Instituto Superior Técnico

Departamento de Engenharia Electrotécnica e de Computadores

Machine Learning

5th Lab Assignment

Shift Tuesday 17:00 Group number 19

Number 78308 Name Rui Daniel Ribeiro Dias

Number 75494 Name Nuno Wallenstein Teixeira

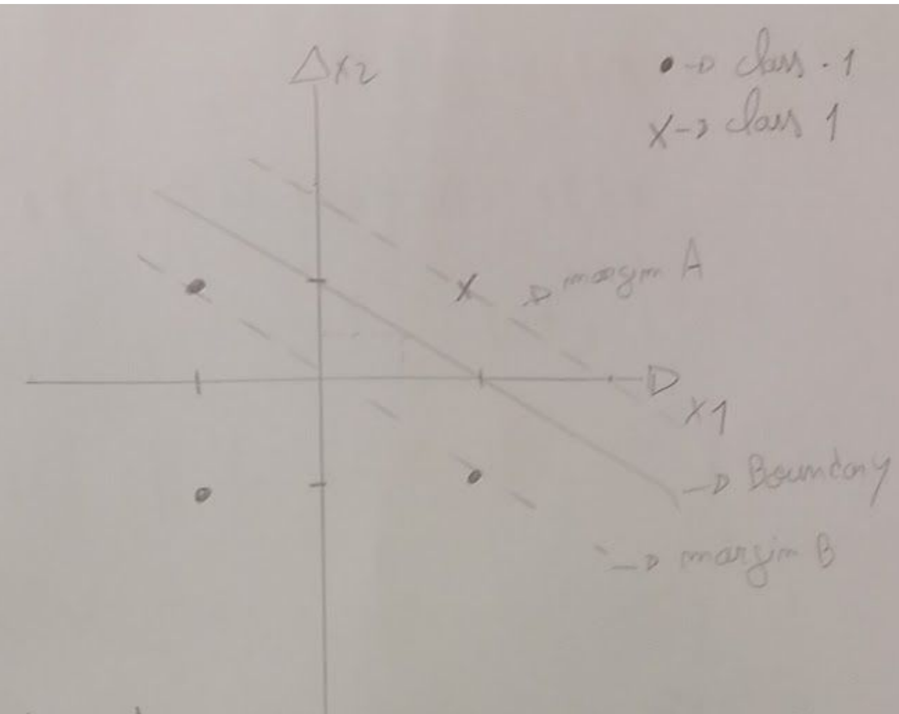
Support Vector Machines for Classification

2.Two simple examples

2.1(T) For the AND function, find (by inspection) the maximum-margin separating straight line, the support vectors and the margin boundaries. Then compute the vector w and the bias b that satisfy the equation

$$(w \cdot xs + b)\, ds = 1 \quad (1)$$

for all support vectors xs, where ds is the desired value corresponding to xs.



Support vectors: [-1, 1]; [1, -1]; [-1, -1]

$$\vec{W} = [a, a] \qquad \begin{cases} wx + b = -1 \\ wx + b = 1 \end{cases} \rightarrow \begin{cases} [a, a]\begin{bmatrix} 1 \\ -1 \end{bmatrix} + b = -1 \\ [a, a]\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = 1 \end{cases}$$

$$\begin{cases} a = 1 \\ b = -1 \end{cases} \Rightarrow \quad \bullet\ f(\ ) = x_1 + x_2 - 1 \Rightarrow f(x) = w^T x - 1$$

$$\bullet\ g(x) = w^T x - 1 = 0$$

Our vector w=[w1,w2] has the respective values w=[1,1]

2.2)

