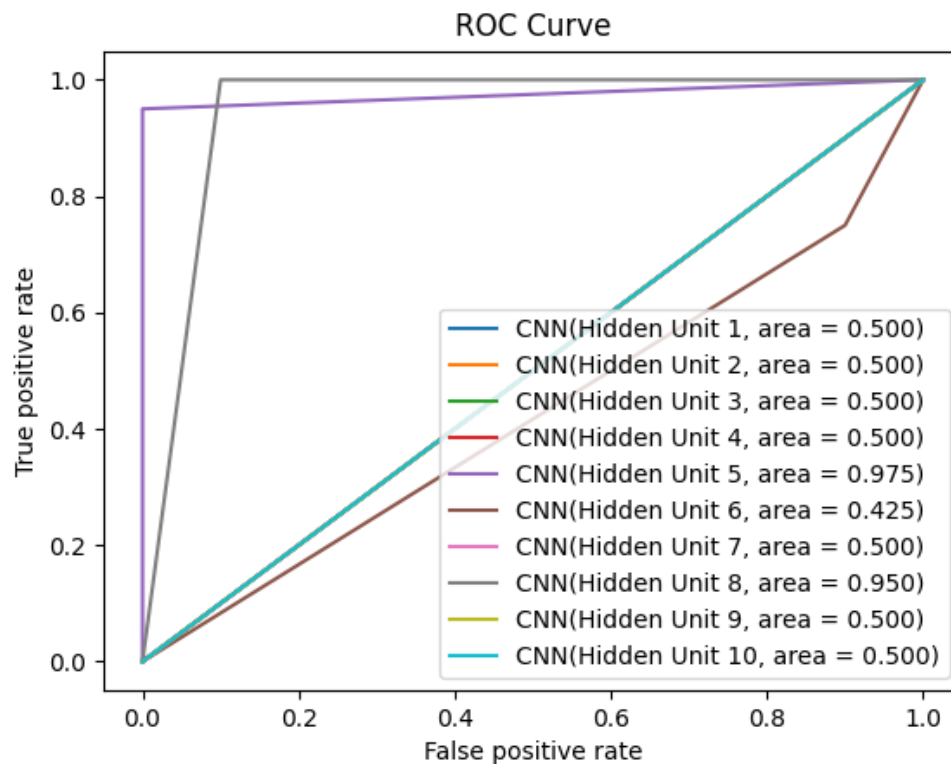


iris dataset (LR 0.00001, B 100, E 1000)

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

```
iris Best Binary Classification Model
Best Hidden Unit: 5
Best Train (By Validation) Accuracy: 100.0
Best Test (By Test Set) Accuracy: 96.66666666666667
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 5: Loss= 0.5759 Best Accuracy = 100.0000
[Final Best Test (By Test Set)] Model Hidden Unit = 5: Loss= 0.6017 Best Accuracy = 96.6667
Cross Validation Details:
Fold 0: 66.66666666666667 %
Fold 1: 41.666666666666664 %
Fold 2: 100.0 %
Fold 3: 29.166666666666668 %
Fold 4: 70.83333333333333 %
Average: 61.666666666666664 %
```

