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

February 22, 2013
Marco Chiarandini

DM825 - Introduction to Machine Learning

Sheet 8, Spring 2013

Exercise 1 – Support vector machines

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

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

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

Plot them and find the optimal separating line. What are the support vectors, and what is the margin.

Exercise 2

Consider the four input points of an XOR function. Find a way to separate out the true cases from the false ones by using a linear separator.

Solution

Introduce a third input:

In 1	In 2	In 3 (1x2)	Output
0	0	0	1
0	1	0	0
1	0	0	0
1	1	1	1

Now we are in three dimensions and the points are linearly separable.
Alternatively: change targets to have -1 and 1 values and introduce a basis of all terms up to quadratic in the two features:

$$1, \sqrt{2}x_1, \sqrt{2}x_2, x_1x_2, x_1^2, x_2^2$$

Then we solve:

$$\max_{\vec{\alpha}} = \sum_{i=1}^{4} \alpha_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} y^i y^j \alpha_i \alpha_j \langle \vec{\phi}(\vec{x}^i), \vec{\phi}(\vec{x}^j) \rangle$$

subject to $\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = 0, \alpha +_i \geq 0, i = 1, \ldots, 4$. It can be solved algebraically and gives a classifier line at $x_1 x_2 = 0$.

Exercise 3

Show that, irrespective of the dimensionality of the data space, a data set consisting of just two data points, one from each class, is sufficient to determine the location of the maximum-margin hyperplane.

Solution

Given a data set of two data points, $x_1 \in C^+(t1 = +1)$ and $x_2 \in C^-(t2 = -1)$, the maximum margin hyperplane is determined by solving

$$\arg\min_{\vec{\theta},\theta_0} \frac{1}{2} \parallel \vec{\theta} \parallel^2$$

subject to the constraints

$$\begin{aligned} \vec{\theta}^T\vec{x}_1 + \theta_0 &= +1 \\ \vec{\theta}^T\vec{x}_2 + \theta_0 &= -1 \end{aligned}$$

We do this by introducing Lagrange multipliers α_1 and α_2, and solving

$$\arg\min_{\vec{\theta},\theta_0} \frac{1}{2} \parallel \vec{\theta} \parallel^2 + \alpha_1(\vec{\theta}^T\vec{x}_1 + \theta_0 - 1) + \alpha_2(\vec{\theta}^T\vec{x}_1 + \theta_0 + 1)$$

Taking the derivative of this w.r.t. $\vec{\theta}$ and θ and setting the results to zero, we obtain

$$\begin{aligned} 0 &= \vec{\theta} + \alpha_1 x_1 + \alpha_2 x_2 \\ 0 &= \alpha_1 + \alpha_2 \end{aligned}$$

which immediately gives $\alpha_1 = -\alpha_2$, and $\vec{\theta} = \alpha_1(x1 - x2)$.
For θ_0, we solve:

$$2\theta_0 = -\vec{\theta}^T(x_1 + x_2)$$

Note that the Lagrange multiplier α_1 remains undetermined, which reflects the inherent indeterminacy in the magnitude of $\vec{\theta}$ and θ_0.