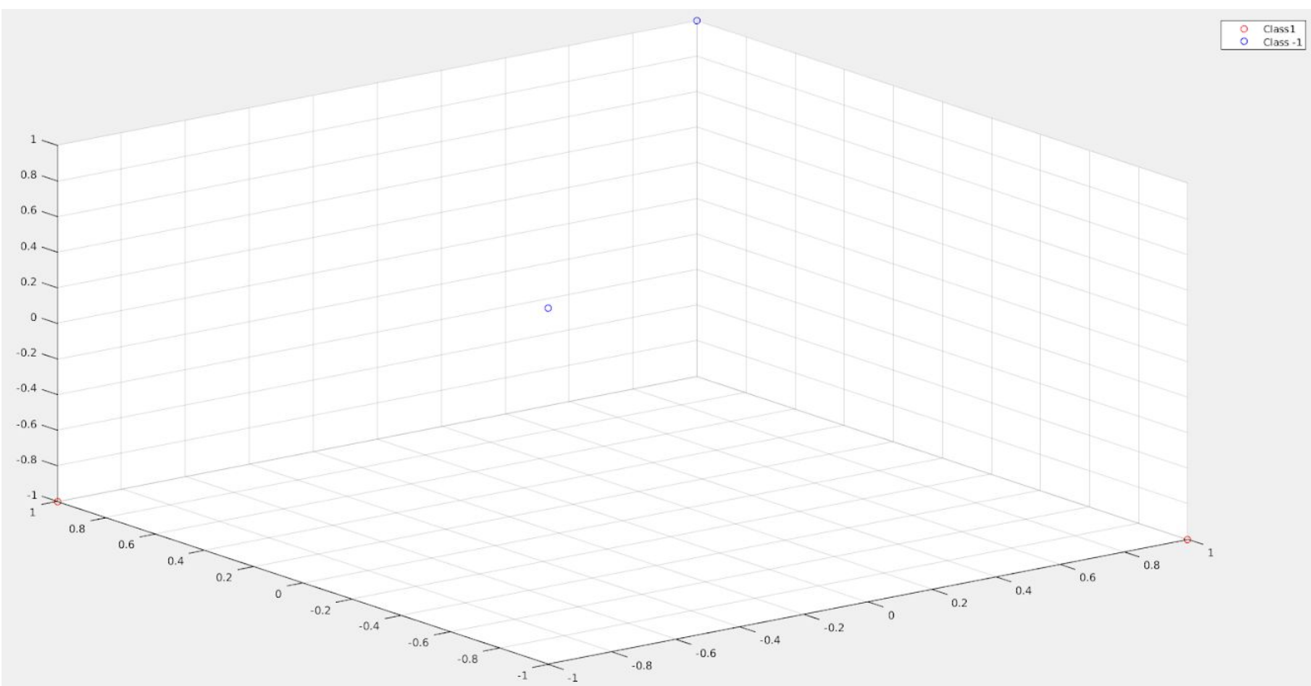
A linear classification can not be performed with the XOR function because the data does not allow that, we have mixed classes. If we go to higher dimensions, and so to a different dimensional space, we will be able to find an hyperplane in that feature space that separates our data.

The Kernel Function that corresponds to the mapping of (4) is the following one:

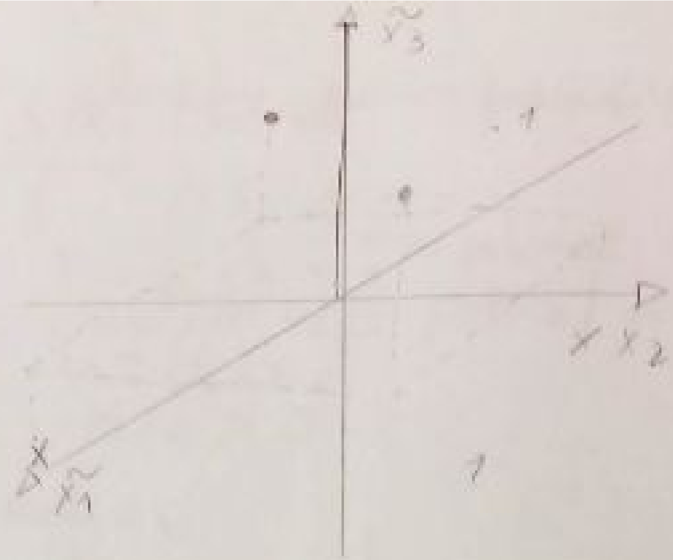
- x= ϕ(x) = (x1, x2, x1x2)T (4)

