KU LEUVEN

Support Vector Machine Assignment

Mohamedhakim Elakhrass

15/07/2016

Contents

1   Objectives

The objectives of this assignment are to learn and work with real life applications of SVM.

2   Simple 2 Gaussian

This sections looks at two simulated datasets. The datasets are created using the MatLab $randn()$ function. One dataset is centered around (1,1) and the other (-1,1)
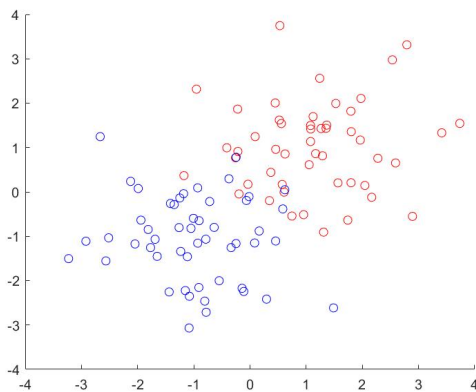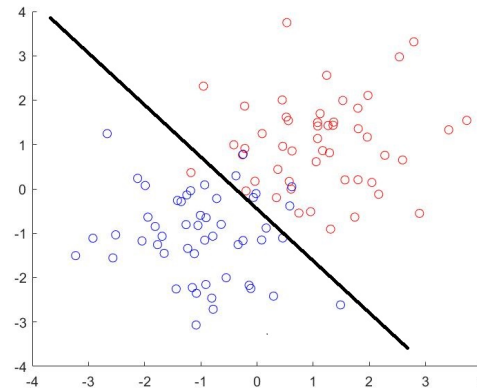

Figure 1: Two Simulated Datasets.

Figure 2: Two Simulated Datasets with Optimal Classifier.

Given this figure, can you make a geometric construction using lines to estimate the optimal classifier? Under which conditions do you think this construction is optimal/valid?

Figure 1:a is the output of the simulated datasets. As shown in figure 1:b it is possible to show an optimal classifier. This classifier is known as the Bayes Classifier. A test observation is assigned with predictor vector x_0 to the class j for which

$$Pr(Y = j | X = x_0)$$

is largest. The classifier is optimal because it produces the lowest possible error rate and allows for some overlap. The classifier is valid because the underlying distribution of the dataset is known. This falls into the special case $\Sigma_{xx1} = \Sigma_{xx2} - \Sigma_{xx}$, the covariance matrices are equal and the decision boundary is linear.

3   The Support Vector Machine

This section will deal with an online demo of a linear and none linear SVM. It will be used to learn the intuition behind changing SVM parameters and changes in the dataset.

Adjust the existing datasets to have at least 10 data points for each class. What do you observe when you are adding data points to the classes? How drastically can classification boundaries change.
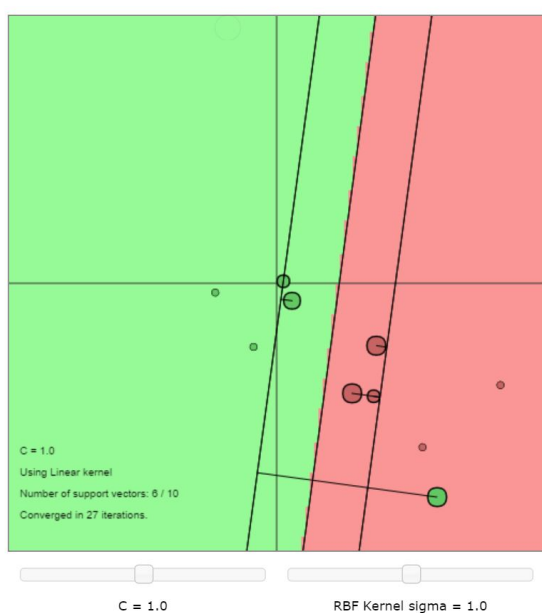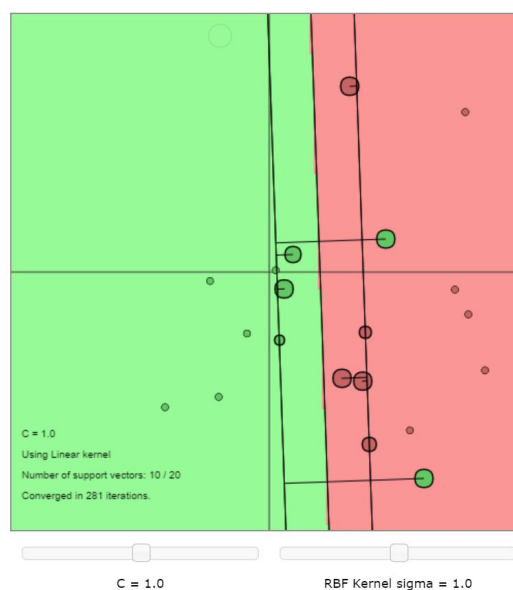
Figure 3: Default Linear Kernal.