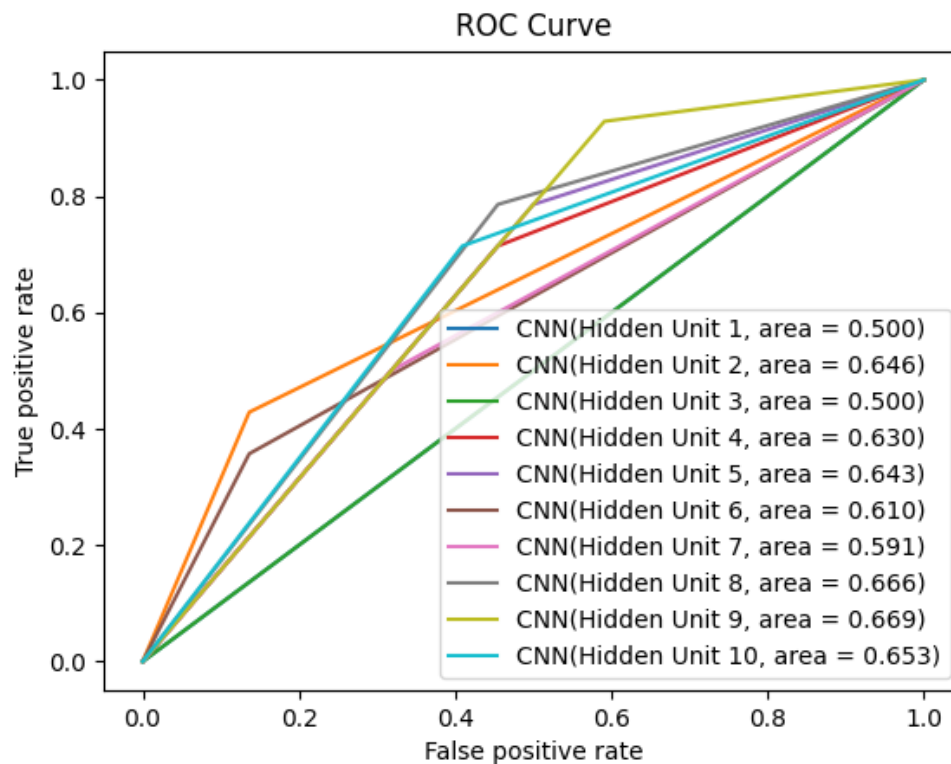


Wine Dataset (LR 0.00001, B 100, E 1000)

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

```
wine Best Binary Classification Model
Best Hidden Unit: 2
Best Train (By Validation) Accuracy: 75.0
Best Test (By Test Set) Accuracy: 69.44444444444444
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 2: Loss= 0.6457 Best Accuracy = 75.0000
[Final Best Test (By Test Set)] Model Hidden Unit = 2: Loss= 0.6616 Best Accuracy = 69.4444
Cross Validation Details:
Fold 0: 51.724137931034484 %
Fold 1: 68.96551724137932 %
Fold 2: 75.0 %
Fold 3: 71.42857142857143 %
Fold 4: 53.57142857142857 %
Average: 64.13793103448276 %
```



LR 0.0001, B 100, E 2000

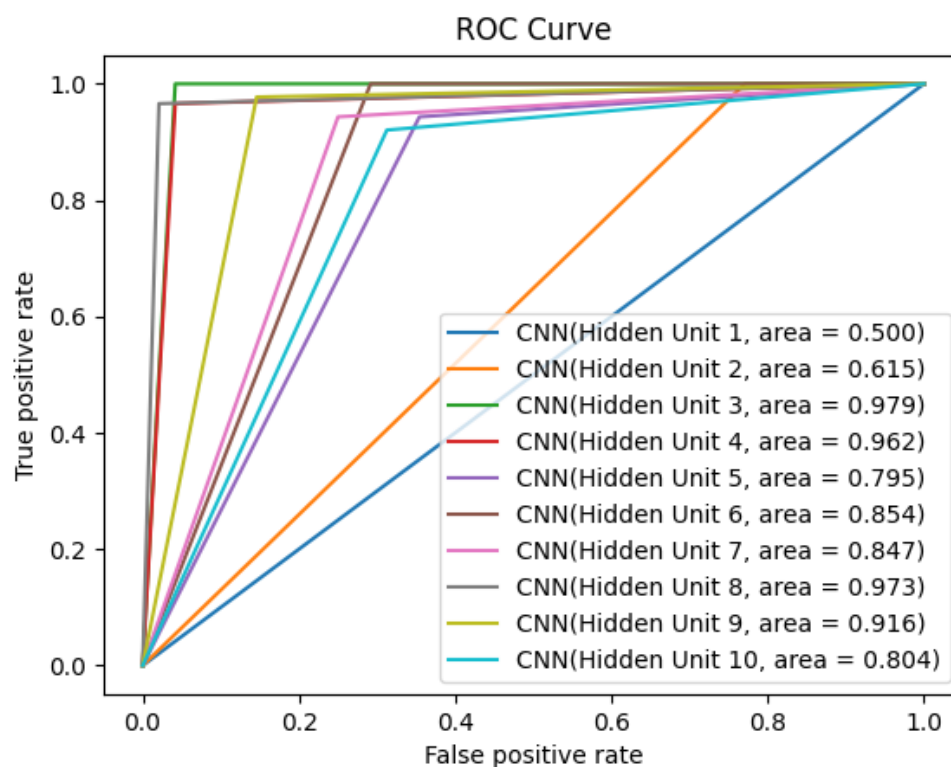
For this parameter, as we notice the result of most dataset being worse after reducing learning rate, therefore we increase learning rate a bit for better learning. For previous parameters, there is finally a different in wine dataset. Therefore, for previous approach and assumption should be right,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

```
breast-cancer Best Binary Classification Model
Best Hidden Unit: 3
Best Train (By Validation) Accuracy: 97.24770642201835
Best Test (By Test Set) Accuracy: 98.52941176470588

[Final Best Train (By Validation)] Model Hidden Unit = 3: Loss= 1.1544 Best Accuracy = 97.2477
[Final Best Test (By Test Set)] Model Hidden Unit = 3: Loss= 1.1452 Best Accuracy = 98.5294
Cross Validation Details:
Fold 0: 62.72727272727273 %
Fold 1: 93.63636363636364 %
Fold 2: 63.30275229357798 %
Fold 3: 97.24770642201835 %
Fold 4: 33.027522935779814 %
Average: 69.9883236030025 %
```