- K(X,Y)= x1y1 + x2y2 + x1x2y1y2

2.3)



2.3



Support vectors: $[-1, -1, 1]$; $[1, 1, 1]$;
$[-1, 1, -1]$; $[1, -1, -1]$

Hyperplane: $\tilde{x}_3 = 0$

$\tilde{W} = [0, 0, a]$

$$\begin{cases} \tilde{w} \cdot \tilde{x} + b = 1 \\ \tilde{w} \cdot \tilde{x} + b = -1 \end{cases} \rightarrow \begin{cases} a = -1 \\ b = 0 \end{cases}$$

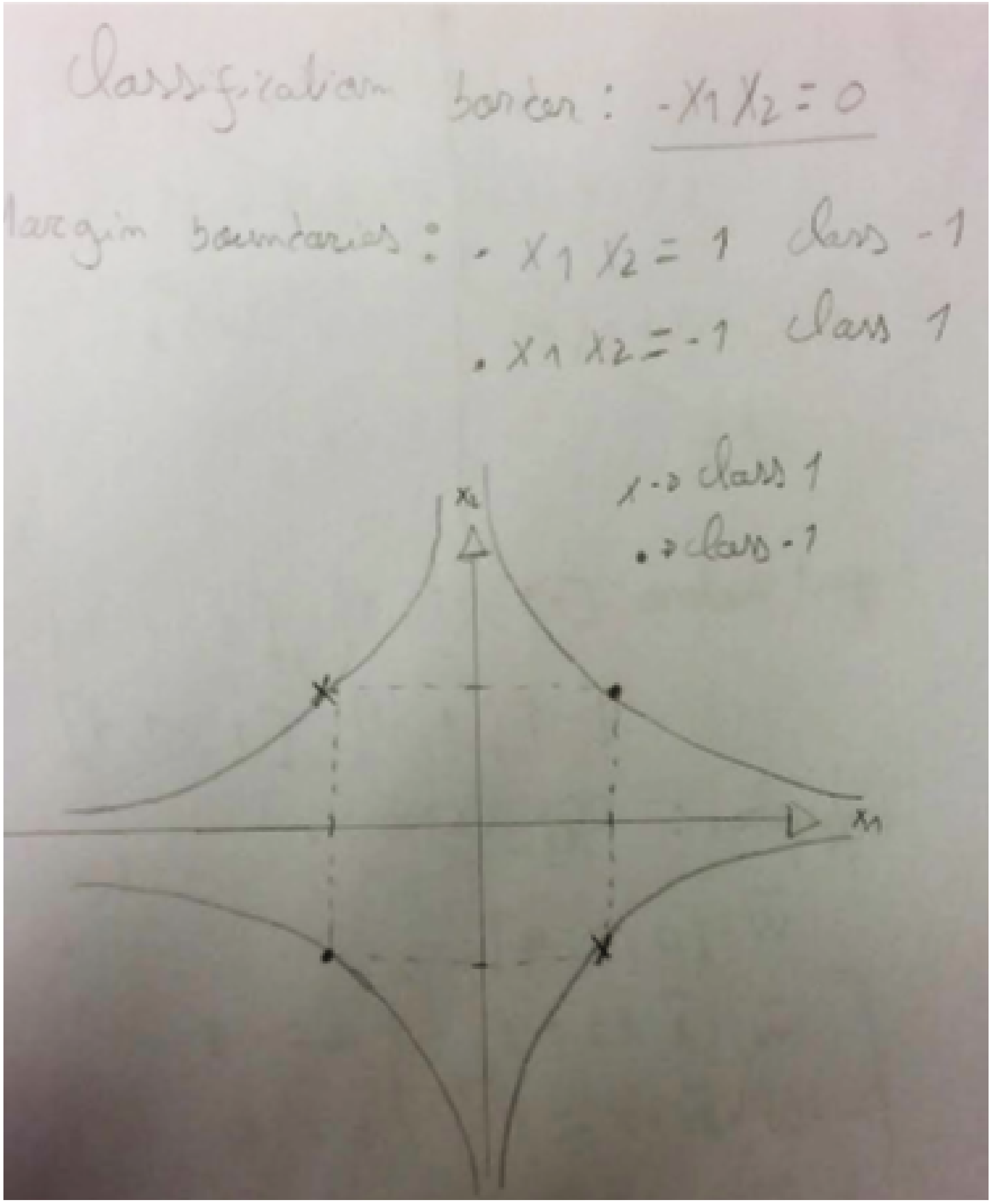
$$g(x) = \tilde{W}^T \tilde{x} + b = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} [\tilde{x}] = 0$$

