Tianchu Li
001826214
https://github.com/WilliamLi822/EECE5644_TianchuLi_Homework/tree/master/Exam2

Exam

Question 1

1. Dataset Distribution

(1) Train Data

a. 100 samples

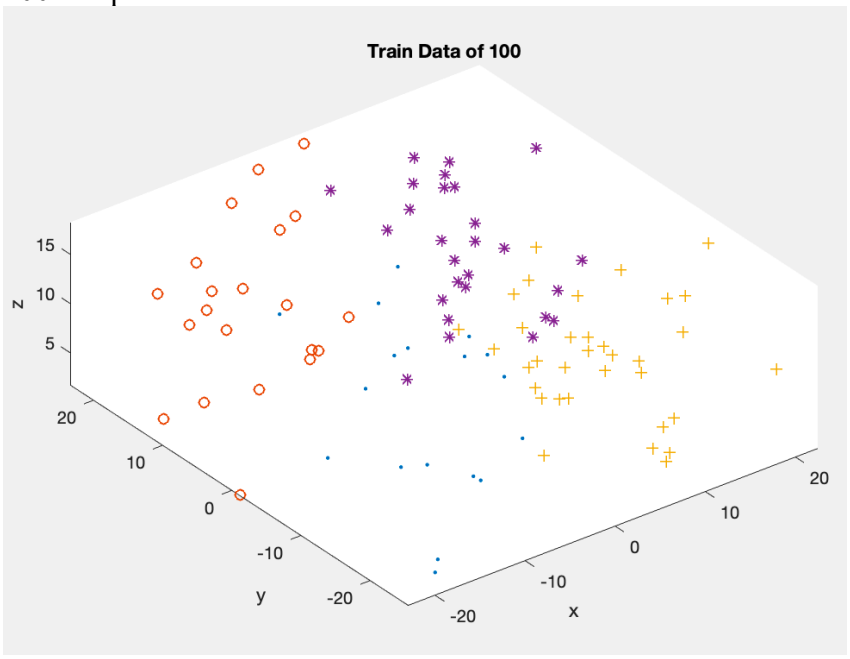

Figure 1. The distribution of 100 samples train data

b. 1000 samples

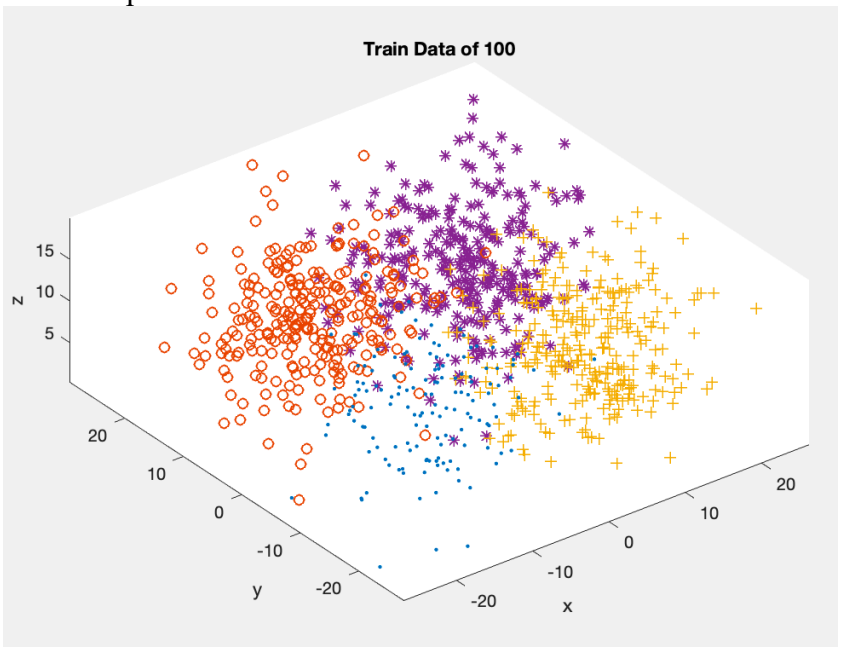

Figure 2. The distribution of 1000 samples train data

c. 10000 samples

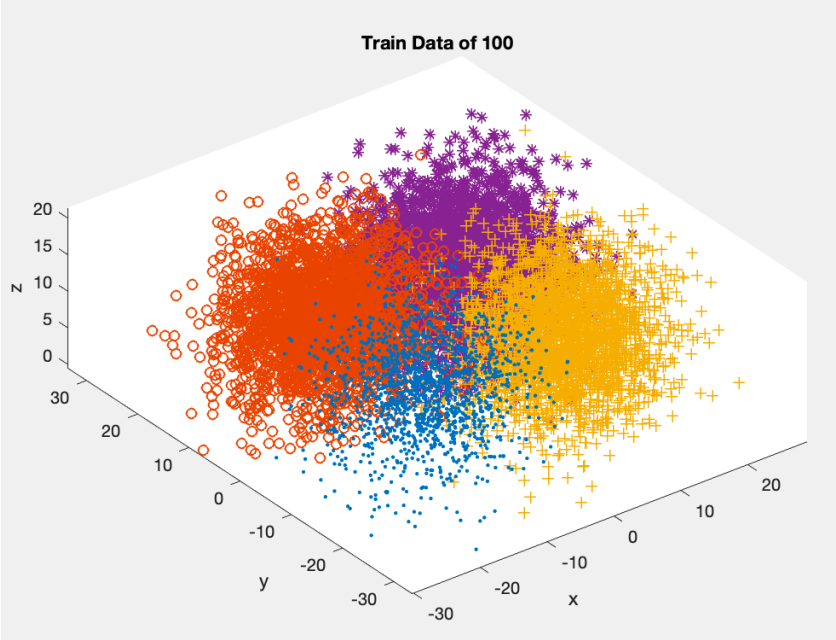

Figure 3. The distribution of 10000 samples train data

(2) Test Data

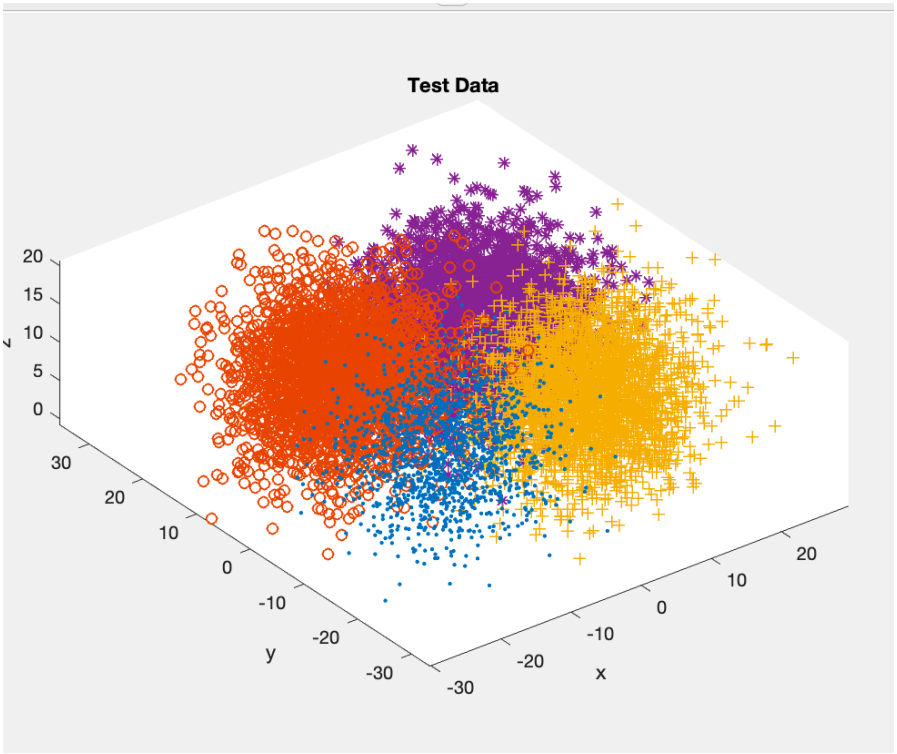

Figure 4. The distribution of 10000 samples test data

2. Data Feature

(1) Train Data

a. 100 samples

Table 1
The feature of 100 samples train data

Class	Class 1	Class 2	Class 3	Class 4
Probability	0.15	0.25	0.30	0.30
Actual number	16	25	28	31
Probability of error by MAP classifier	0.125	0.12	0.0714	0.129

```
ConfusionMatrix =

    14     2     0     0
     1    22     1     1
     0     2    26     0
     0     2     2    27
```

b. 1000 samples

Table 2
The feature of 1000 samples train data

Class	Class 1	Class 2	Class 3	Class 4
Probability	0.15	0.25	0.30	0.30
Actual number	145	246	323	286
Probability of error by MAP classifier	0.0828	0.0894	0.0619	0.059

```
ConfusionMatrix =

   133     6     6     0
    12   224     7     3
     4     8   303     8
     0     6    10   270
```

c. 10000 samples

Table 3
The feature of 10000 samples train data

Class	Class 1	Class 2	Class 3	Class 4
Probability	0.15	0.25	0.30	0.30
Actual number	1504	2454	3014	3028
Probability of error by MAP classifier	0.0559	0.0656	0.0617	0.0641

```
ConfusionMatrix =

        1420          50          34           0
          62        2293          83          16
          13          83        2828          90
           0          78         116        2834
```