Figure 4: 10 Data Point Linear Kernal.

Data points added inside the margin drastically change the decision boundary and become support vectors. Data points added to the side of the opposing color are also automatically support vectors but remain misclassified. Data points added to the same side as its own color have very little effect on the decision boundary, although the closer to the boundary the larger the effect.

What if you add an outlying datapoint which lies on the wrong side of the classification boundary? How does it affect the classification hyperplane?
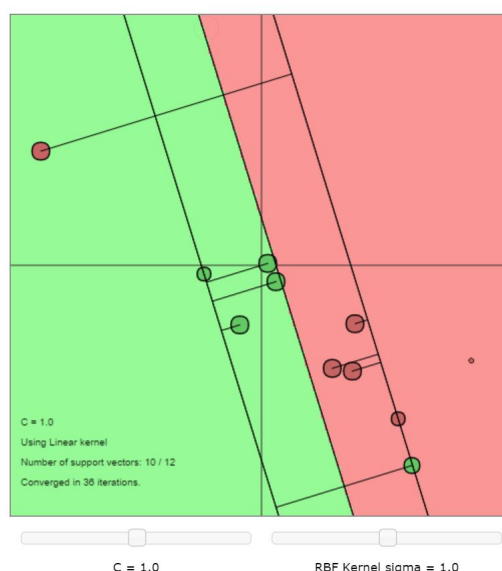

Figure 5: Outlier data point

A data point added to the wrong side of the boundary changes the direction of the hyperplane towards that point. If too many points are added the classifier misclassified all points of the opposing class.

Try different values of C regularization hyperparameter. How does it affect the classification outcome? What is the role of it?

Using the intial dataset the effects of the regularization hyperparemter can be seen. The parameter control the slack of the SVM model. When C is high there is less tolerance for misclassification and therefore a smaller margin. When C is large there is higher tolerance for misclassification and a larger margin.
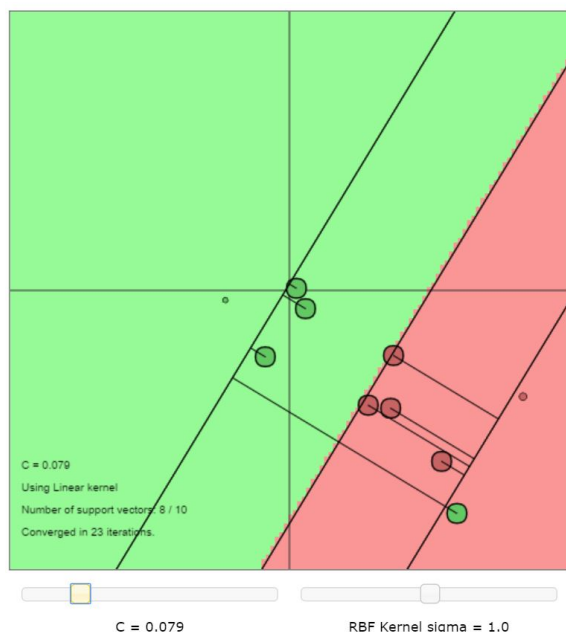