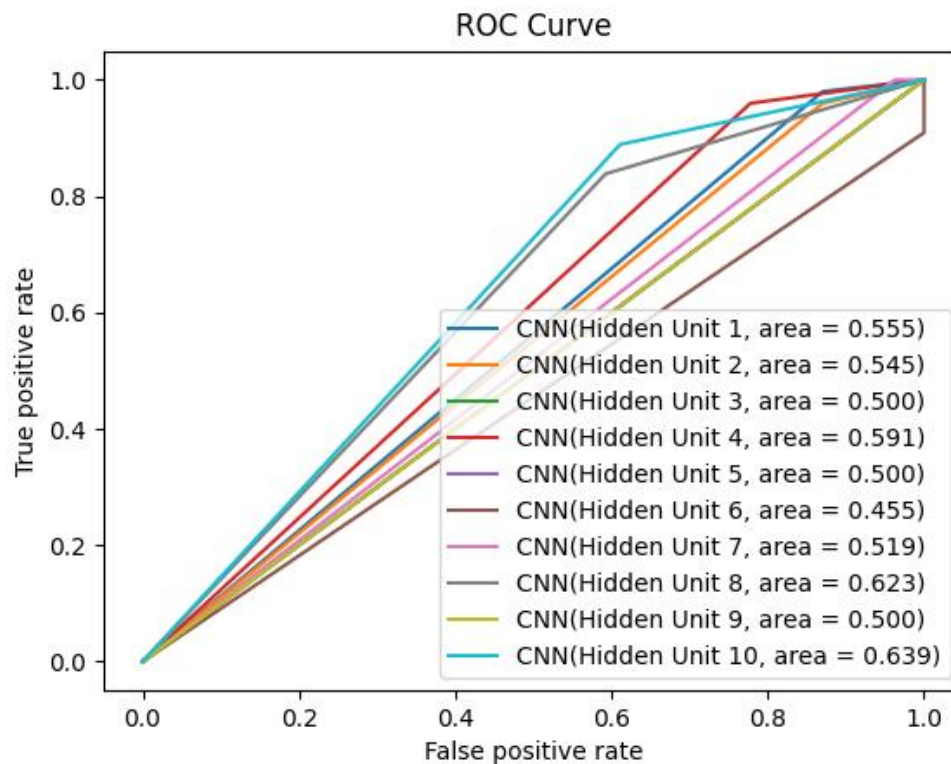


Diabetes dataset (LR 0.0001, B 100, E 2000)

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

```
diabetes Best Binary Classification Model
Best Hidden Unit: 10
Best Train (By Validation) Accuracy: 74.79674796747967
Best Test (By Test Set) Accuracy: 71.24183006535948
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 10: Loss= 1.3371 Best Accuracy = 74.7967
[Final Best Test (By Test Set)] Model Hidden Unit = 10: Loss= 1.3480 Best Accuracy = 71.2418
Cross Validation Details:
Fold 0: 65.85365853658537 %
Fold 1: 60.97560975609756 %
Fold 2: 69.91869918699187 %
Fold 3: 61.78861788617886 %
Fold 4: 74.79674796747967 %
Average: 66.66666666666667 %
```

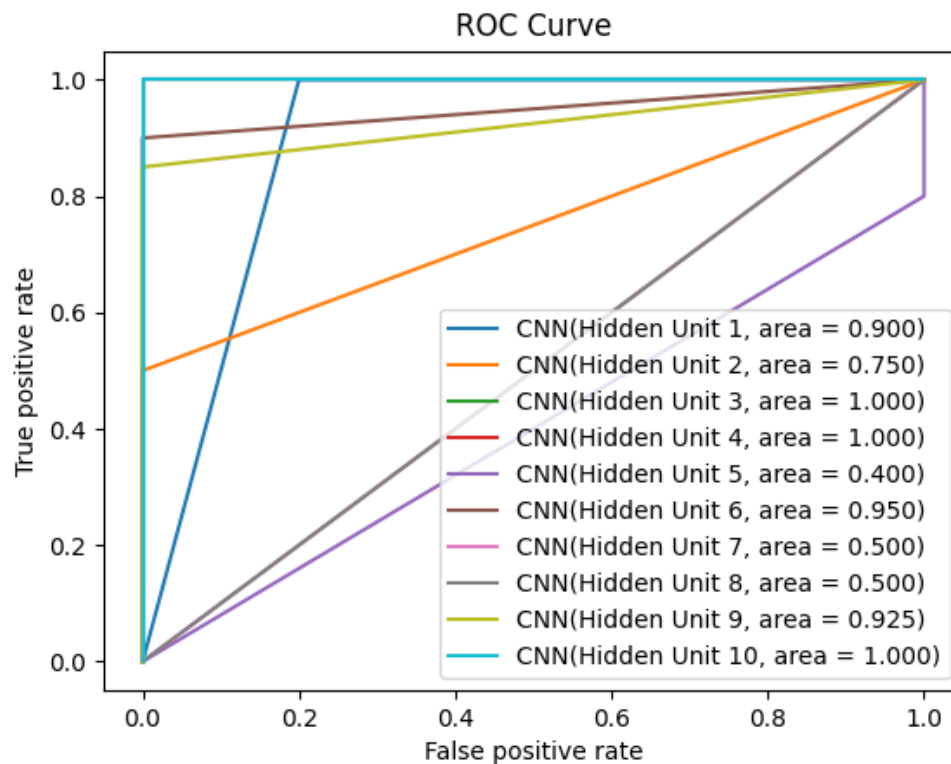


iris dataset (LR 0.0001, B 100, E 2000)