(2) Test Data

Table 4
The feature of 10000 samples test data

Class	Class 1	Class 2	Class 3	Class 4
Probability	0.15	0.25	0.30	0.30
Actual number	1517	2475	3039	2969
Probability of error by MAP classifier	0.0587	0.0634	0.0628	0.0606

```
ConfusionMatrix =

        1428          62          27           0
          62        2318          85          10
          15         108        2848          68
           0          76         104        2789
```

3. Train Model & Parameters Selection

The model contains two layers of perceptrons, the math equation is

$$Y=W2(W1X + B1) + B2$$

And the structure of this model is as shown in figure 5.

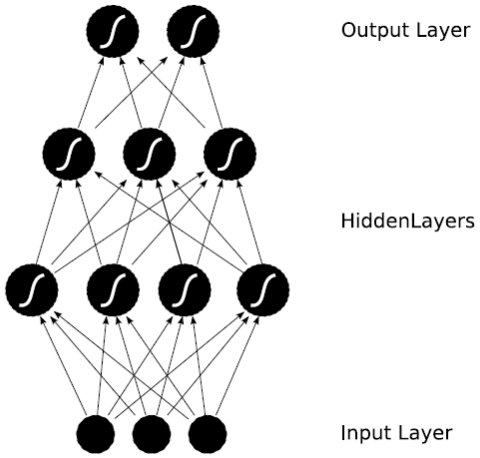

Figure 5. The structure of MLP model

After building the model, k-fold (k=10) is used to do the cross-validation to choose the best performance of different number of perceptron in a single layer. The selection range is from 1 to 10. The activation function here use softplus and SGD is used to optimize the parameters with learning rate = 0.01. And the loss function is categorical cross-entropy Figure 6 to 8 show the probability of correct decisions based on different datasets.

