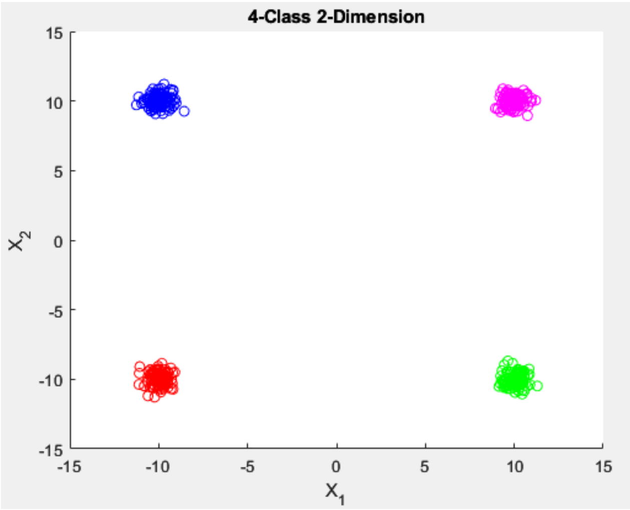
111061702 常安彥

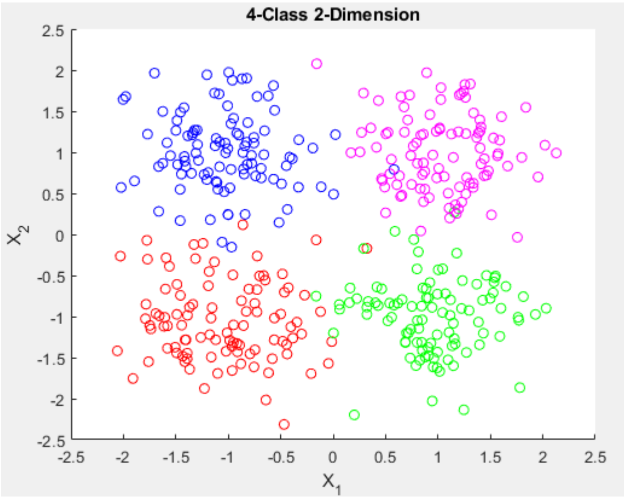
Pattern Recognition ex3

1. Scatter Matrices and Criterion J3

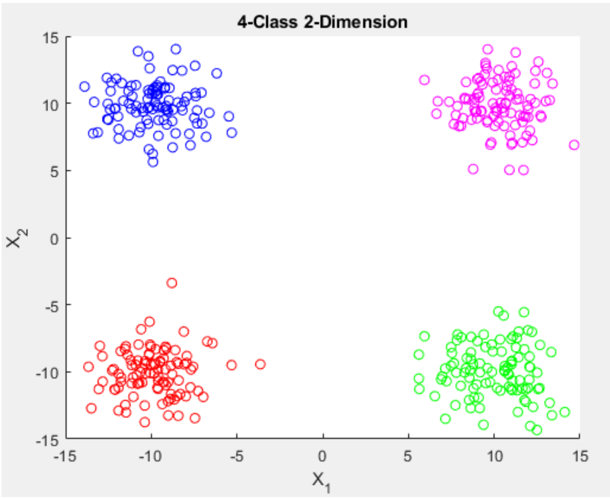
(a) [-10 -10; -10 10; 10 -10; 10 10], 0.2 eye(2)



(b) [-1 -1; -1 1; 1 -1; 1 1], 0.2 eye(2)



(c) [-10 -10; -10 10; 10 -10; 10 10], 1.0 eye(2)



Results:

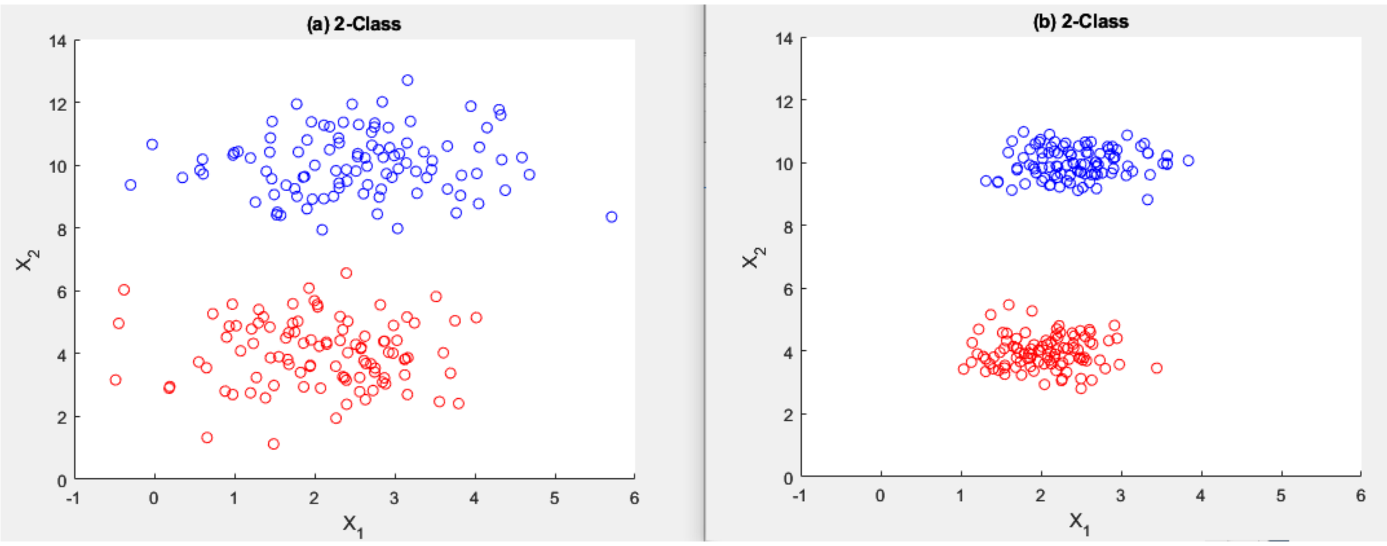
```
>> ex3_1
(a)
Sw: [0.207183 0.00675828; 0.00675828 0.202538]
Sb: [99.7442 -0.42041; -0.42041 100.136]
Sm: [99.9514 -0.413652; -0.413652 100.338]
J3: 979.035
(b)
Sw: [0.21784 -0.0156265; -0.0156265 0.218595]
Sb: [1.02527 0.0381738; 0.0381738 1.01658]
Sm: [1.24311 0.0225472; 0.0225472 1.23518]
J3: 11.4305
(c)
Sw: [3.19029 -0.106465; -0.106465 3.16818]
Sb: [99.6931 1.31687; 1.31687 98.701]
Sm: [102.883 1.21041; 1.21041 101.869]
J3: 64.5006
```

Class a's within-class spread is small and between-class spread is large, so it has the largest value.

Class c's between-class spread is large, but within-class spread is larger than class a, so it has a smaller value.

Class b is the worst with large within-class spread and small between-class spread.

2. FDR



Results:

```
>> ex3_2
(a):
FDR1: 0.0978121
FDR2: 1.78363
(b):
FDR1: 0.29577
FDR2: 1.93878
```

Fisher's discriminant ratio is used to quantify the separability capabilities of individual features.