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

```
iris Best Binary Classification Model
Best Hidden Unit: 3
Best Train (By Validation) Accuracy: 100.0
Best Test (By Test Set) Accuracy: 100.0
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 3: Loss= 0.5544 Best Accuracy = 100.0000
[Final Best Test (By Test Set)] Model Hidden Unit = 3: Loss= 0.5528 Best Accuracy = 100.0000
Cross Validation Details:
Fold 0: 70.83333333333333 %
Fold 1: 70.83333333333333 %
Fold 2: 70.83333333333333 %
Fold 3: 41.666666666666664 %
Fold 4: 100.0 %
Average: 70.83333333333333 %
```

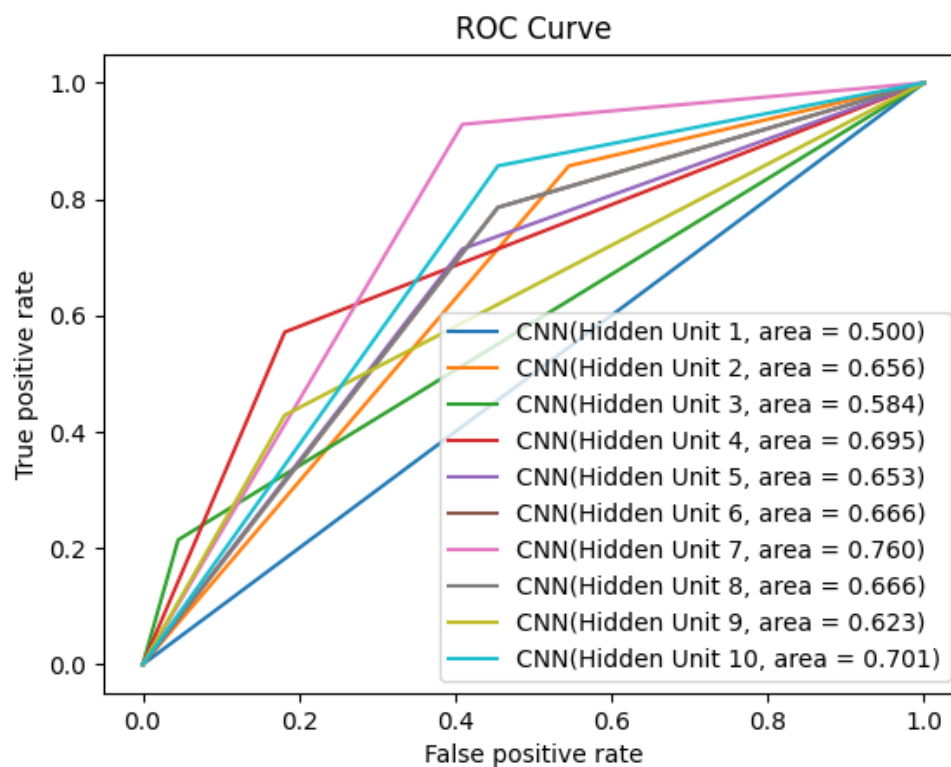


Wine Dataset (LR 0.0001, B 100, E 2000)

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

```
wine Best Binary Classification Model
Best Hidden Unit: 4
Best Train (By Validation) Accuracy: 89.65517241379311
Best Test (By Test Set) Accuracy: 72.22222222222223
```

```
[Final Best Train (By Validation)] Model Hidden Unit = 4: Loss= 0.3584 Best Accuracy = 89.6552
[Final Best Test (By Test Set)] Model Hidden Unit = 4: Loss= 0.5602 Best Accuracy = 72.2222
Cross Validation Details:
Fold 0: 89.65517241379311 %
Fold 1: 34.48275862068966 %
Fold 2: 78.57142857142857 %
Fold 3: 64.28571428571429 %
Fold 4: 42.85714285714285 %
Average: 61.970443349753694 %
```



Multi-class Classification

The following 4 set of parameters have been applied.