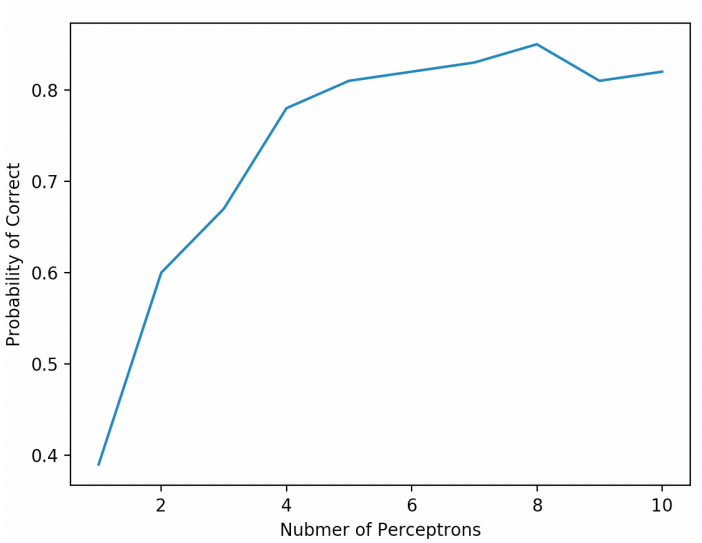

Figure 4. 10-fold model selection on 100 samples train dataset

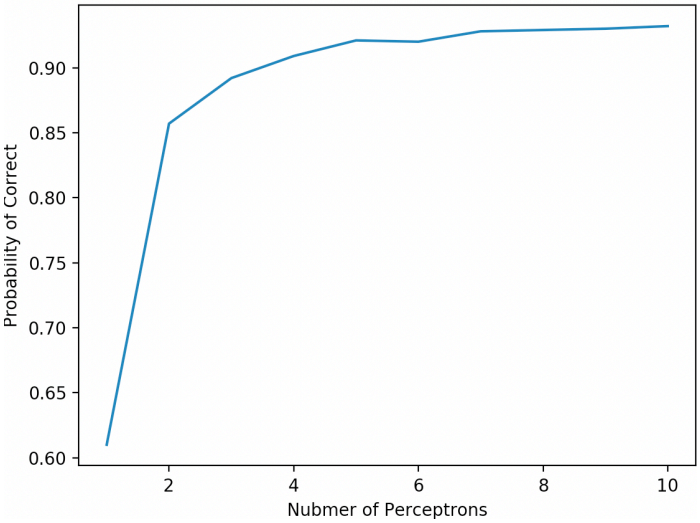

Figure 4. 10-fold model selection on 1000 samples train dataset

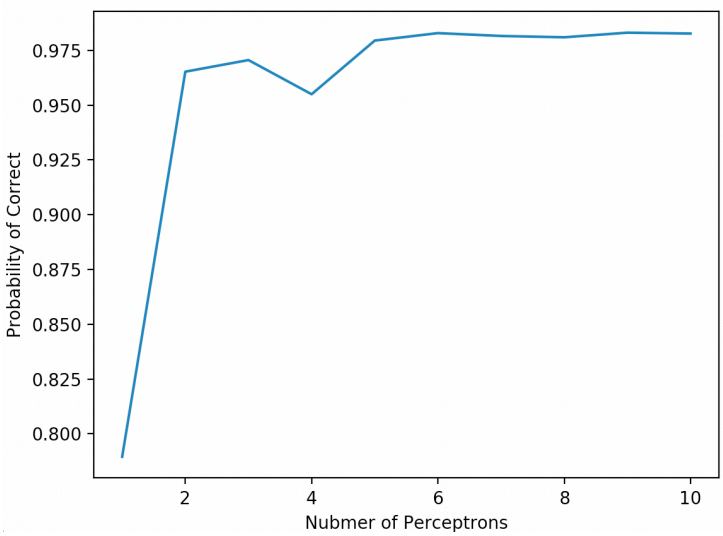

Figure 4. 10-fold model selection on 10000 samples train dataset

From the changing tendency of probability of correct, the models can be selected as below:

(1) 100 samples train dataset
Number of perceptron = 8
Then run the model on 10000 samples test dataset

```
Probability of Correct on test dataset with best model:
0.9105

Probability of error of test dataset:
[0.13971050975456267, 0.09003601440576225, 0.06641816333446293, 0.08510638297872342]

Confusion matrixof test dataset:
[[1367   84   65    1]
 [ 166 2274   16   19]
 [  45    7 2755  232]
 [  11  134  115 2709]]
```

(2) 1000 samples train dataset
Number of perceptron = 10
Then run the model on 10000 samples test dataset

```
Probability of Correct on test dataset with best model:
0.9317

Probability of error of test dataset:
[0.11667765326301915, 0.088006230529595, 0.06681657565049792, 0.025695931477516032]

Confusion matrixof test dataset:
[[1340  150   27    0]
 [ 118 2342    4   11]
 [  57   16 2905   61]
 [   2   60  177 2730]]
```

(3) 10000 samples train dataset
Number of perceptron = 9
Then run the model on 10000 samples test dataset

```
Probability of Correct on test dataset with best model:
0.9895

Probability of error of test dataset:
[0.0401785714285714, 0.00987654320987652, 0.003956478733926794, 0.002020882452004047

Confusion matrixof test dataset:
[[1505   12    0    0]
 [  63 2406    6    0]
 [   0   12 3021    6]
 [   0    0    6 2963]]
```

From result of model on test dataset, we can see that MLP classifiers performs better than or at least similar with MAP classifiers. Besides, with more perceptrons in a single layer, the probability of correct also become higher. But in turn, the time of training the model will also increase dramatically.

Question 2

1. Train Model & Parameters Selection

The model contains two layers of perceptrons, the math equation is

$$X2=W2(W1X1 + B1) + B2$$

And the structure of this model is as shown in figure 5.

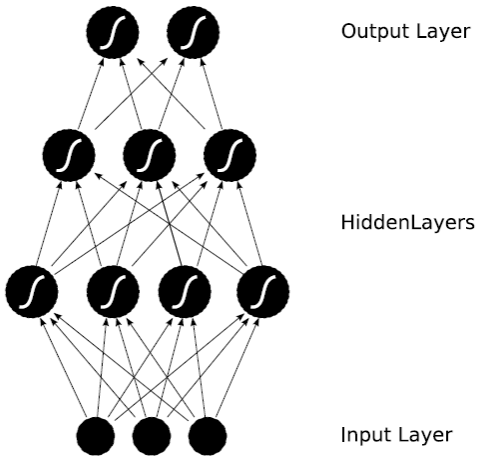

Figure 5. The structure of MLP model

After building the model, k-fold (k=10) is used to do the cross-validation to choose the best performance of different number of perceptron and different activation function. The selection range is from 1 to 10. The alternative activation functions are softplus and sigmoid

Adam is used to optimize the parameters. And the loss function is mean squared error. Figure 9 show the training results.

