

Lab Assignment 3

Steven Truong

Task 1

\mathbf{x} is on the decision boundary between classes i and j when the i -th and j -th component of $f(\mathbf{x}, W, \mathbf{w}_0)$ are the same, i.e.,

$$f_i(\mathbf{x}, W, \mathbf{w}_0) = f_j(\mathbf{x}, W, \mathbf{w}_0).$$

Note that $f_i(\mathbf{x}, W, \mathbf{w}_0) = \mathbf{e}_i^\top W \mathbf{x}$, so the equation for the boundary is

$$\mathbf{e}_i^\top W \mathbf{x} = \mathbf{e}_j^\top W \mathbf{x}.$$

where \mathbf{e}_i is one of the canonical unit vectors.

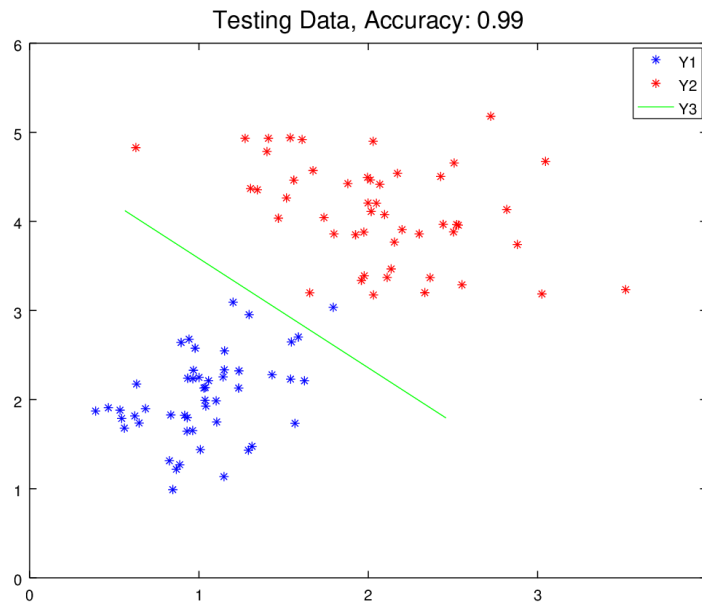
Task 2

Yes, it does make sense to use a linear classifier to classify this data since both data sets occupy different “sides” of the plane.

Task 3



Task 4



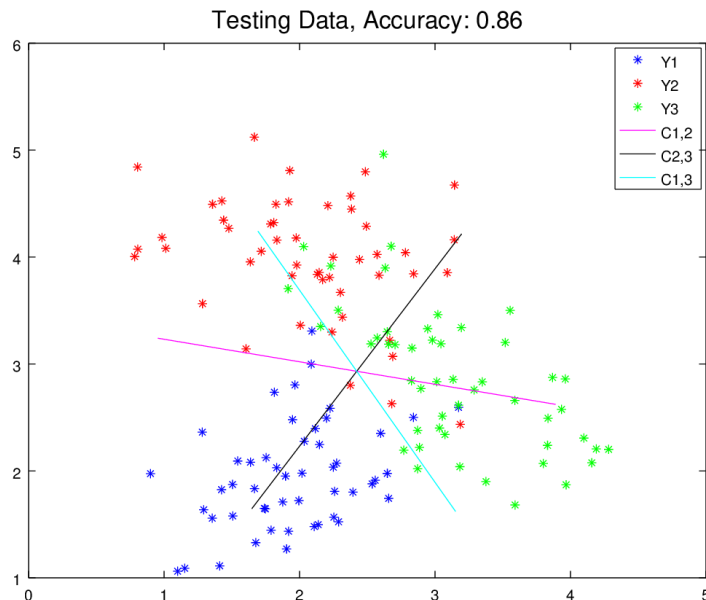
The classification accuracy is $99/100 = 99\%$, since it accurately classified all but one point generated from $\mathcal{N}(\mu_1, \Sigma_1)$. This result makes sense, since the decision boundary separates the points nearly perfectly from our original sample.

Task 5



D1, D2, and D3 refer to class 1, class 2, and class 3, respectively. $C_{i,j}$ is the decision boundary between classes i and j .

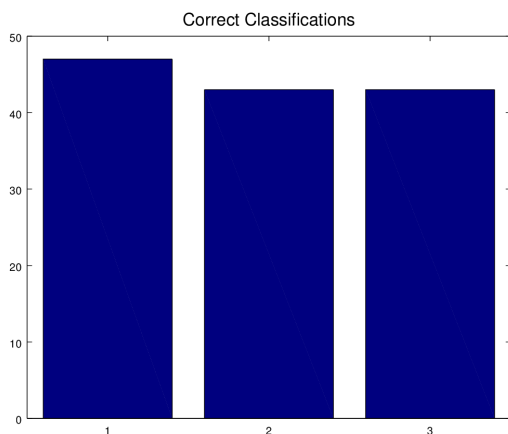
Task 6



Y1, Y2, and Y3 refer to the testing samples of class 1, class 2, and class 3, respectively. $C_{i,j}$ is the decision boundary between classes i and j .

The classification accuracy is about 86%, which makes sense since in the original data, class 2 and class 3 mixed quite a bit, so we expected the linear classifier to not perform as well for this data set compared to the 2 class data.

Task 7



Class	Correct Classifications
1	47
2	43
3	43

k -nearest neighbors worked best when $k = 9$, which gave an accuracy of 88.67%, which is slightly better than the linear classifiers, which had an accuracy of 86%.

This is due to the fact that k NN is able to classify less extreme outliers, i.e., outliers which are still not too far from the rest of its class. On the other hand, the linear classifier is unable to do so, and both classifiers are unable to classify more extreme outliers, which is why k NN only performs slightly better.