5. Learning Rate 0.1, Batch Size 1, Epoch 50
6. Learning Rate 0.1, Batch Size 100, Epoch 2000
7. Learning Rate 1, Batch Size 100, Epoch 100
8. 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.

```

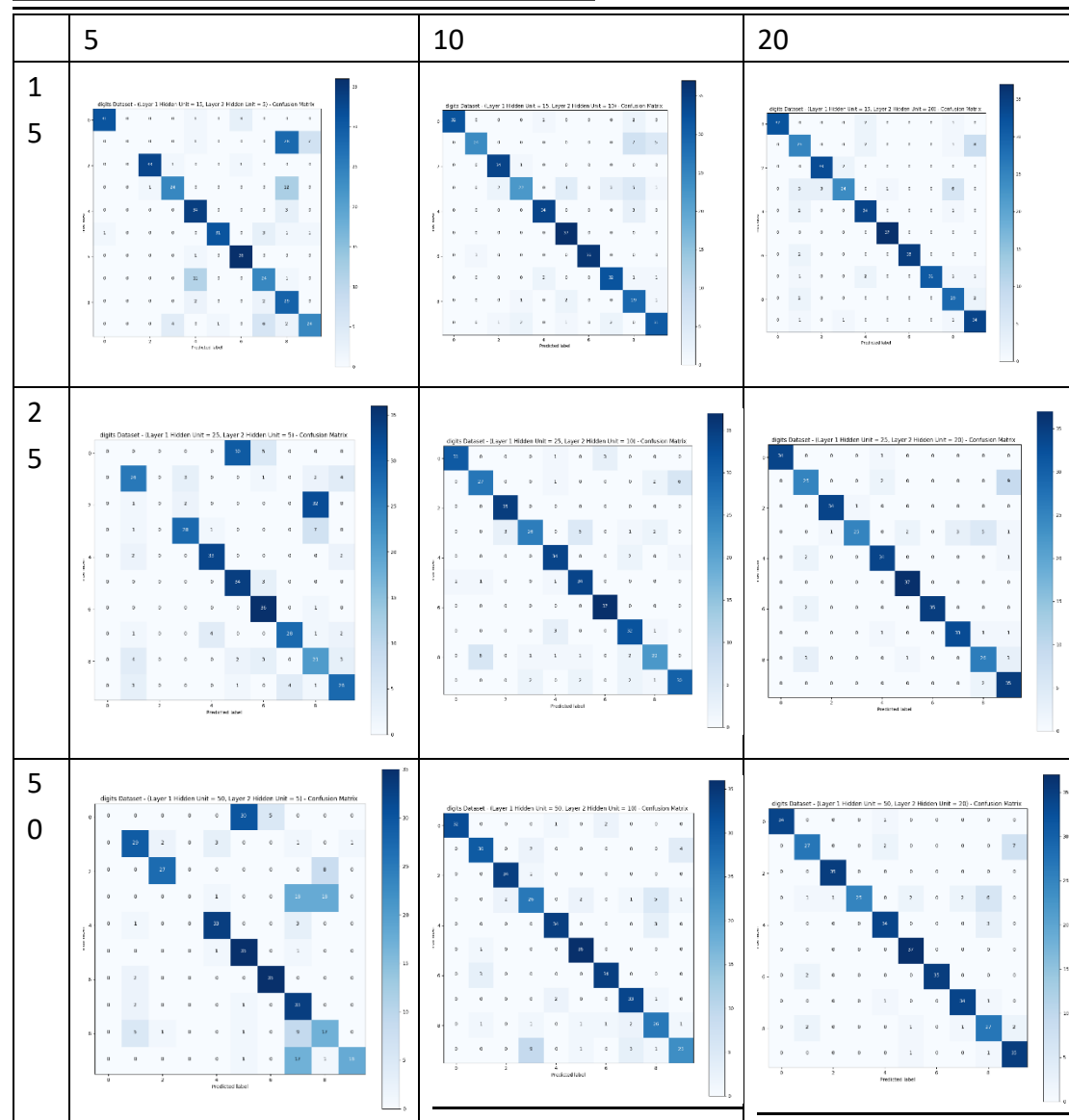
=====
digits Best Multi Classification Model
Best Hidden Unit: Layer 1 Hidden Unit = 50, Layer 2 Hidden Unit = 20
Best Train (By Validation) Accuracy: 98.26388888888889
Best Test (By Test Set) Accuracy: 89.72222222222223
=====

```

```

[Final Best Train (By Validation)] Model (Layer 1 Hidden Unit = 50, Layer 2 Hidden Unit = 20): Loss= 4.5390 Best Accuracy = 98.2639
[Final Best Test (By Test Set)] Model (Layer 1 Hidden Unit = 50, Layer 2 Hidden Unit = 20): Loss= 6.2763 Best Accuracy = 89.7222
Cross Validation Details:
Fold 0: 97.22222222222223 %
Fold 1: 98.26388888888889 %
Fold 2: 96.86411149825784 %
Fold 3: 96.51567944250871 %
Fold 4: 97.5609756097561 %
Average: 97.28537553232675 %
=====