$$g(x) = -\tilde{x}_3 = -x_1 x_2 = 0$$

$$f(x) = -x_1 x_2$$

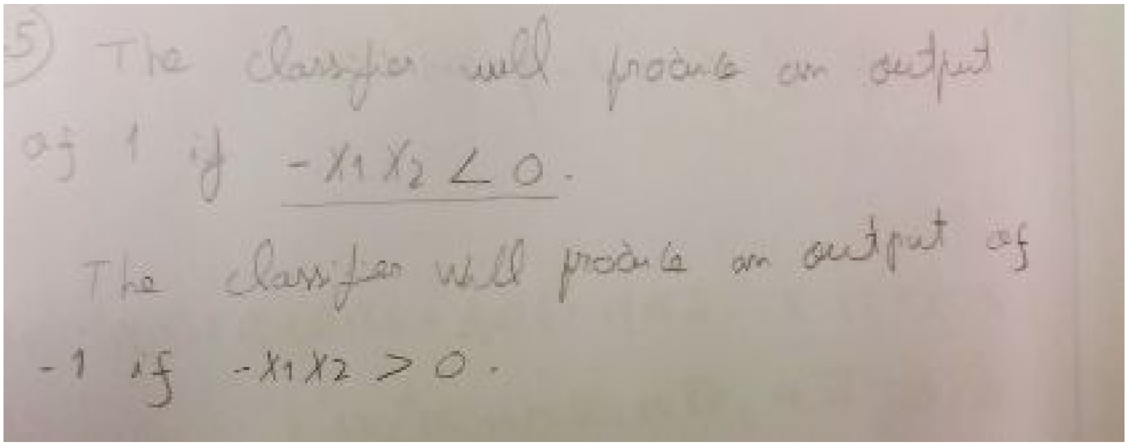
Our vector w=[w1,w2,w3] has the respective values w=[0,0,-1]

2.4)



Our classification condition is based on the sign of (-x1x2). The sign of the previous operation will tell us if a given point is classified as being in class 1 or -1.

2.5)