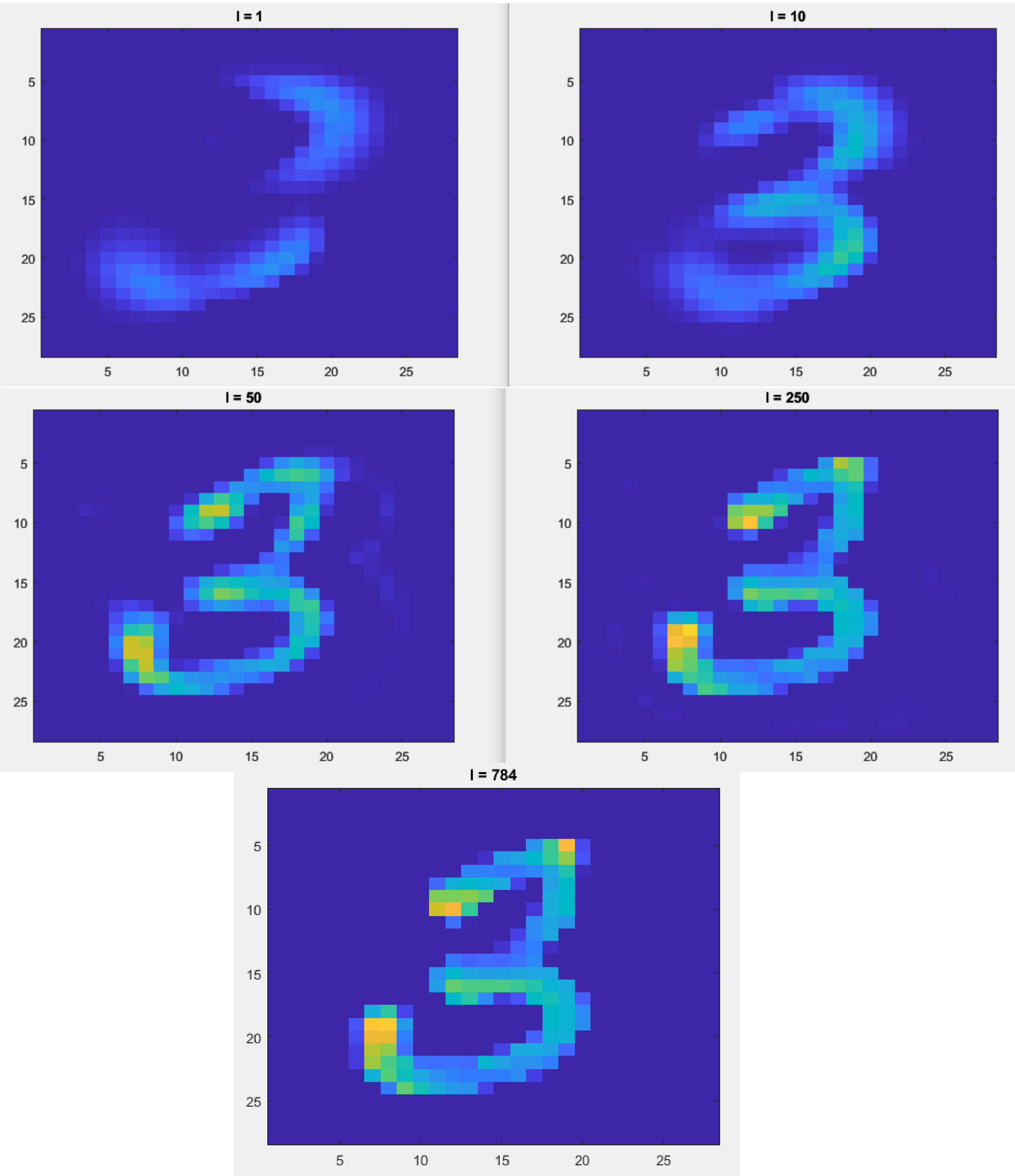
First, let's compare the features, X1 and X2. The FDR of X2 is always larger than the FDR of X1. In both classes, the distribution of X2 is much tighter than X1, and it is obviously observed through above images.

Next, let's compare between two datasets. Both features got higher FDR score in the second dataset, which has a smaller covariance matrix. The distribution is obviously tighter than the first dataset's, and therefore easier to separate.

3. MNIST

When l=10, it is obvious already.

When l = 50, there’s barely different from the origin image.



Result

```
>> ex3_3
N = 500, l = 1, MSE = 7477.673918
N = 500, l = 10, MSE = 2193.390327
N = 500, l = 50, MSE = 755.167479
N = 500, l = 250, MSE = 122.762091
N = 500, l = 784, MSE = 0.000000
N = 1000, l = 1, MSE = 7463.150634
N = 1000, l = 10, MSE = 2056.518439
N = 1000, l = 50, MSE = 677.292931
N = 1000, l = 250, MSE = 88.231810
N = 1000, l = 784, MSE = 0.000000
N = 1500, l = 1, MSE = 7514.999442
N = 1500, l = 10, MSE = 1949.039542
N = 1500, l = 50, MSE = 630.867404
N = 1500, l = 250, MSE = 70.498335
N = 1500, l = 784, MSE = 0.000000
N = 2000, l = 1, MSE = 7347.916712
N = 2000, l = 10, MSE = 1946.028734
N = 2000, l = 50, MSE = 628.985818
N = 2000, l = 250, MSE = 66.289092
N = 2000, l = 784, MSE = 0.000000
```

N:

As the training data (N) increases, the evaluation error decreases when L = 10, 50 or 250. N doesn't affect the result when L = 1 or 784, because L=1 is has little impact, and L=784 is the origin image.

L:

As L increase, the loss significantly decreases when L is still low. There are great impact on the first few coefficients, and made the error decreased a lot.