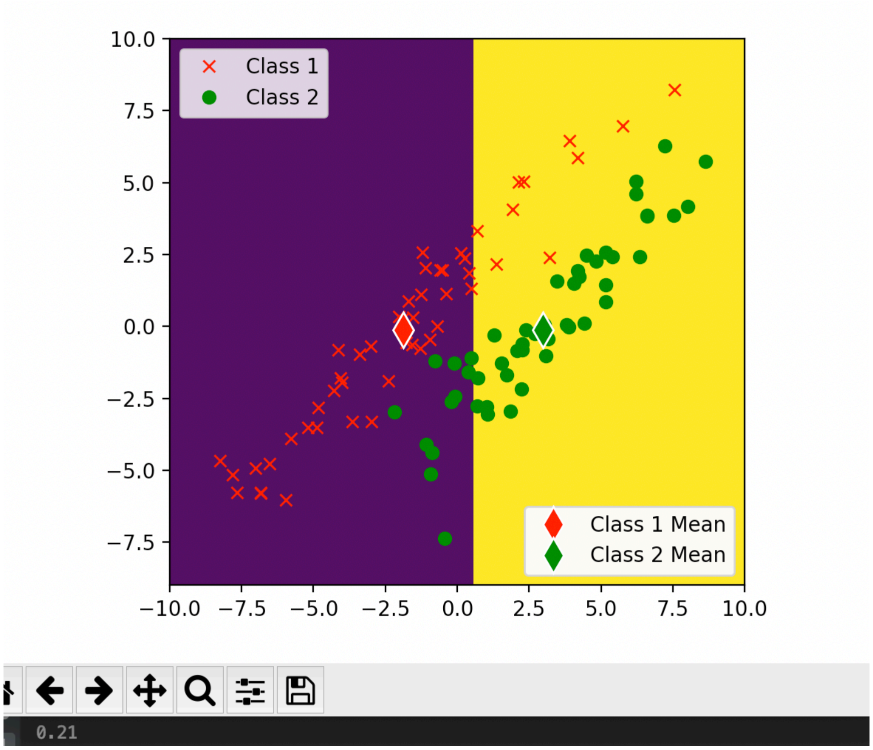
Jian Xu

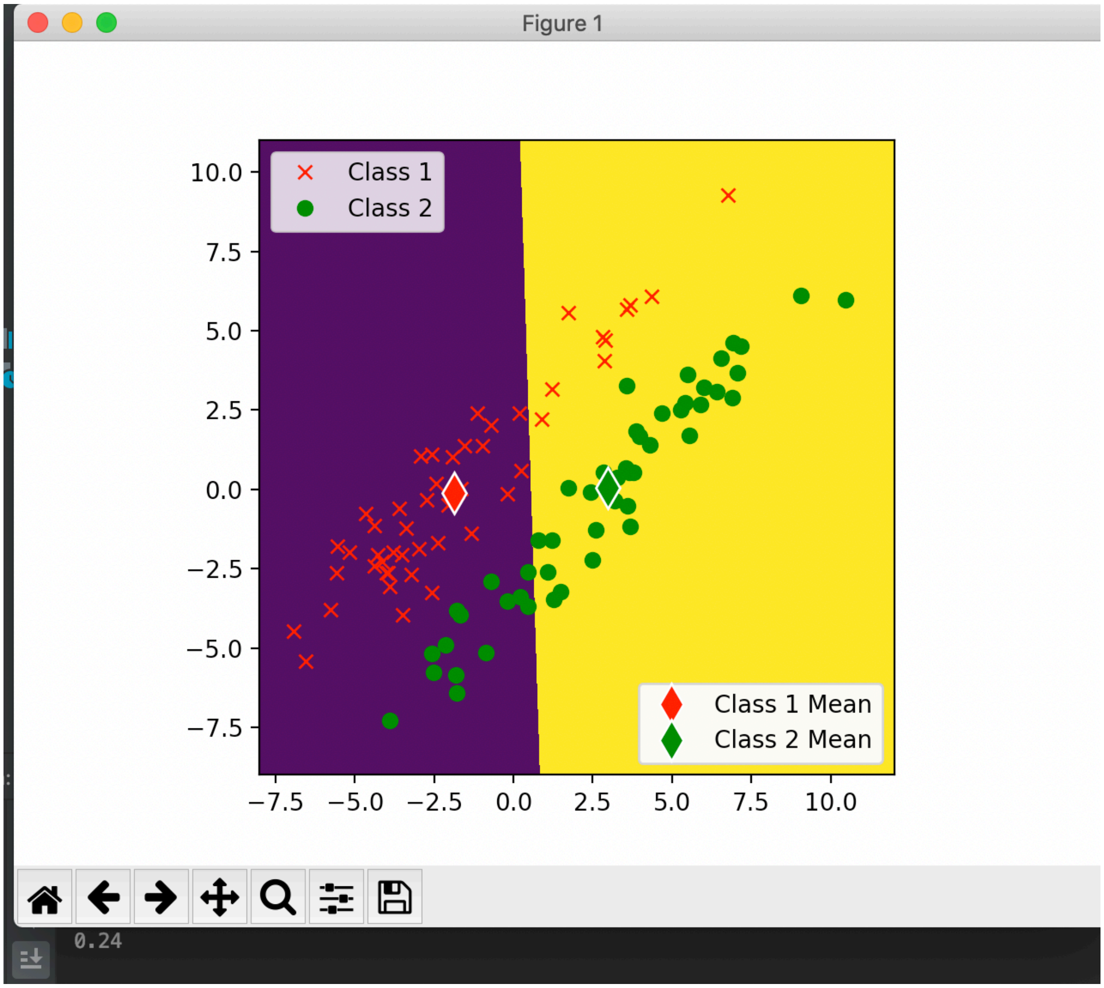
jxu72364@usc.edu

(a)