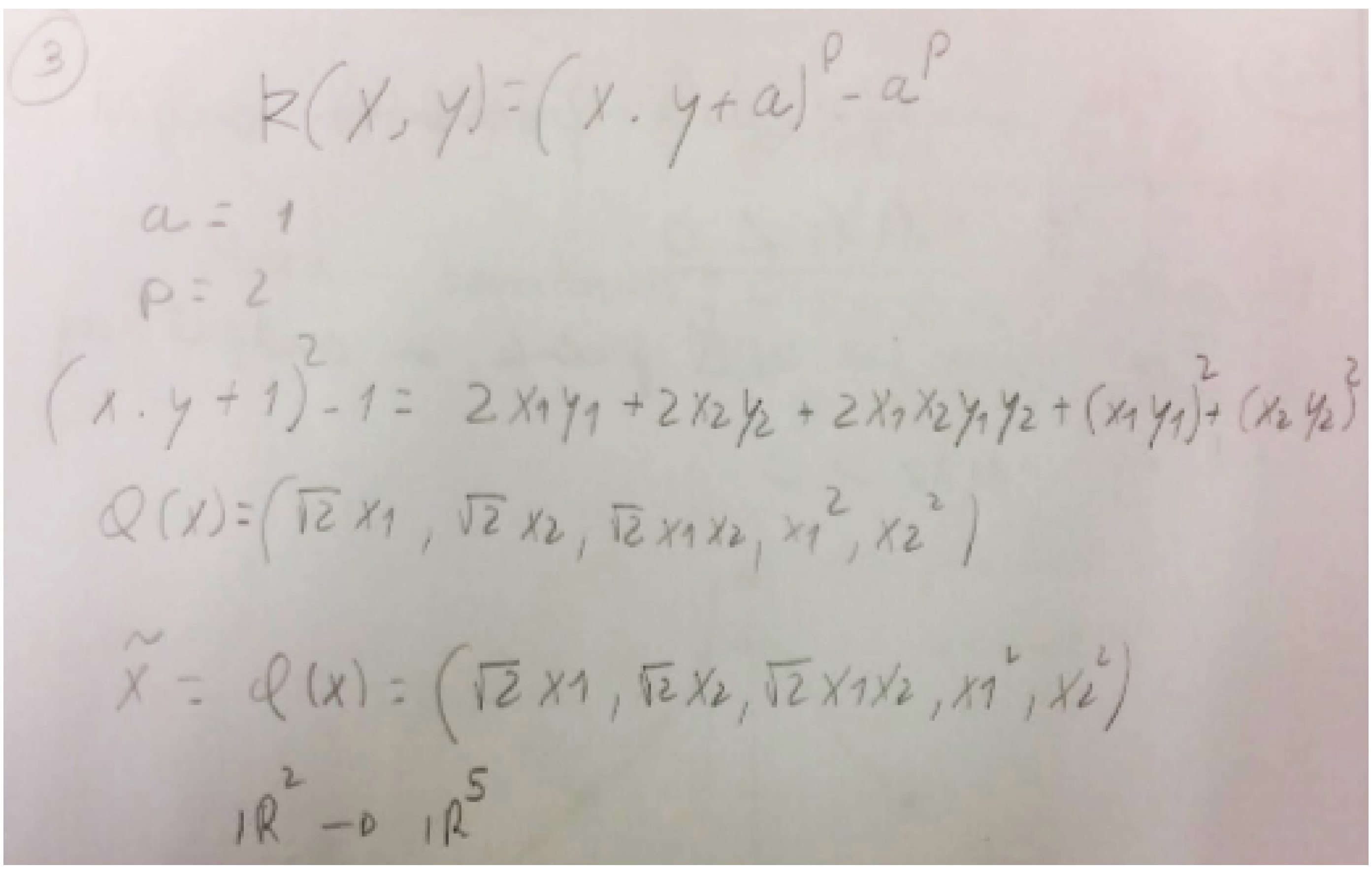

5) The classifier will produce an output of 1 if $-x_1 x_2 < 0$.

The classifier will produce an output of -1 if $-x_1 x_2 > 0$.

Our boundary as the following expression : $g(X)=-x1x2=0$ and $f(x)=-x1x2$. The sign of $f(x)$ will give the class of the input point. If $f(x)<0$ it will be classified as class -1 , if $f(x) >0$ it will be classified as class 1.

3. Classification using SVMs

3.1)


3)

$$k(x,y)=(x\cdot y+a)^p-a^p$$

$a=1$

$p=2$

$$(x\cdot y+1)^2-1 = 2x_1y_1+2x_2y_2+2x_1x_2y_1y_2+(x_1y_1)^2+(x_2y_2)^2$$

$$\varphi(x)=(\sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$

$$\tilde{x}=\varphi(x)=(\sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$

$$\mathbb{R}^2 \to \mathbb{R}^5$$

The new feature space is a 5D space, and so it is composed of 5 dimensions.

