Department of Mathematics and Computer Science
University of Southern Denmark, Odense

February 22, 2013
Marco Chiarandini

DM825 - Introduction to Machine Learning

Sheet 9, Spring 2013

Exercise 1

Suppose that the following are a set of points in two classes:

$$\text{class 1: } \begin{bmatrix}0\\0\end{bmatrix} \begin{bmatrix}1\\2\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix}$$

$$\text{class 2: } \begin{bmatrix}1\\1\end{bmatrix} \begin{bmatrix}1\\0\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix}$$

Plot them and find out which of the basis functions seen at lecture separate this data.

Solution

```
c1 <- matrix(c(0,0,1,2,2,1),ncol=2,byrow=TRUE)
c2 <- matrix(c(1,1,1,0,0,1),ncol=2,byrow=TRUE)
plot(c(0,3),c(0,3),type="n")
text(c1,"x")
text(c2,"o")
library(e1071)
D <- data.frame(rbind(c1, c2), y = c(rep("x", 3), rep("o", 3)))
m <- svm(y~.,data=D,type="C-classification",kernel="polynomial",degree=2,
    cost=1000)
plot(m,D)
```

A polynomial of degree 2 suffices.

Exercise 2

Task 3 of Exam 2010.

Solution We would need to recognize a Kernel that corresponds to the feature mapping ϕ in such a way that $K(\mathbf{x}_j, \mathbf{x}_l) = \phi(\mathbf{x}_j)^T \cdot \phi(\mathbf{x}_l)$.
The computations where the Kernel trick enters are (**??**) and in prediction $h(\mathbf{x}, \hat{\boldsymbol{\theta}}) = \text{sign}(\hat{\boldsymbol{\theta}}^T \mathbf{x}) = \text{sign}(\sum_j \hat{\alpha}_j y^j \mathbf{x}^j \cdot \mathbf{x})$, which becomes $\text{sign}(\sum_j \hat{\alpha}_j y^j K(\mathbf{x}^j, \mathbf{x}))$.

Solution $(\mathbf{x} \cdot \mathbf{z})^2 = (x_1 z_1)^2 + 2(x_1 x_2)(z_1 z_2) + (x_2 z_2)^2$ so that

$$K(\mathbf{x}, \mathbf{z}) = x_1 z_1 + x_2 z_2 + 4(x_1 z_1)^2 + 8(x_1 x_2)(z_1 z_2) + 4(x_2 z_2)^2 \quad (1)$$
$$= [x_1, x_2, 2x_1^2, 2\sqrt{2}x_1 x_2, 2x_2^2] \cdot [x_1, x_2, 2x_1^2, 2\sqrt{2}x_1 x_2, 2x_2^2] \quad (2)$$

Thus $\phi(\mathbf{x}) = [x_1, x_2, 2x_1^2, 2\sqrt{2}x_1 x_2, 2x_2^2]$.

Solution I b (not a because less margin)

II a
III c, d
IV e

Exercise 3
In R support vector machines are implemented in the function `svm` of the `e1071` package. Read the documentation of that function, try out the examples and understand them with reference to the theory.

Solution See answer to exercise 1