1 Motivation

The programming language is python. The basic algorithm is the nearest centroid mean algorithm. The algorithm initially compute the mean for each class. The mean is a vector which is computed by averaging all features of training data. The algorithm compute the Euclidean distance between vector of mean trained by the classifier and vector of test data.

2 Problem solution

2.1 problem a

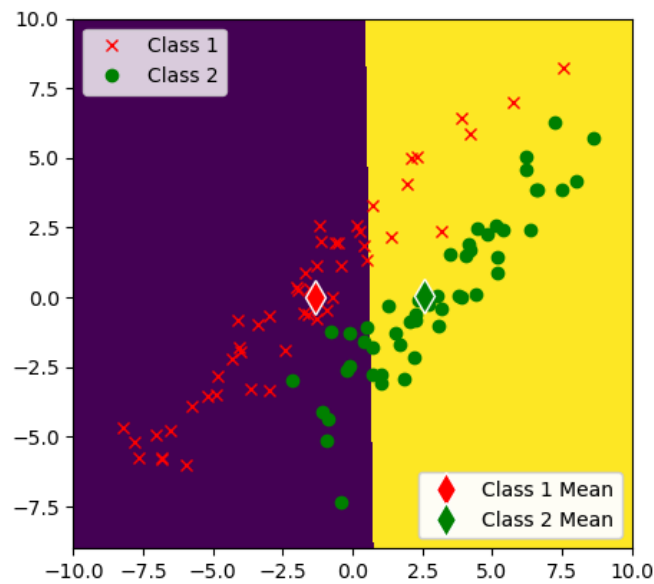Two figures is generated by using *PlotDecBoundaries()*. The CSV file The Fig-

Figure 1: synthetic train1

ure **??** shows the boundary is horizontal

2.2 problem b

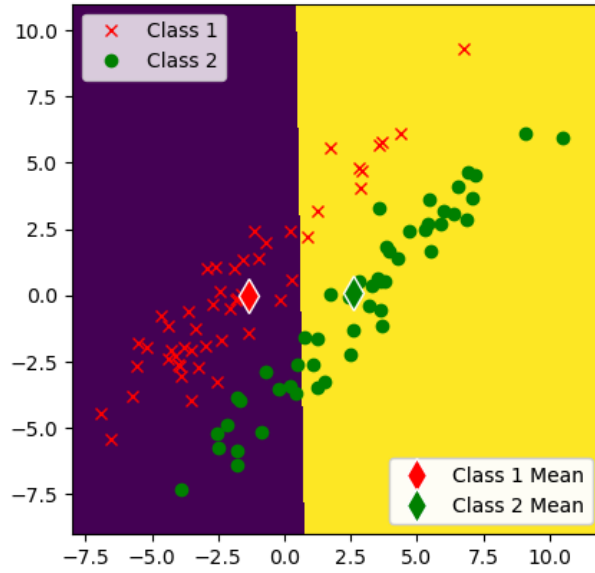The error rate of both synthetic data set is the same.

Figure 2: synthetic test1

Data	Error rate	Test samples
synthetictrain1	0.24	100
synthetictest1	0.24	100
synthetictrain2	0.04	100
synthetictest2	0.04	100

Table 1: Error rate

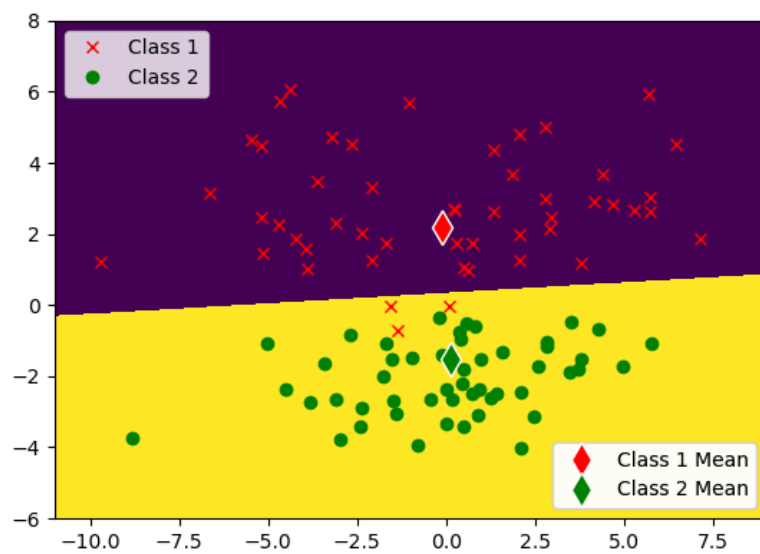## 2.3   problem c

3   Summary
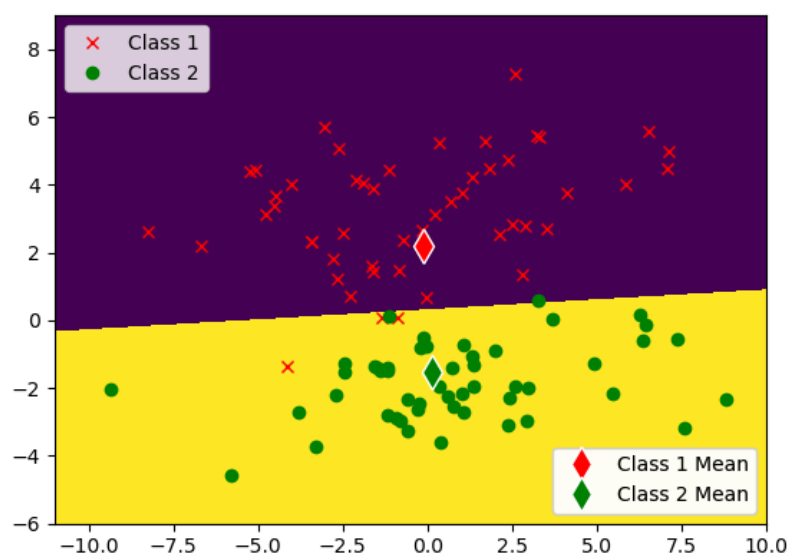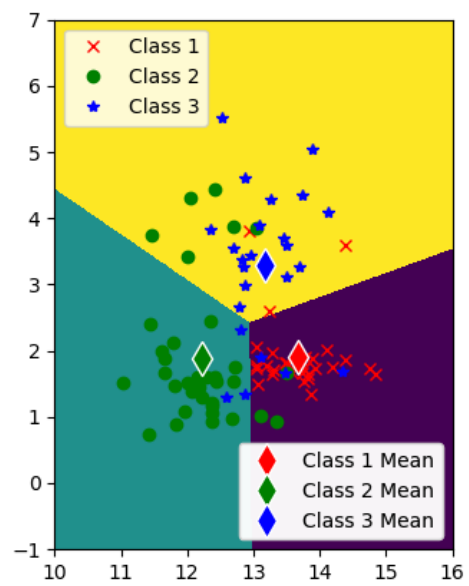
Figure 3: synthetic train2

Figure 4: synthetic test2

4

Figure 5: the data scatter plot of wine data set whose feature is 0 and 1