Synthetic Data Set 1

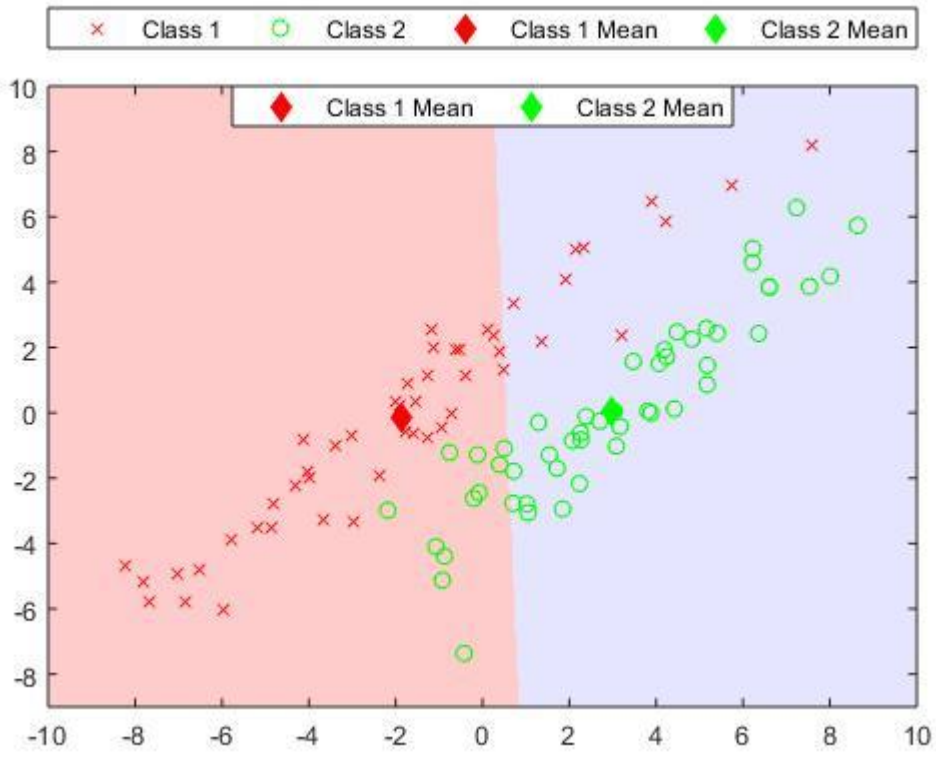

Fig 1. Training Plot

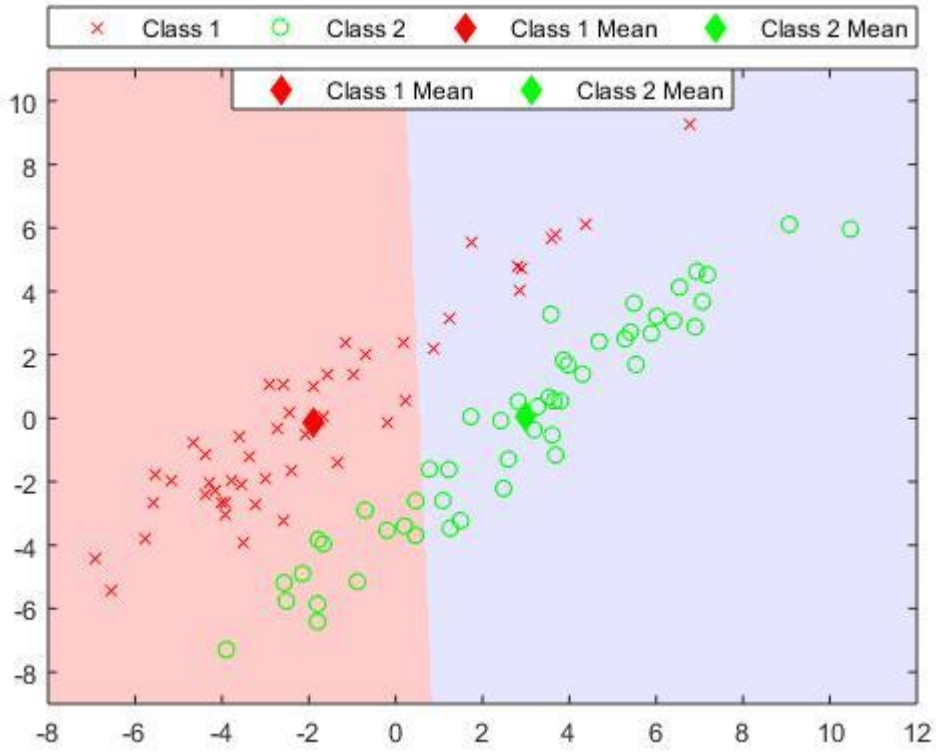

Fig 2. Test Plot

Error Percentage Rates

```
>> ep

ep =

    21              Error Percentage for Training

>> ep1

ep1 =

    24              Error Percentage for Testing
```