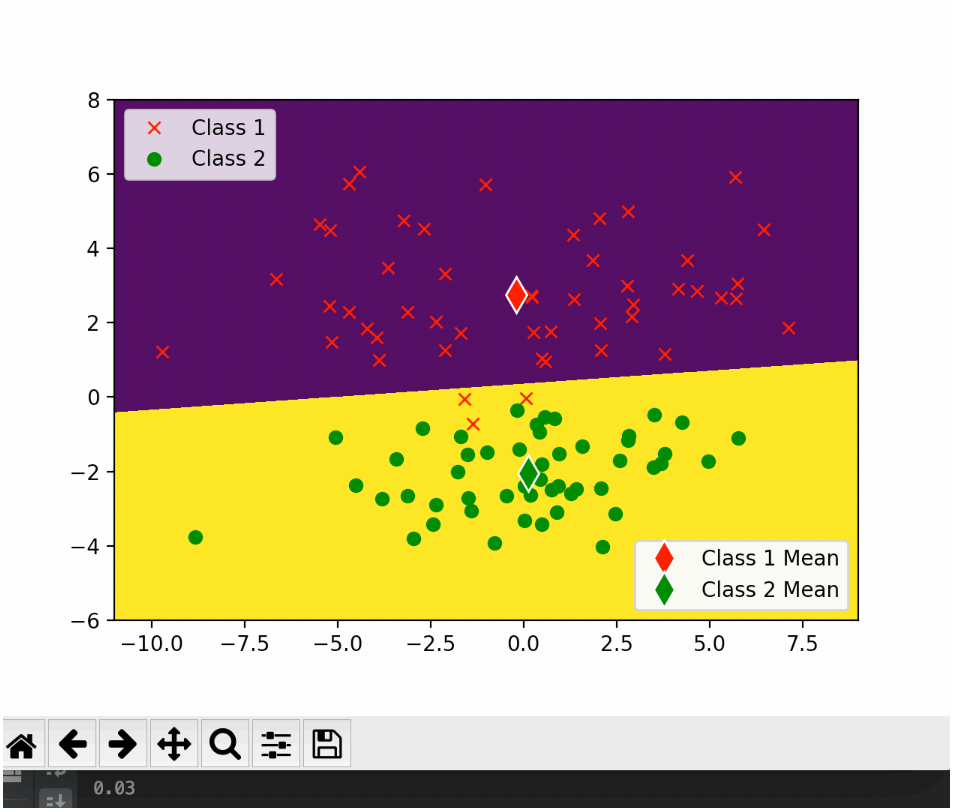
```
0.21
```

the classification error rate on the training set sythetic1_train.csv is 0.21



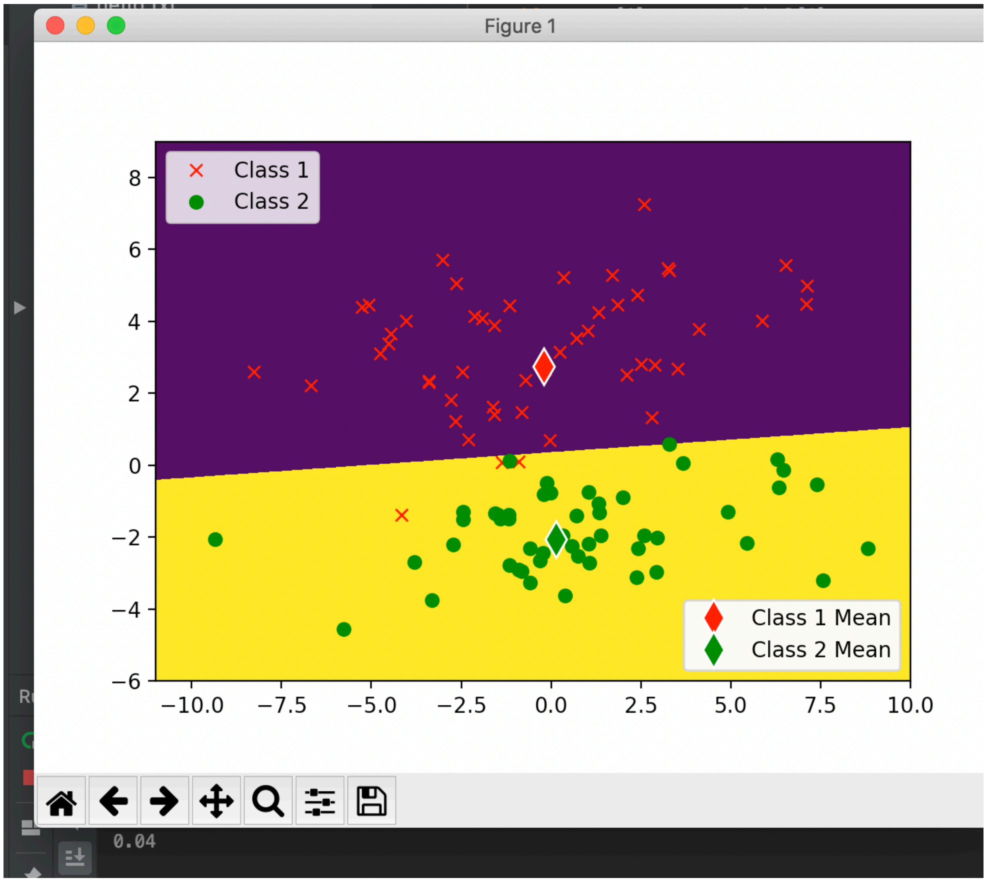
```
0.24
```

the classification error rate on the test set sythetic1_test.csv is 0.24



```
0.03
```