3.2)

Using the following expression:

$w\varphi(x)=1$ and assuming that we have the p=1 and a=1. For w to represent this new feature space but with the same classification borders and margins:

$$w\,(\sqrt{2}x1,\ \sqrt{2}x2\ ,\sqrt{2}x1x2, x1^2, x2^2) = (w1, w2, w3, w4, w5)^T(\sqrt{2}x1,\ \sqrt{2}x2\ ,\sqrt{2}x1x2, x1^2, x2^2)$$

From out input space we know that -x1x2=1 is a condition for one of the margins. Substituting this in the previous equation we get:

$(w1, w2, w3, w4, w5)^T(\sqrt{2}x1,\ \sqrt{2}x2\ ,\sqrt{2}x1x2, x1^2, x2^2)$ =(0,0,-x1x2,0,0)

For that to happen $(w1 = 0, w2 = 0, w3 =- 1/\sqrt{2}, w4 = 0, w5 = 0)$ and so our vector w in this feature space is w=[0,0,$-1/\sqrt{2}$,0,0]

4. Experiments

4.1)

P (order)	Error	Support Vectors
6	0	28
8	0	45
10	0	55
11	0	62
12	5	58
20	29	38

Table1. Error and Number of Support Vector Obtained as function of the polynomial order.

It is clear that the order of polynomial has a huge impact in the error obtained. The purpose is always to get the best classification and increasing the order of polynomial increases the flexibility of the boundary to adapt to our data. This means that increasing the order will leads to an overfitting of the boundary and so the margin will be smaller, leading to an increase in the error and so miss classification.