Synthetic Data Set 2

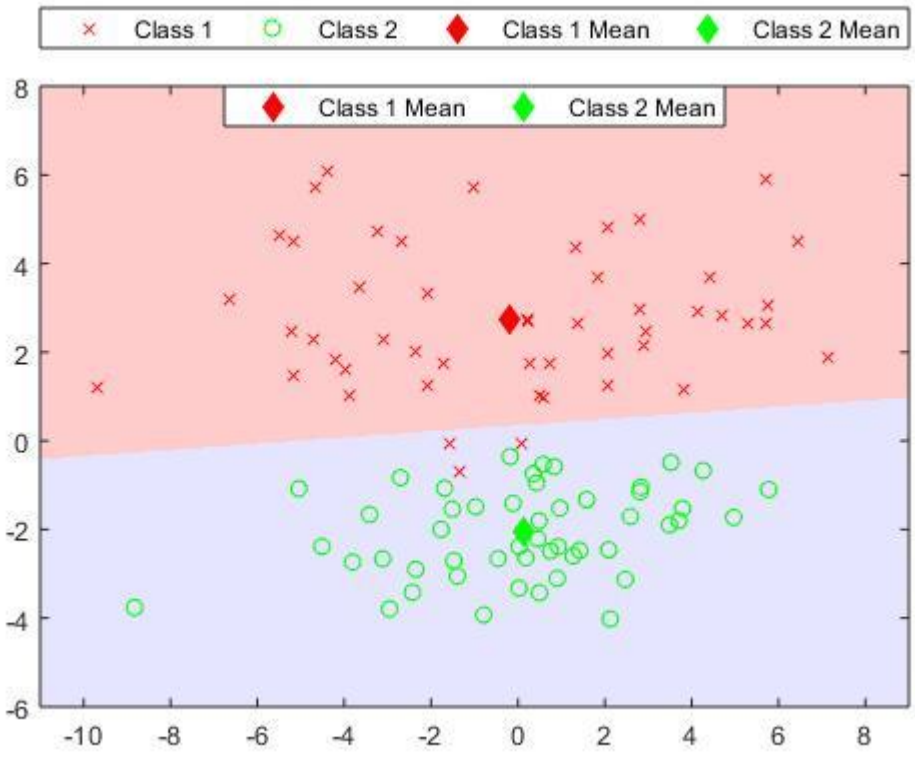

Fig 3. Training Plot

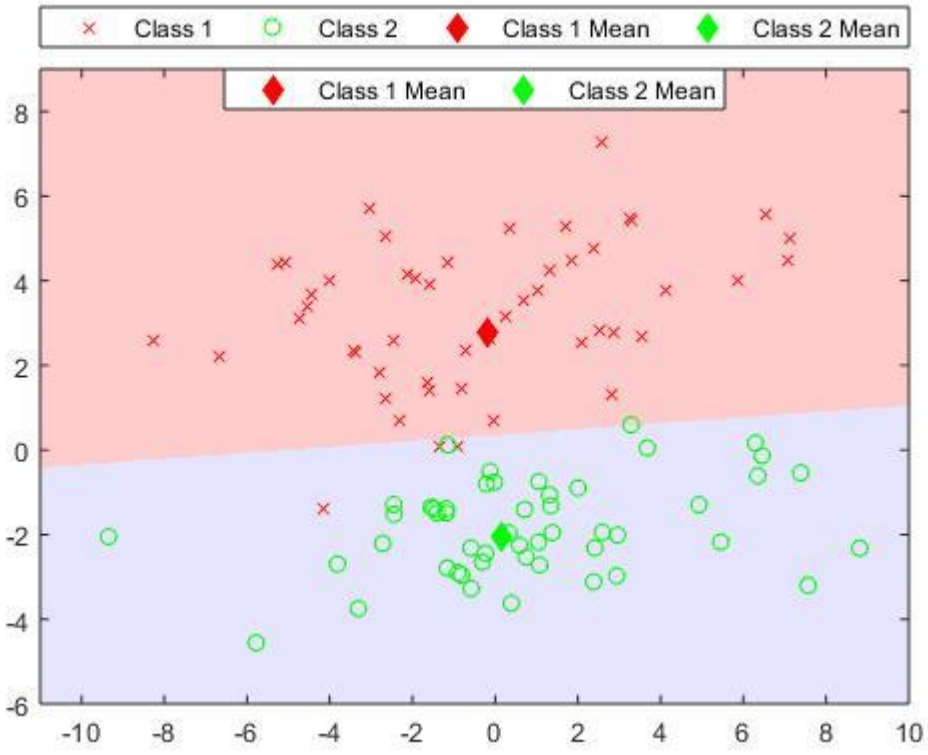

Fig 4. Test Plot

Error Percentage Rates

```
>> ep

ep =

    3              Error Percentage for Training

>> ep1

ep1 =

    4              Error Percentage for Testing
```

ANSWERS:

A. Plotted the graph along with feature space, decision boundary and also mentioned error rate percentage for synthetic data set 1 and 2.

B. Yes, there is much difference in percentage error rate of synthetic data set 1 and 2. This is because more the clustering in feature space, it is difficult to plot the decision boundary. As, we can see that in Synthetic Data set 1, the data points are much closer to each other. So, we get higher error rate. While for Synthetic Data set 2, data points are spread out making it easier to make decision and this gives less error percentage.

