Lab5

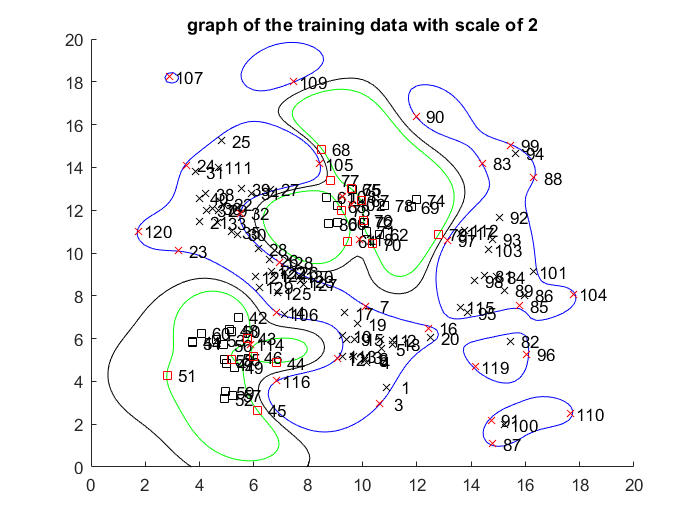
Kevin Wang

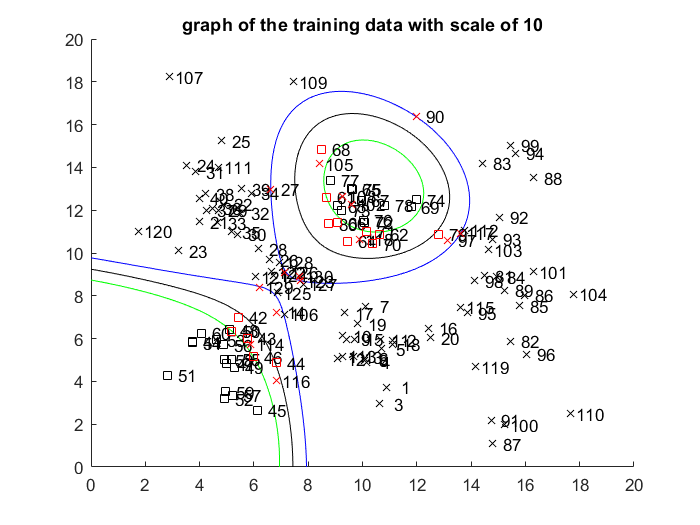
Chelsey Yin

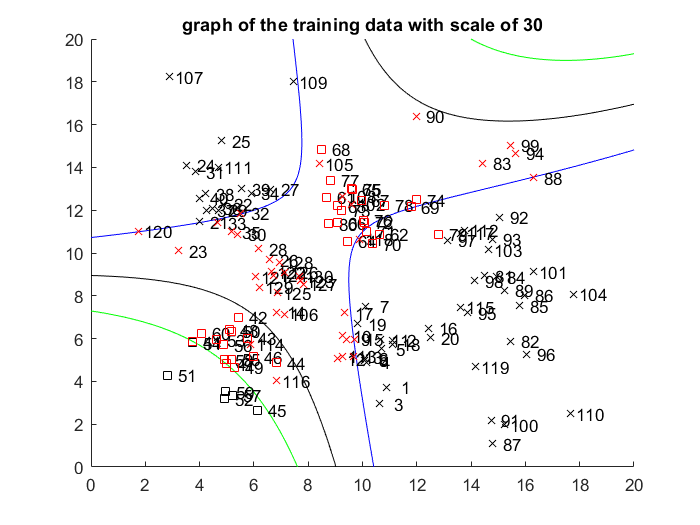
In this lab, the kernel function we used is “rbf”, the box constraint we used is 100. For the kernel scale, we used 3 different numbers to see their difference. They are 2, 10 and 30. We feel that 10 is the reasonable number that we should use. 2 is too small, and it strictly apply to the training set. 30 is too board, and it has more errors.

Below is the table of data we get from the testing set. 2, 10 and 30 are different kernel scales.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2 | 10 | 30 |
| true positive | 40 | 39 | 20 |
| true negative | 80 | 87 | 86 |
| false positive | 10 | 3 | 20 |
| false negative | 0 | 1 | 4 |
| TPR | 1 | 0.975 | 0.5 |
| FPR | 0.111 | 0.033 | 0.044 |

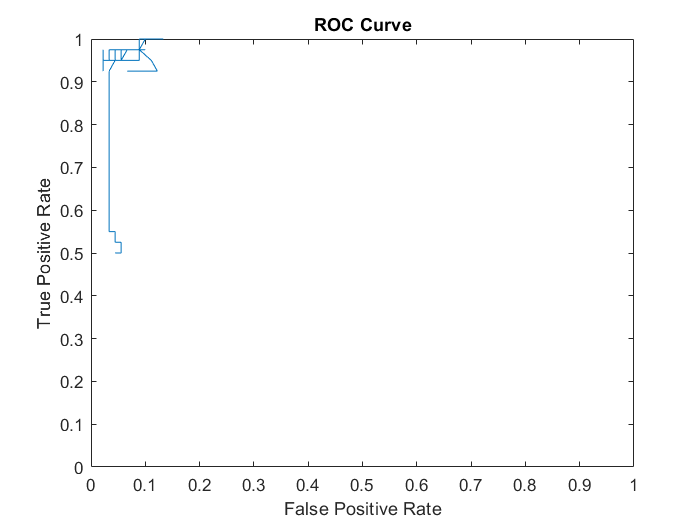






We believe that our data is quite reasonable. We have a high true positive rate and a low false positive rate. From the graph, they show their properties under different scales. Our classifier does not achieve 100% accuracy. But it can get most of them correct when we choose 10 as the kernel scale. In the training data, some of the positive and negative data are quite close. This makes it hard to separate them by clear vectors. So it is almost not possible to achieve 100% accuracy.

Below is our ROC Curve



Our roc curve has similar shape of example roc curve, but some part of our roc curve is not look like the example. Our small test set might be the reason. Another reason maybe we didn’t pick the correct range of data so some of them goes beyond the usual range. We only change the scale (sigma of rbf kernel) here. It will be better if we also test the box constraint (c of rbf kernel).