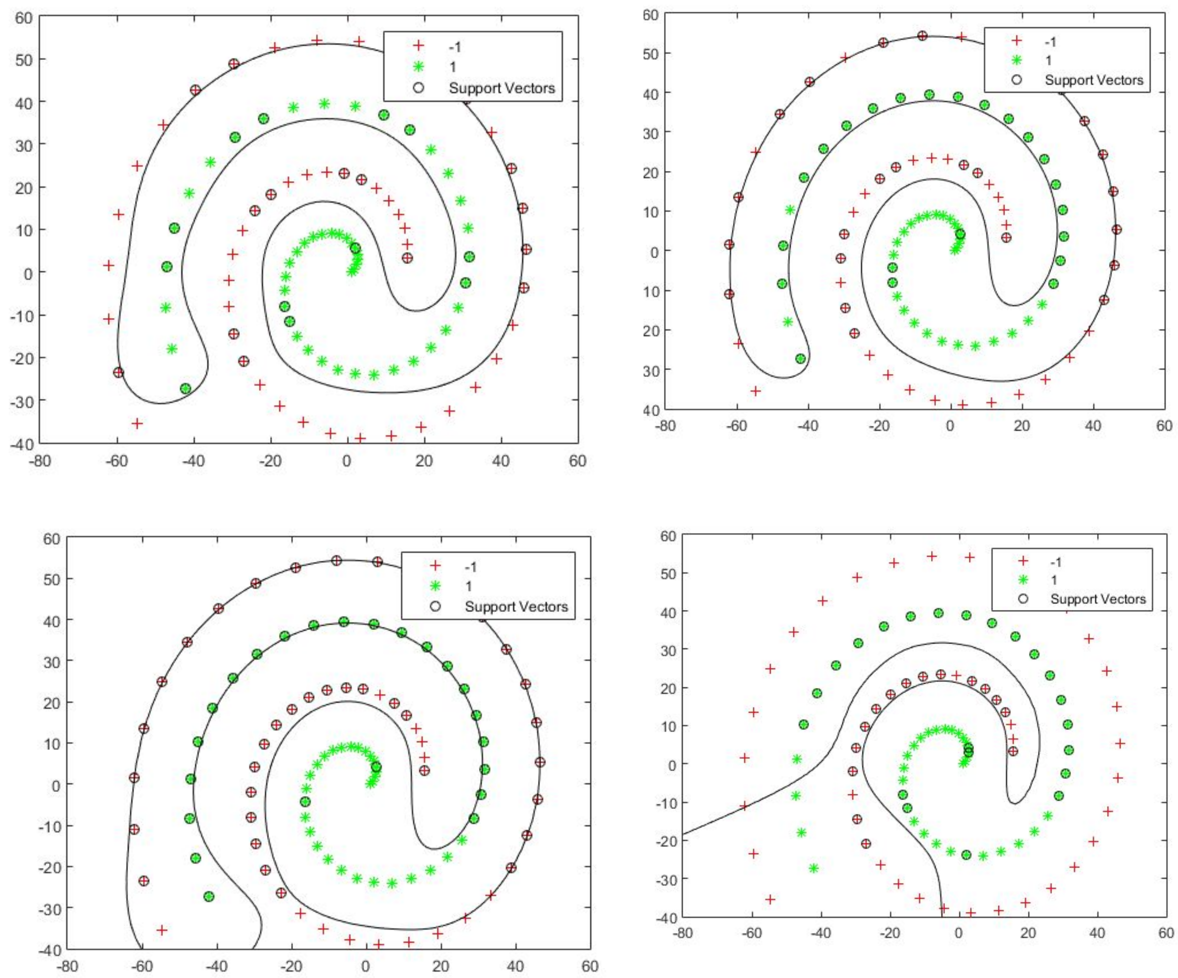


Fig1: Plot for “polyorder” 6,8 (Top) 12 and 20 (Bottom).

4.2)

σ	Error	Support Vectors
0.1	0	97
0.5	0	70
1	0	48
2.5	20	55
5	32	64
7.5	31	70

Table 2. Error and number of Support Vector as function of σ.

Using this method we have one value (σ) that we have to choose to get a good result. If the value is to low (close to 0) we will end up with an overfitting, which is not desired.

4.3)

σ	Error	Support Vectors
0.1	0	64
0.5	0	23
1	0	10
2.5	0	11
3	0	12
3.5	0	12

Table 3:Error and number of Support Vector as function of σ.

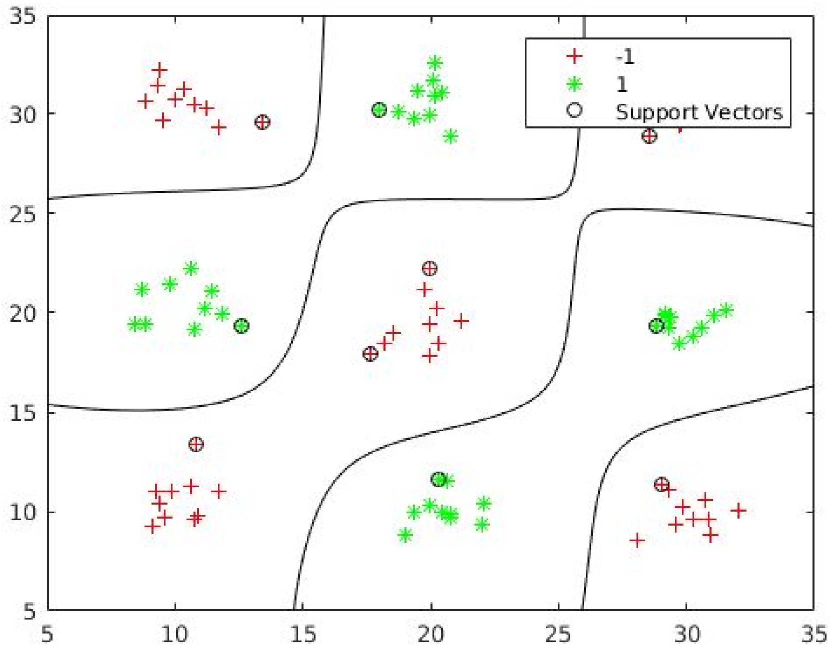
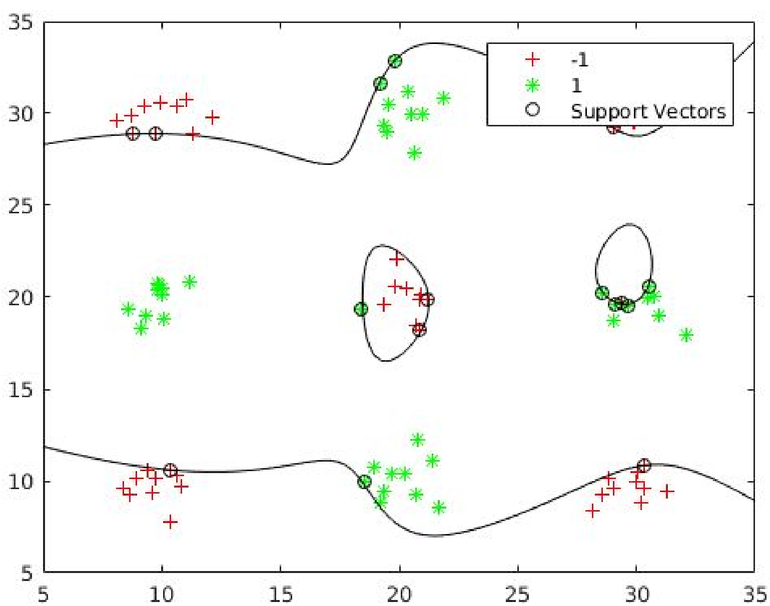