Wine Data Set

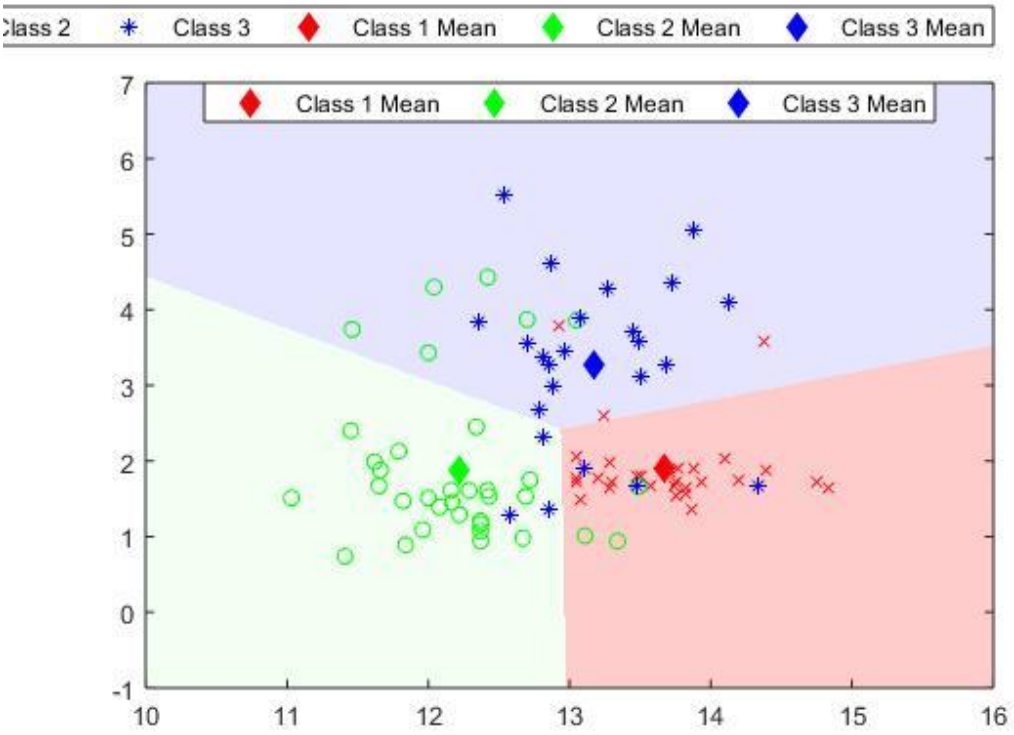

Fig 5. Training Plot

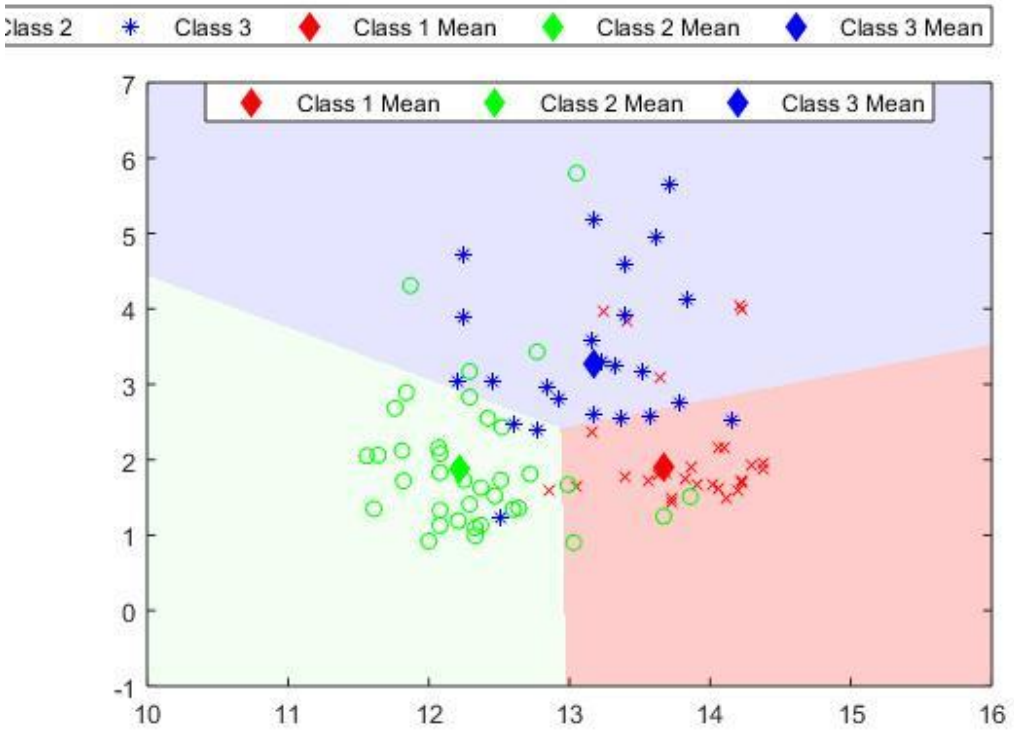

Fig 6. Test Plot

Error Percentage Rates

```
>> ep
ep =
   20.2247                Error Percentage for Training

>> ep1
ep1 =
   22.4719                Error Percentage for Testing

>>
```

ANSWERS:

C. Using Column 1 and column 2 of feature_train and feature_test, I have plotted graph which includes data set points and decision boundary. I have also included error rate percentage.

D. To find best feature among the 13 that we have, I did iterate say k to 1 to 12 and m to k+1 to 13. And applied this nested for loops to the entire code written above. The output of this error rate matrix of 12*13. By analysing the array, I found out that Column 1 and Column 12 gives least error on training data set.
So, using these features for testing data set as well.

