

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 Figure 1 shows the

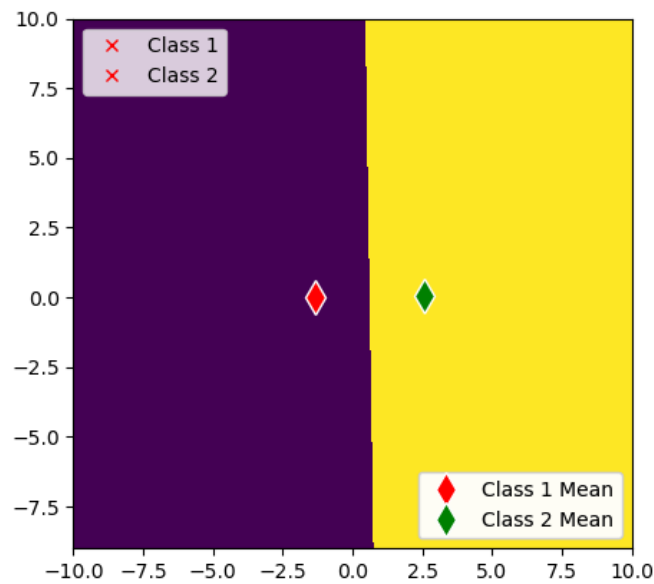


Figure 1: synthetic1

boundary is horizontal sdfsd

2.2 problem b

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

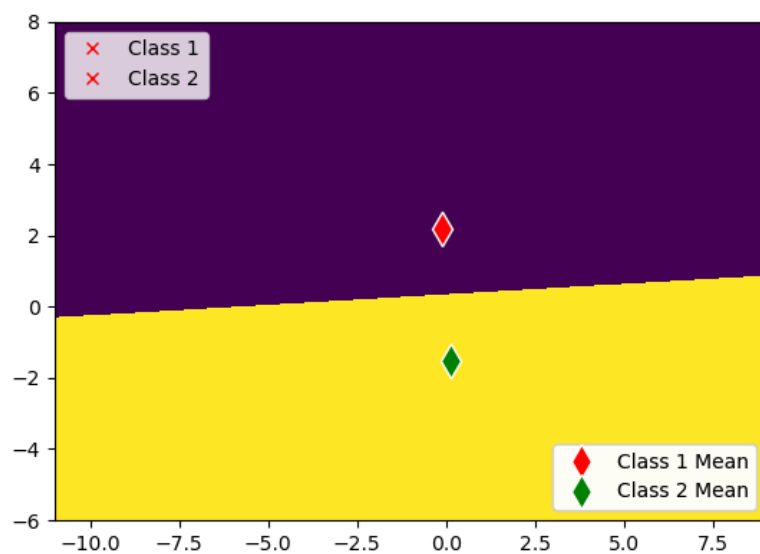


Figure 2: synthetic2

Data	Error rate	Sample number
synthetic1	0.24	100
synthetic2	0.04	100

Table 1: Error rate

3 Summary