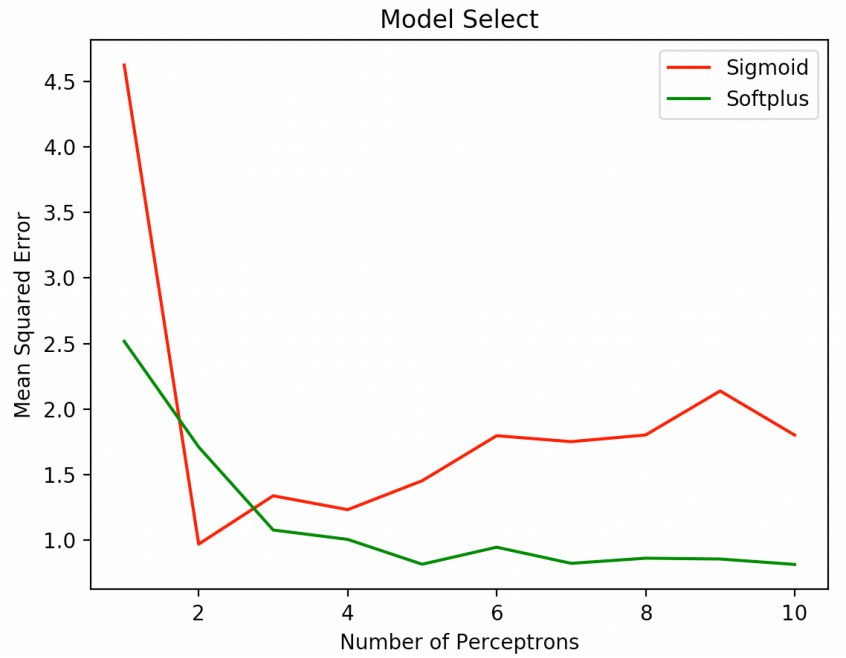

Figure 9. 10-fold model selection on 1000 samples train dataset

According to the result, the model can be selected as below:

```
the number of perceptron of best perform model:  9
the activation function of best perform model:  softplus
```

Run this model on the test dataset, the MSE will be:

```
MSE of test dataset with best model:  0.6266267706871033
```

Then use this model to predict the distribution of X2 from the X1 of test dataset and compare it with the true distribution of X2.

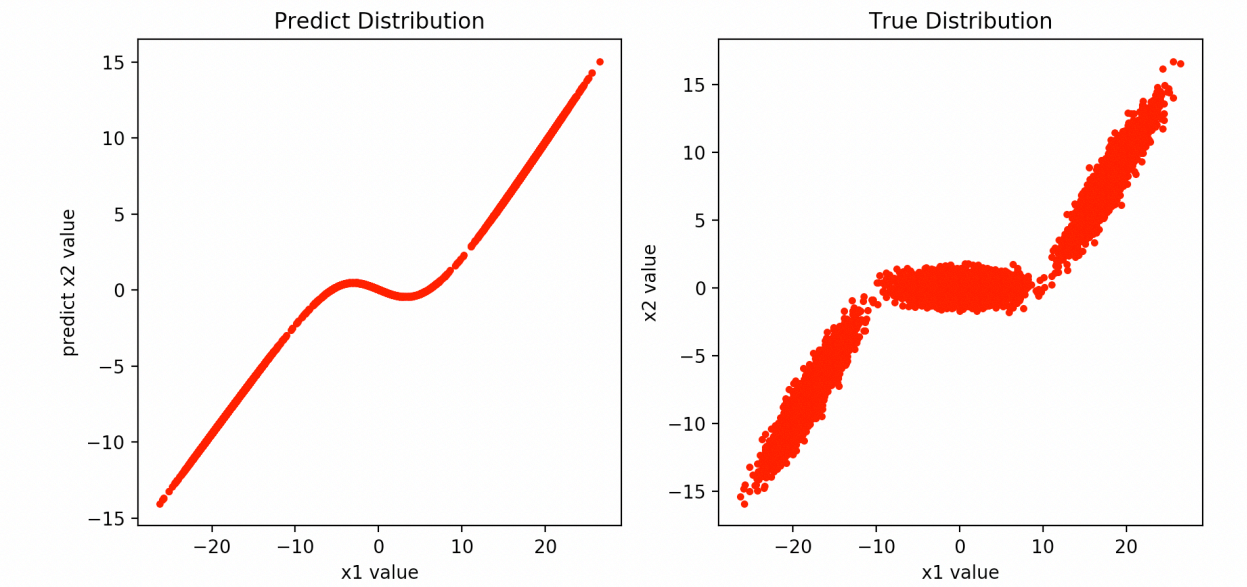

Figure 10. The comparison of predicted and true distribution of (x1, x2)