```



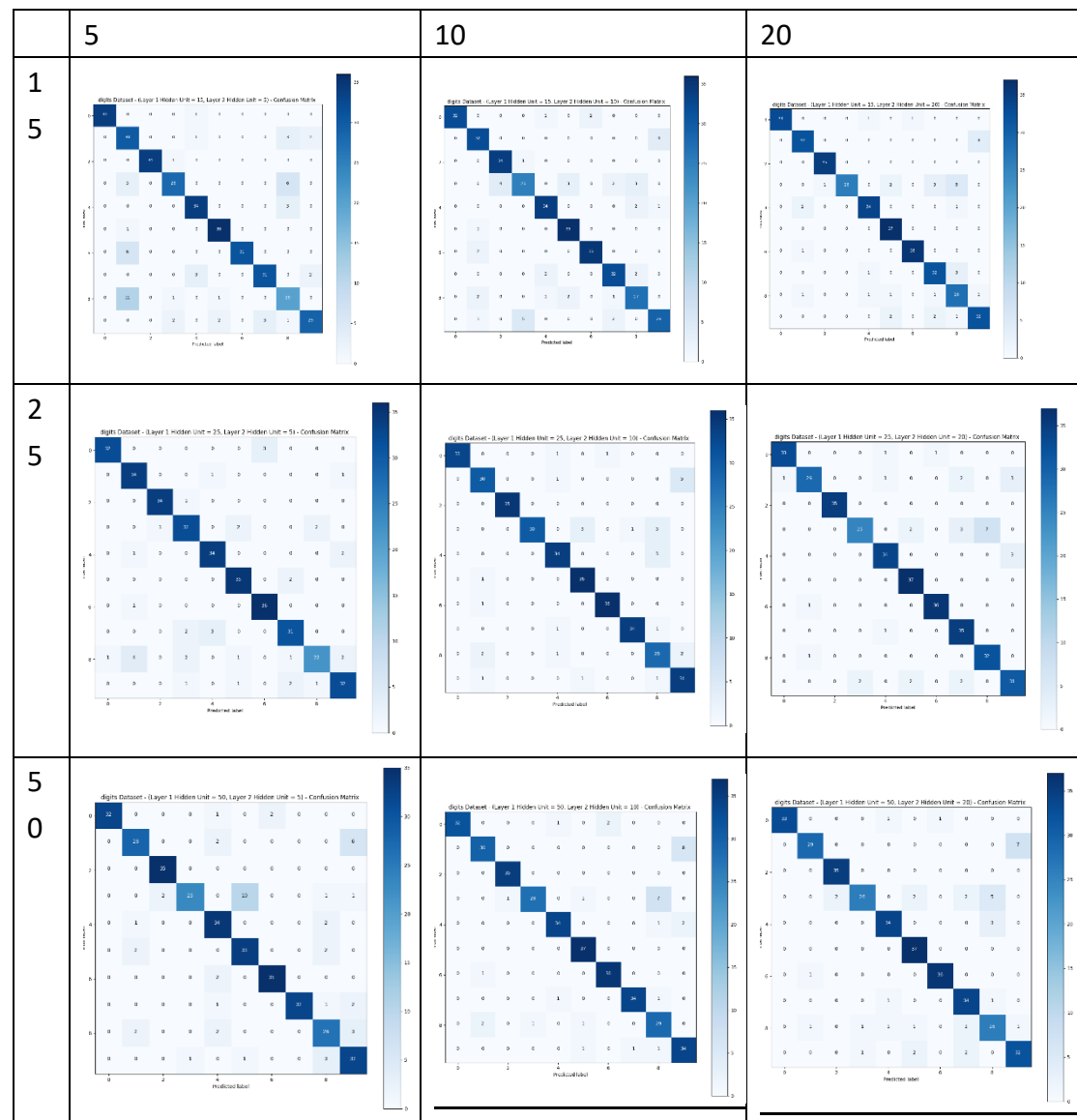
LR 0.1, B 100, E 2000

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.

```
digits Best Multi Classification Model
Best Hidden Unit: Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10
Best Train (By Validation) Accuracy: 98.95833333333333
Best Test (By Test Set) Accuracy: 91.66666666666667
```

```
[Final Best Train (By Validation)] Model (Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10): Loss= 4.4611 Best Accuracy = 98.9583
[Final Best Test (By Test Set)] Model (Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10): Loss= 6.2735 Best Accuracy = 91.6667
Cross Validation Details:
Fold 0: 98.61111111111111 %
Fold 1: 98.95833333333333 %
Fold 2: 96.16724738675958 %
Fold 3: 98.25783972125436 %
Fold 4: 97.90940766550523 %
Average: 97.98078784359272 %
```