Here, are the graphs and error rates for the same.

```
>> ep
ep =
    7.8652
>> ep1
ep1 =
   12.3596
```

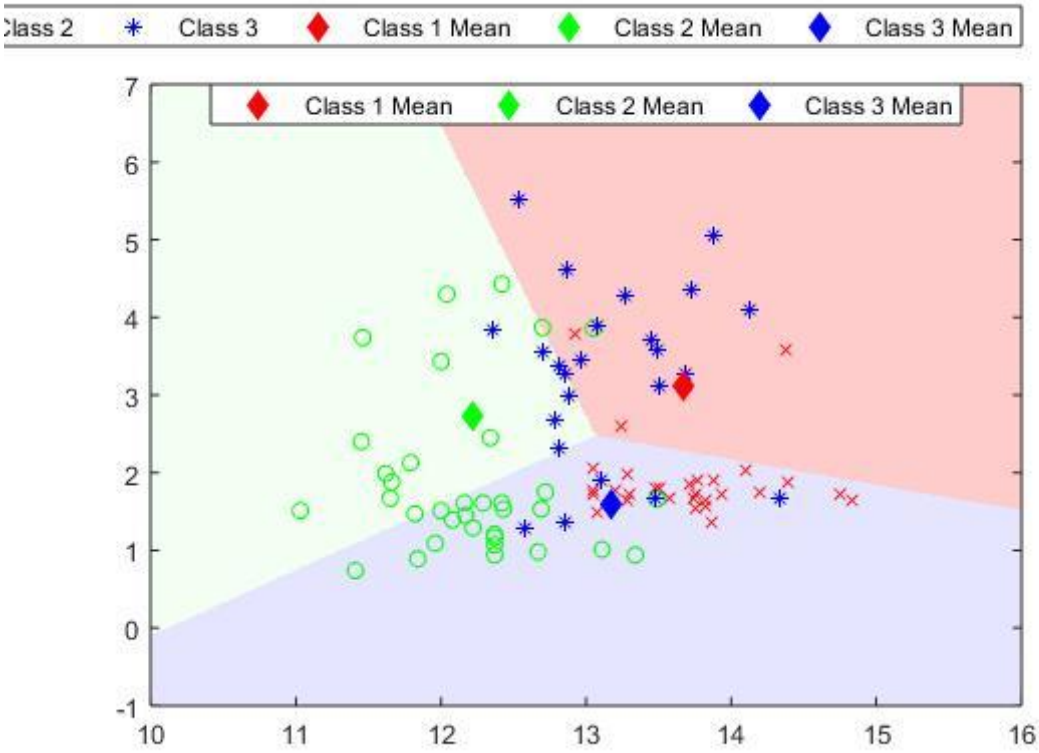

Fig 7. Training Plot

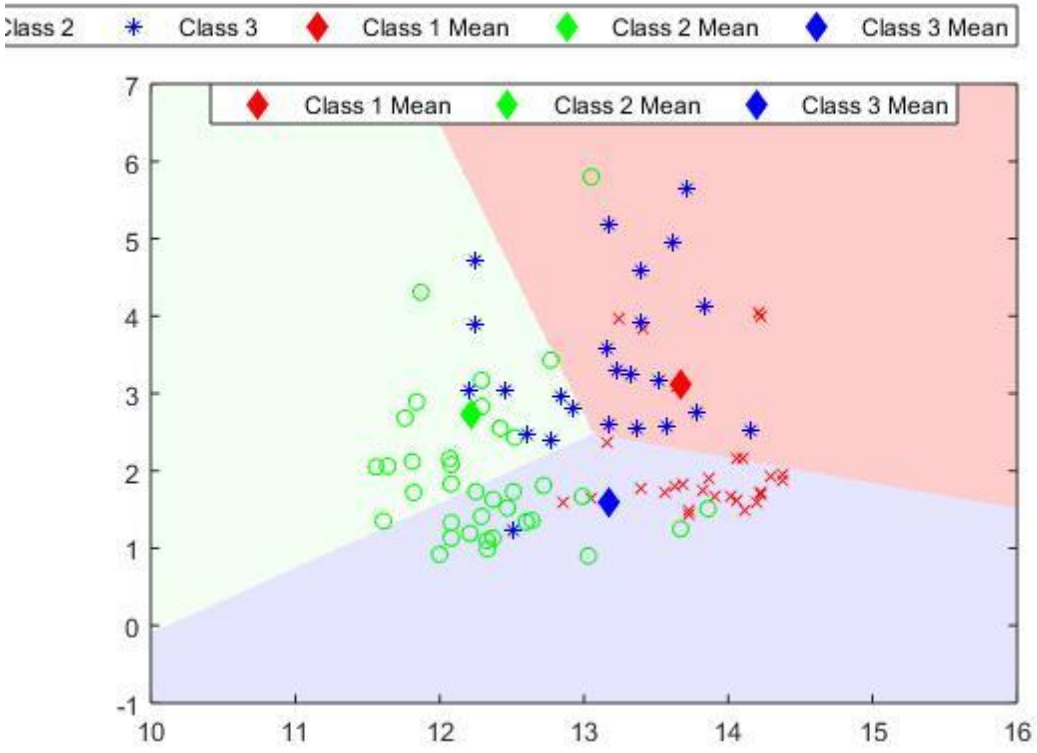

Fig 8. Testing Plot

E. Giving error rates for training data set

Variables - ep

class_check | feature_train | ep

12x13 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	7.8652	20.2247	31.4607	44.9438	56.1798	14.6067	8.9888	33.7079	16.8539	25.8427	25.8427	7.8652	24.7191	
2	0	0	39.3258	39.3258	57.3034	29.2135	20.2247	32.5843	40.4494	24.7191	38.2022	42.6966	24.7191	
3	0	0	0	47.1910	57.3034	32.5843	14.6067	51.6854	38.2022	30.3371	30.3371	29.2135	24.7191	
4	0	0	0	0	42.6966	47.1910	42.6966	47.1910	44.9438	21.3483	47.1910	43.8202	24.7191	
5	0	0	0	0	0	56.1798	56.1798	57.3034	56.1798	49.4382	57.3034	56.1798	24.7191	
6	0	0	0	0	0	0	22.4719	34.8315	40.4494	28.0899	32.5843	24.7191	24.7191	
7	0	0	0	0	0	0	0	16.8539	16.8539	21.3483	15.7303	13.4831	24.7191	
8	0	0	0	0	0	0	0	0	42.6966	30.3371	33.7079	40.4494	24.7191	
9	0	0	0	0	0	0	0	0	0	30.3371	37.0787	34.8315	24.7191	
10	0	0	0	0	0	0	0	0	0	0	30.3371	26.9663	24.7191	
11	0	0	0	0	0	0	0	0	0	0	0	40.4494	24.7191	
12	0	0	0	0	0	0	0	0	0	0	0	0	24.7191	
13														
14														
15														
16														
17														

Fig 9. Training Error Rate

Variables - ep1

class_check | feature_train | ep | ep1