the classification error rate on the training set sythetic2_train.csv is 0.03



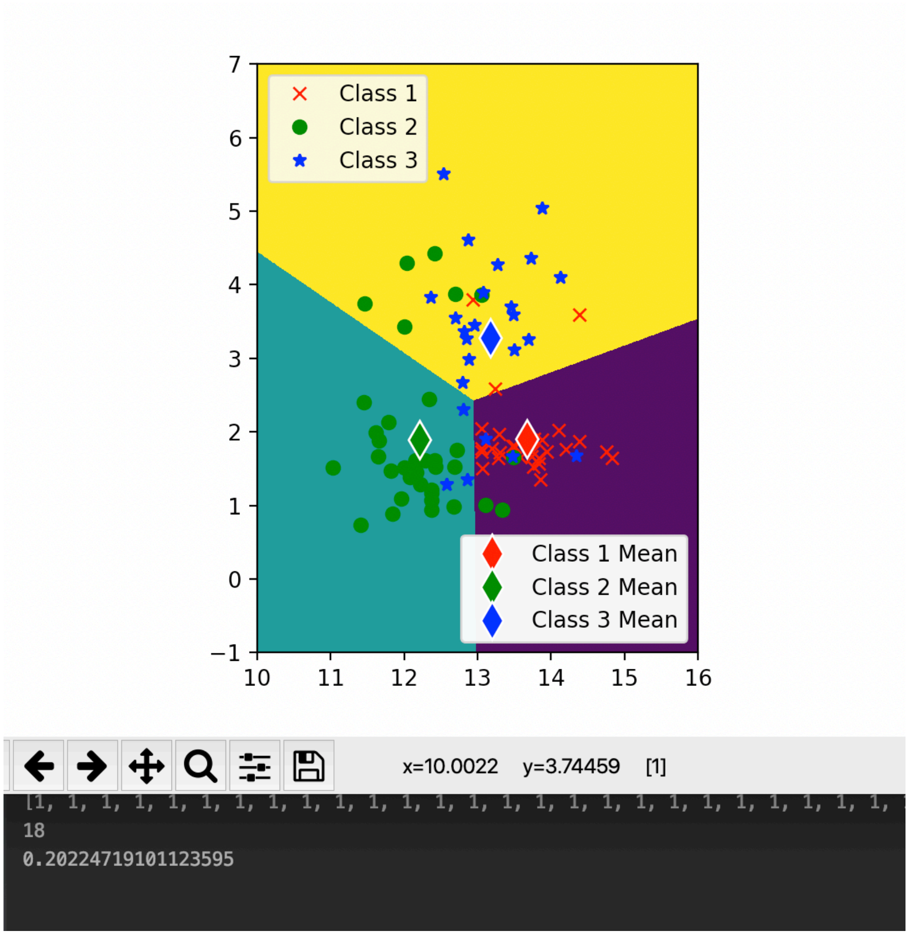
```
0.04
```

the classification error rate on the test set sythetic2_test.csv is 0.04

(b)

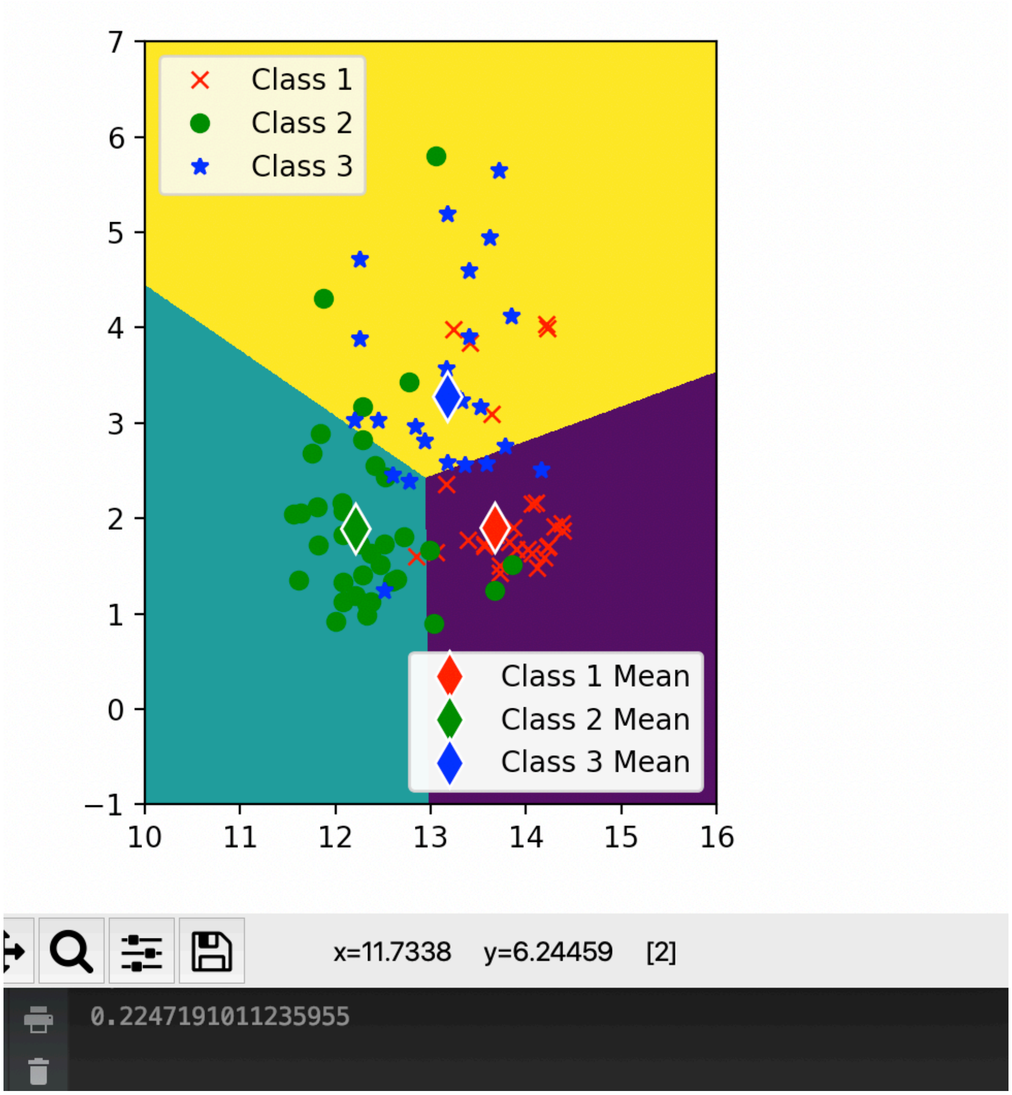
Yes, because the distribution of the datasets is different.
The first reason is that for some datasets, probably it is better to class by linear, but some is not. We cannot always class feature space using linear.
The second reason is that for some data points, we cannot using the means of these points and class by the distance from these average means. Judging by average means may work for some datasets, by not for all datasets
Therefore, we can see that using this method for sythetic2, the error rate is smaller than sythetic1.