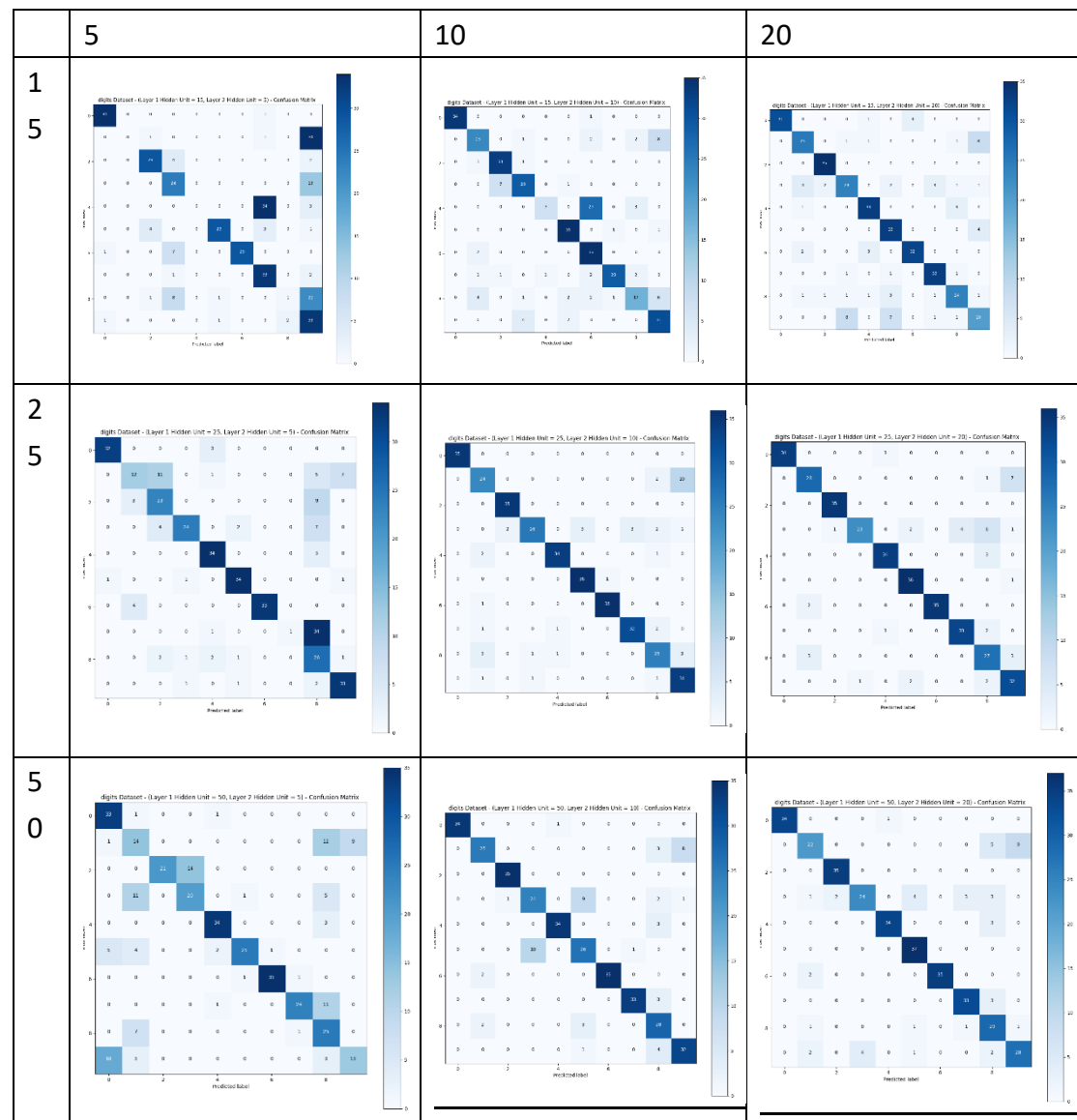
LR 0.001, B 100, E 2000

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.

```
digits Best Multi Classification Model
Best Hidden Unit: Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10
Best Train (By Validation) Accuracy: 88.50174216027875
Best Test (By Test Set) Accuracy: 88.05555555555556
```

```
[Final Best Train (By Validation)] Model (Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10): Loss= 4.7260 Best Accuracy = 88.5017
[Final Best Test (By Test Set)] Model (Layer 1 Hidden Unit = 25, Layer 2 Hidden Unit = 10): Loss= 6.4736 Best Accuracy = 88.0556
Cross Validation Details:
Fold 0: 79.51388888888889 %
Fold 1: 81.94444444444444 %
Fold 2: 79.79094076655052 %
Fold 3: 88.50174216027875 %
Fold 4: 84.3205574912892 %
Average: 82.81431475029036 %
```



