Document your kernel and parameters

We used an RBF (Gaussian Radial-basis function) as our kernel function to separate the data when training on the training set. We used 8 as our common length scale parameter and the default value (1) for the coefficient’s upper bound (C) to the RBF function through trial and error while analyzing the results with different values. We found that a common length scale of 8 and C=1 provided the highest accuracy and TPR, while 20 and C=1 provided the highest precision and minimized FPR. This data is shown in table 2 and table 3.

Results

The description of the results is using the first of each table.

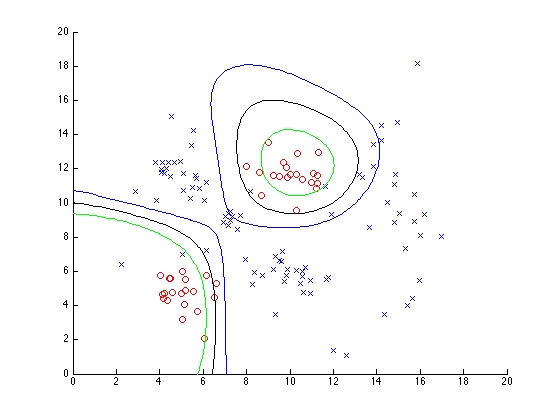
The training set with the RBF kernel function produced the decision boundary shown in figure 1. Any data points in a region between black and green lines or surrounded by green lines is classified > 0 and anything between black and blue lines or surrounded by blue lines is classified as < 0. We then used this decision boundary to classify each point in the test data set. The final collective results of these classifications are shown in table 1. There were 40 instances of true positives, 0 false negatives, 4 false positives, and 86 true negatives.

As shown is table 3, we calculated the true positive rate (TPR) to be 100%, false positive rate (FPR) to be 4.44%, accuracy to be 96.92%, and precision to be 90.91%. TPR was calculated by dividing the number of true positives (40) by the total that was actually positive (40+0=44). FPR was calculated by dividing the number of false positives (4) by the total number that was actually negative (4+86=90). Accuracy was calculated by dividing the total number correct (40+86=126) by the total number of data points (40+86+0+4=130). Precision was calculated by dividing the number of true positives (40) by the total detected as positive (40+4=44).

The classifier will never achieve 100% accuracy, no matter the data. By looking at the training data, we can see that data points from one classification are intermixed with another classification, making it impossible to separate and classify. To achieve 100% accuracy, removing the points mixed in a cluster of opposing classification needs to be done.

We believe the results are reasonable for the given data due to the high level of accuracy and precision, as well as the low false positive rate (FPR).

Contour plot (Figure 1)



Final Results (table 1)

|  |  |  |
| --- | --- | --- |
| **Detected Class**  **True Class** | 1 | -1 |
| 1 | 40 | 0 |
| -1 | 4 | 86 |

Results with different parameters (table 2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Upper Bound (C)** | **Common Length Scale** | **True Positive** | **False Negative** | **False Positive** | **True Negative** |
| Default (1) | 8 | 40 | 0 | 4 | 86 |
| 100 | 8 | 39 | 1 | 5 | 85 |
| Default (1) | 20 | 38 | 2 | 3 | 87 |
| 100 | 20 | 39 | 1 | 4 | 86 |

Result calculations with different parameters (table 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Upper Bound (C)** | **Common Length Scale** | **Accuracy** | **TPR** | **Precision** | **FPR** |
| Default (1) | 8 | .9692 | 1 | .9091 | .0444 |
| 100 | 8 | .9538 | .9750 | .8864 | .0556 |
| Default (1) | 20 | .9615 | .95 | .9268 | .0333 |
| 100 | 20 | .9615 | .9750 | .9070 | .0444 |