

**Information Mining - winter semester 2020****Exercise sheet 10****Exercise 1: Classification with SVM**

Assume you have given the following data:

x1	x2	type
0.5	3.5	1
1.0	1.0	1
1.0	2.5	1
2.0	2.0	1
3.0	1.0	1
3.5	1.2	1
4.0	5.8	-1
3.5	3.0	-1
4.0	4.0	-1
5.0	5.0	-1
5.5	4.0	-1
6.0	3.0	-1

Table 1: Some classification data.

When this data is plotted it looks like as shown in Figure 1. In the figure the red dots are the positive classes and the black dots the negative ones.

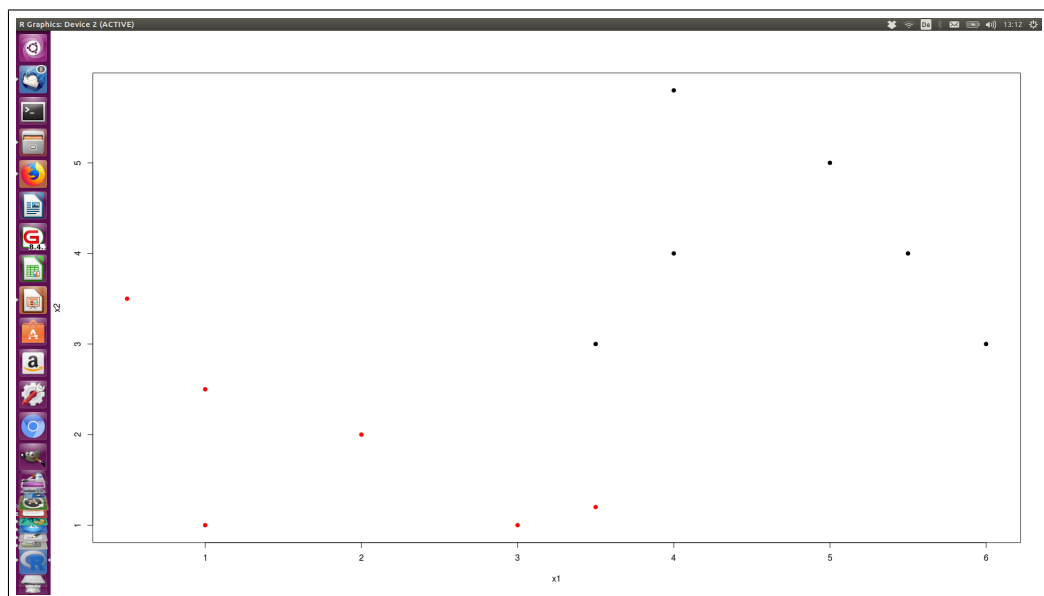


Figure 1: Some classification data plot.

- For this data perform a SVM linear (soft margin) classification using R. SVM in R can be installed using `install.packages("e1071")`. Once this is installed you can use it by loading `library(e1071)`.
- Once you have trained the linear SVM model you should plot it along with the data. Ideally the plot should show the hyperplane (separating line) and the support vectors.
- Make prediction for the following points: P1(1,4) and P2(3.5,3.5).
- Now add the following points to the above data shown in Table 1: P3(6,2,1) and P4(3,5.5,1) where the third number indicates the class of the point. Re-do the SVM linear training with the extended data. What do you see?
- Now perform a kernel trick and select a polynomial kernel with degree 2. What changed on the support vectors? Is the data again separable?
- On this data play with the parameter C (cost). Set it to 0.001 and then to 100000. What do you observe?
- Finally add to the extended data the point P5(2.5, 2.5, -1) (keep C equals 100000) and retrain your polynomial SVM. What do you observe?