FACULTY OF BIOSCIENCE ENGINEERING
Masters of Bioinformatics
Support Vector Machines

KU LEUVEN

Support Vector Machine Assignment

Mohamedhakim Elakhrass

15/07/2016

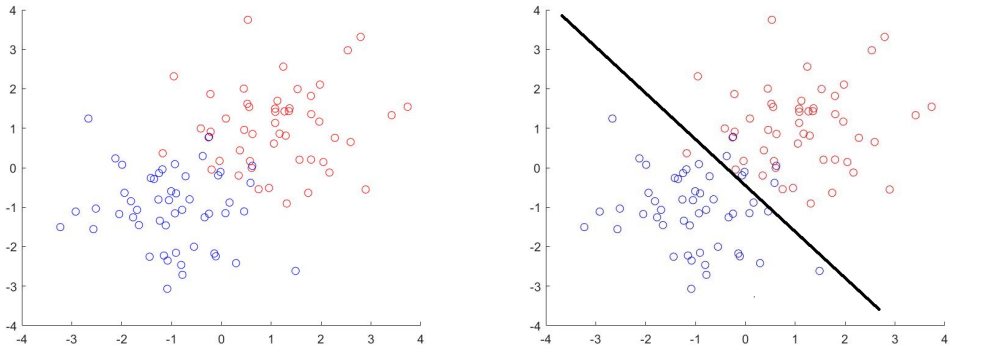
Contents

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)



(a) Two Gaussians (b) Two Gaussians with optimal classifier

Figure 1: (a) Two Simulated Datasets. (b) 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?

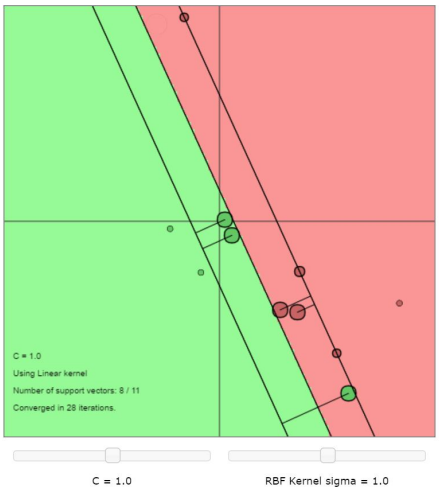
Figure 1:a is the output of the simulated datasets. As shown in figure 1:b it is possible to show an optimal classifier. This classifier is known as the Bayes Classifier. A test observation is assigned with predictor vector x_0 to the class j for which

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

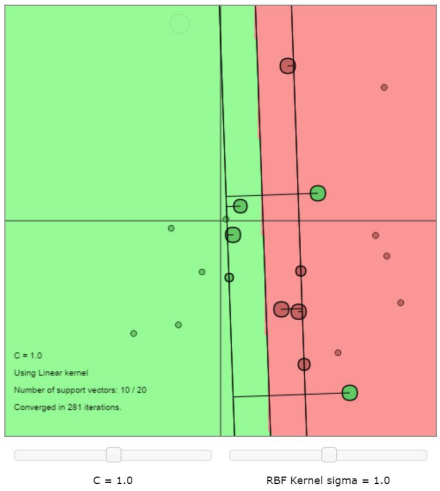
is largest. The classifier is optimal because it produces the lowest possible error rate and allows for some overlap. The classifier is valid because the underlying distribution of the dataset is known.

3 The Support Vector Machine

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



(a) Default Linear Kernal



(b) 10 Data Point Linear Kernal