Fig 2. Plot of the boundaries with the separation of the data with σ=1.

We can clearly see that this method with the correct parameters leads to a good separation of the data.

4.4)erro=0 Support vectors 16



We can clearly see that this data is not correctly separated by the SVM algorithm. Since we have an outlier pattern we need a soft-margin algorithm which is the same as decreasing the boxconstraint value. Setting the value of the boxconstraint to Inf doesn't lead to the best results because it is not correctly separating the data.

4.5)

BoxConstraint	Error	Support Vectors
10	2.222	24
1000	1.1111	13
10^5	1.1111	16
10^7	0	16
10^9	0	16
10^{11}	0	16

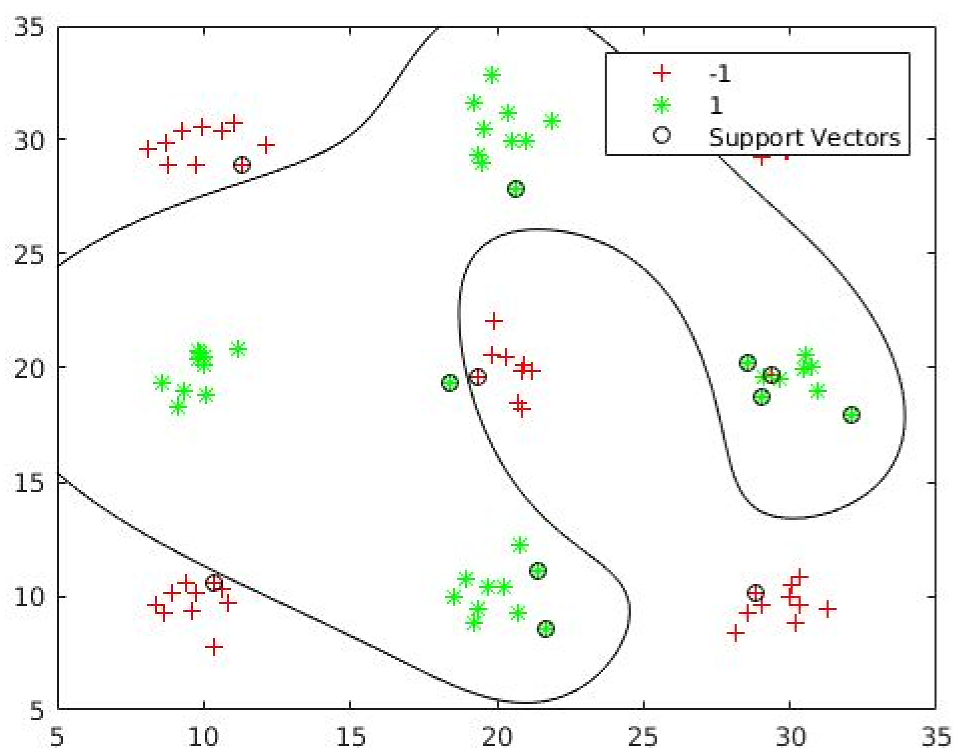


Figure - Boxconstraint 1000

With a boxconstraint of 1000, the soft margin algorithm is able to seperate the data even when the classes are very close to each other. Although, when the box decreases too much the error increases. There exists an optimal value for the boxconstraint.