CASE 1: Diagonal covariance matrix $\Sigma_1 = \Sigma_2 = I$
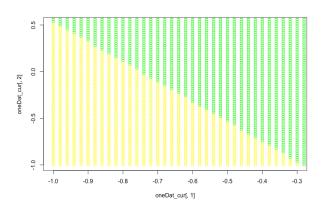

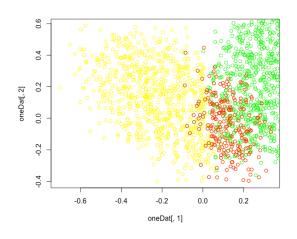Figure 1(a): Plot of training data, for class 1 and class 2 and the mistakes

Figure 1(b): Plot of decision boundary

Here the table shows the MAP probabilities and predicted labels for the first 3 test points

No.	MAP probability: $P(C=0|\mathbf{x})$	MAP probability $P(C=1|\mathbf{x})$	Predicted label
1	0.385604	0.614396	1
2	0.8086006	0.1913994	0
3	0.5325342	0.4674658	0

Table 1: MAP probabilities for case 1

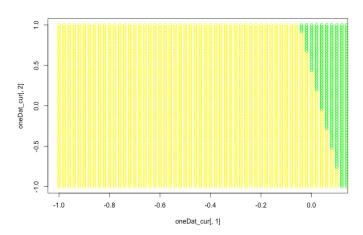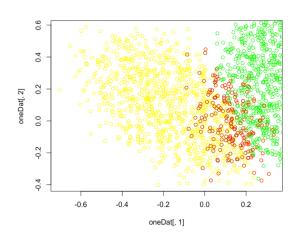**CASE 2:** Common covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma$

Figure 2(a): Plot of training data, for class 1 and class 2 and the mistakes

Figure 2(b): Plot of decision boundary

The table shows the MAP probabilities and predicted labels for the first 5 test points for case 2

No.	MAP probability: P(C=0|**x**)	MAP probability P(C=1|**x**)	Predicted label
1	0.008802201	0.991197799	1
2	0.9881458	0.01185418	0
3	0.05099577	0.9490042	1
4	0.8129269	0.1870731	0
5	0.788086	0.211914	0

Table 2: MAP probabilities for case 2

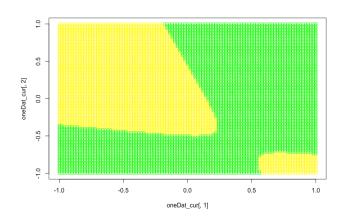**CASE 3:** Full covariance matrix for both classes

Figure 3(a): Plot of training data, for class 1 and class 2 and the mistakes

Figure 3(b): Plot of decision boundary

Here the table shows the MAP probabilities and predicted labels for the first 4 test points

No.	MAP probability: P(C=0|**x**)	MAP probability P(C=1|**x**)	Predicted label
1	1.000000e+00	5.861444e-10	0
2	1	8.687032E-20	0
3	1.426095E-05	0.9999857	1
4	0.570659	0.429341	0

Table 3: MAP probabilities for case 1

Discussion/comparison of results for all cases
The accuracy increases in ascending order from case 1 to case 3 with the least number of mistakes in full covariance matrix and most in identity matrix. The full covariance matrix also learns a better non linear decision boundary as can be seen from the plot Figure 3(b) while the identity matrix equally distributes the two classes.