12x13 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	59.5506	22.4719	28.0899	40.4494	44.9438	15.7303	11.2360	28.0899	24.7191	22.4719	26.9663	12.3596	30.3371	
2	0	0	38.2022	42.6966	43.8202	29.2135	23.5955	39.3258	37.0787	22.4719	44.9438	37.0787	30.3371	
3	0	0	0	50.5618	46.0674	28.0899	22.4719	32.5843	40.4494	23.5955	26.9663	28.0899	30.3371	
4	0	0	0	0	38.2022	47.1910	41.5730	50.5618	48.3146	29.2135	50.5618	44.9438	30.3371	
5	0	0	0	0	0	44.9438	41.5730	46.0674	44.9438	43.8202	46.0674	43.8202	30.3371	
6	0	0	0	0	0	0	24.7191	35.9551	32.5843	22.4719	28.0899	25.8427	30.3371	
7	0	0	0	0	0	0	0	24.7191	26.9663	15.7303	24.7191	17.9775	30.3371	
8	0	0	0	0	0	0	0	0	47.1910	23.5955	34.8315	32.5843	30.3371	
9	0	0	0	0	0	0	0	0	0	24.7191	39.3258	31.4607	30.3371	
10	0	0	0	0	0	0	0	0	0	0	23.5955	22.4719	30.3371	
11	0	0	0	0	0	0	0	0	0	0	0	32.5843	30.3371	
12	0	0	0	0	0	0	0	0	0	0	0	0	30.3371	
13														
14														
15														
16														

Fig 10. Testing Error Rate

Yes, as we can see from fig 9 that, there is difference in error rate for the different features.

The error difference depends upon the feature space graph on how the data points are arranged. If there are more clustered, it will account for more error.

The error array for testing data points is also shown in fig 10. Here, too we can see the error difference.