(c)



```
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
18
0.20224719101123595
```

Using x1 and x2, the classification error on the training set is about 0.202247



```
x=11.7338   y=6.24459   [2]
```

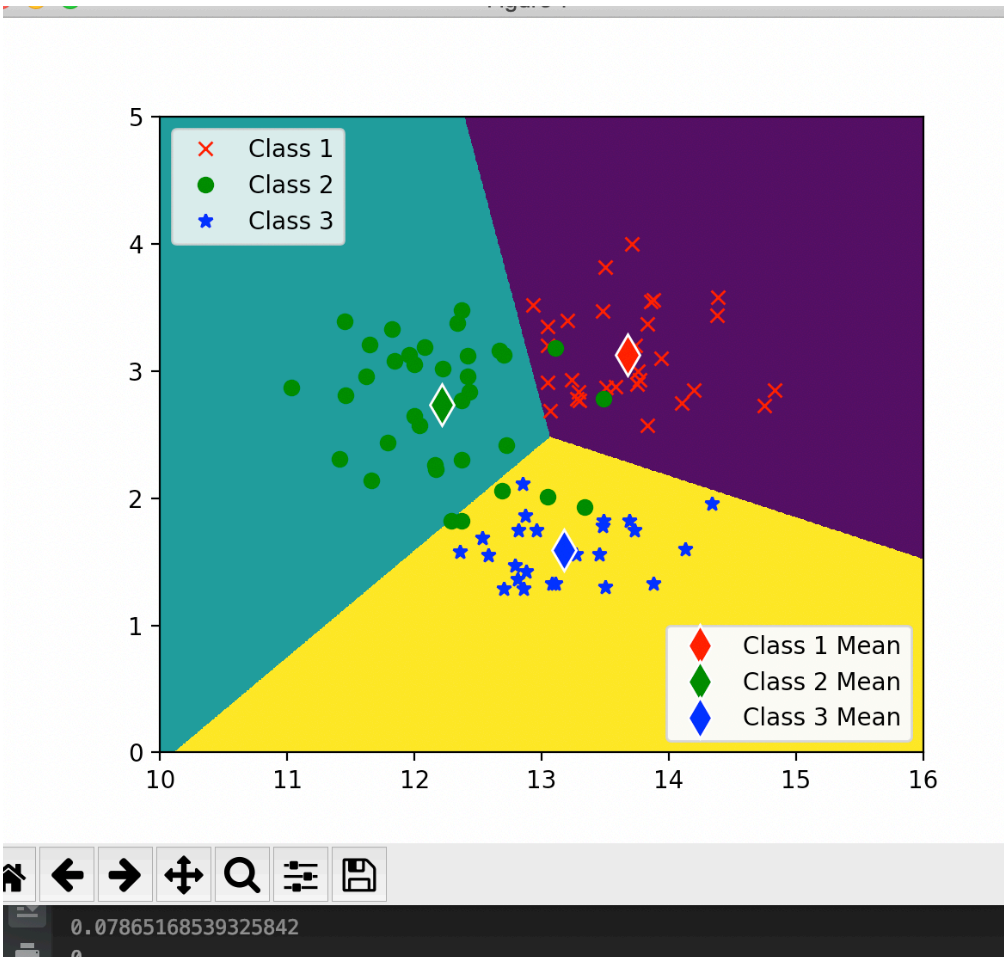
```
0.2247191011235955
```

Using x1 and x2, the classification error on the test set is about 0.22471910

(d)