LR 1, B 100, E 100

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

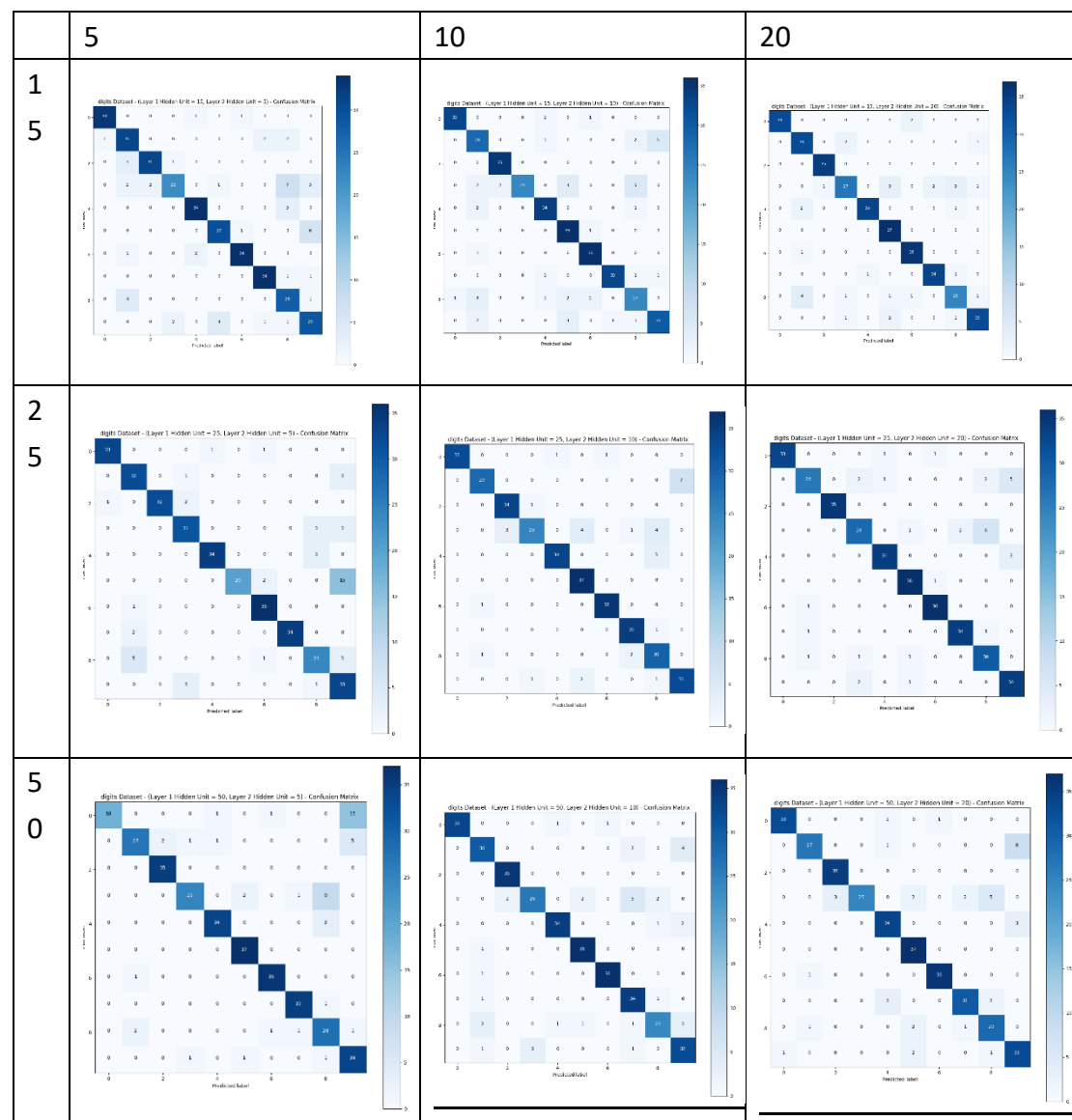
Program output, we found the best model with layer 1 with 25 hidden unit, and layer 2 with 10 hidden unit, with accuracy 91.6.

```

digits Best Multi Classification Model
Best Hidden Unit: Layer 1 Hidden Unit = 15, Layer 2 Hidden Unit = 20
Best Train (By Validation) Accuracy: 98.60627177700349
Best Test (By Test Set) Accuracy: 90.83333333333333
  
```

```

[Final Best Train (By Validation)] Model (Layer 1 Hidden Unit = 15, Layer 2 Hidden Unit = 20): Loss= 4.4399 Best Accuracy = 98.6063
[Final Best Test (By Test Set)] Model (Layer 1 Hidden Unit = 15, Layer 2 Hidden Unit = 20): Loss= 6.1833 Best Accuracy = 90.8333
Cross Validation Details:
Fold 0: 97.56944444444444 %
Fold 1: 97.22222222222223 %
Fold 2: 98.60627177700349 %
Fold 3: 97.5609756097561 %
Fold 4: 98.60627177700349 %
Average: 97.91303716608596 %
  
```



For any details or original output, 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.