Figure 6: C = .079

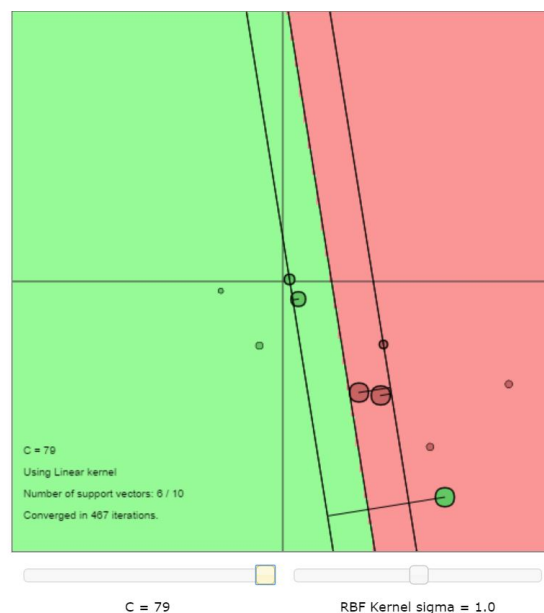
Figure 7: C = 79

Follow the instructions and switch back to RBF kernel by toggling the k button. Compare to the classification outcome of the linear case.
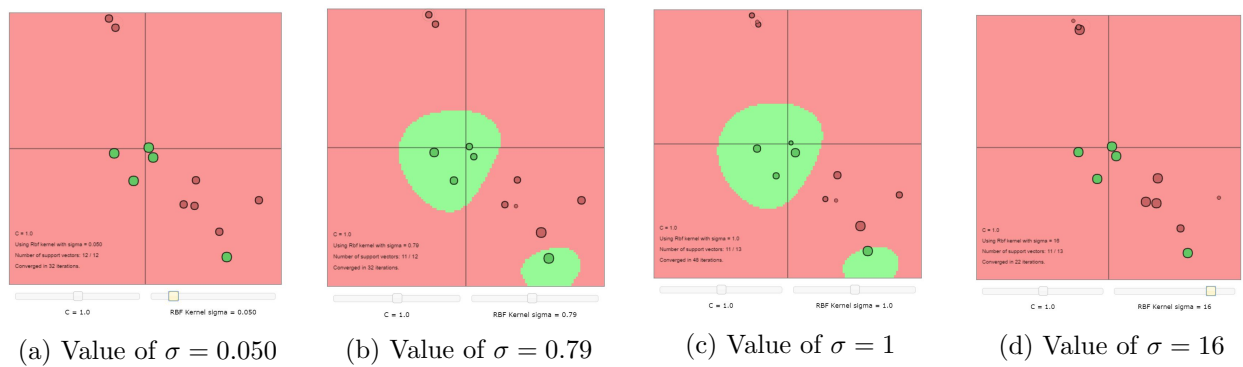
In this section, we were asked to switch away from a linear kernel and opt for the Radial Basis Function kernel (RBF). RBF are localized function, centered at a particular value, and use a gaussian fit with a certain spread around the center.

The same dataset as defined above can be seen on the graphics on the right. Here, one can observe that all points have been correctly classified, but the area defined are very wobbly, displaying clear signs of overfitting.

This can be tuned by changing the value of σ, which controls the bandwidth of the gaussian function, see figure 9. Larger values of σ leading to more linear decision boundaries.
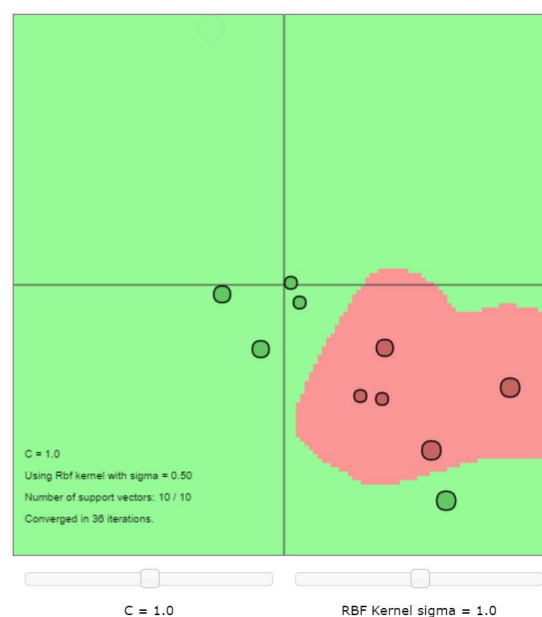
The RBF kernal has no misclassification as opposed to the linear classifier. This is because the data set is not linearly seperable. The RBF kernal as the ability to wrap around the data.The RBF is a localized funtcion

(a) Value of $\sigma = 0.050$ (b) Value of $\sigma = 0.79$ (c) Value of $\sigma = 1$ (d) Value of $\sigma = 16$

Figure 9: How σ effects RBF classifier with C = 1.0