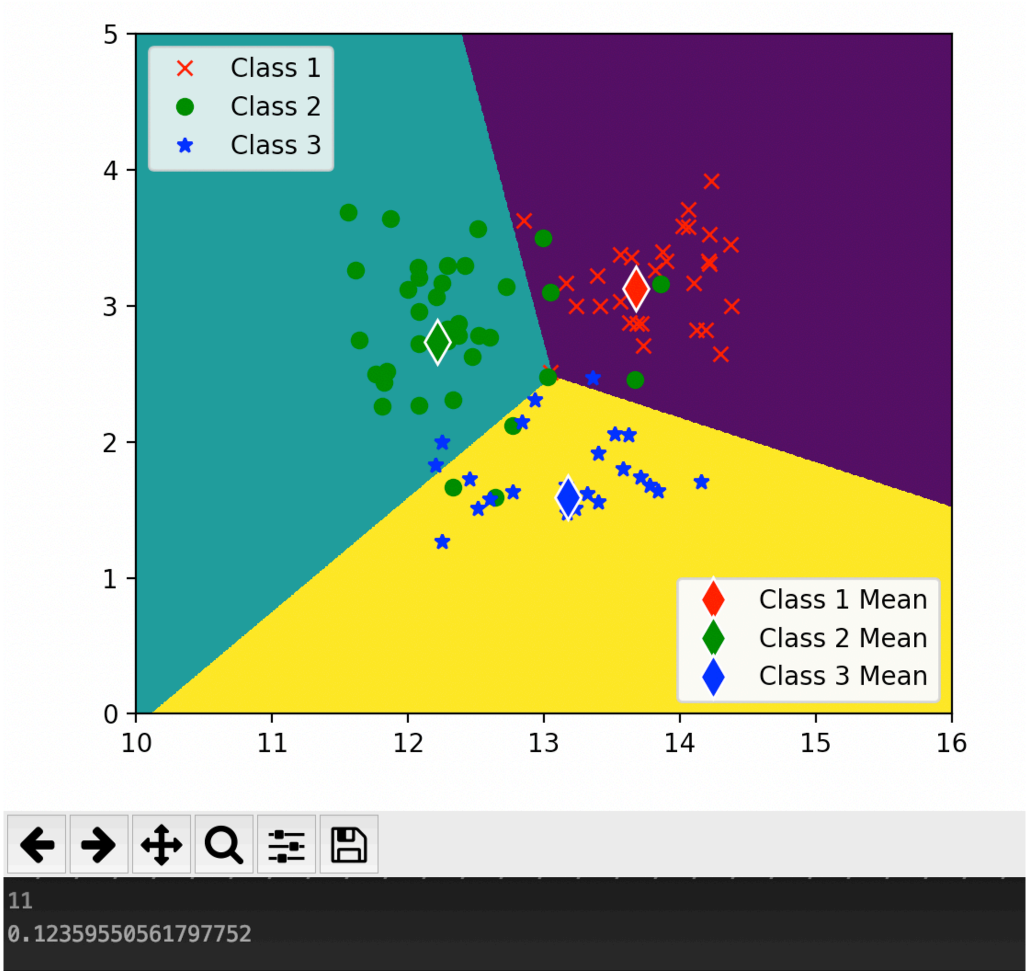
I find that using feature 1 and feature 12, I can get the minimum classification error on the training set.

My method is using “combination”, for 13 features totally 78 combinations, so I traversed all 78 possibilities and found the minimum pair of features is 1 and 12.



```
0.07865168539325842
0
```

Using x1 and x12, the classification error on the training set is about 0.07865



```
11
0.12359550561797752
```

Using x1 and x12, the classification error on the test set is about 0.123596

(e)

Yes, I think there is much difference in both training-set error rate and test-set error rate for different pairs of features.

For training-set I calculated the standard deviation of error rates over all possible pairs of features and found that the standard deviation is 0.1292, but the mean of these errors is about 0.3358. Therefore I think it is much difference in training-set error rate for different pairs of features.

Based on standard deviation 0.1292, and the first part that using x1 and x2 the error rate is 0.22471910 but test-set is 0.123596, I think it is much difference in test-set error rate for different pairs of features.