Machine Learning Worksheet 7

Kernels and Support Vector Machines

You need to solve at least one Problem from Section 1, one from Section 3 and one from Section 4 to get full credit for this assignment. Please hand in your solutions by Dec. 21st, 2011.

1 Optimisation & convexity

As it happens with most machine learning methods, training an SVM means in the end to optimise a cost function / cost functional. And remember: one of the biggest problems with optimisation is local extrema - there is no general way of solving it. It turns out that one of the most interesting characteristics of SVMs is that their cost functional is *convex*.

Problem 1. What is a convex function? What is a *strictly* convex function? Give a precise mathematical definition.

An alternative, very comfortable and intuitive definition of convexity is: a convex function will always lie *above* its tangent linear manifolds - a convex function of one variable lies above all its tangent lines; a convex function of two variables lies above all its tangent planes, etc.

Problem 2. What about the function's extremum points?

Problem 3. Give examples of a non-strictly convex function and a strictly convex function. Definitions *and* graphs please; Matlab graphs are good, too.

2 The SVM cost function $\star$

Let us consider the classical SVM formulation for classification of two classes. Given our old friend the data set $\mathcal{D} = \{x_i, z_i\}_{i=1}^n$ where z_i is either 1 or -1, we want to minimise

$$L_P = \frac{1}{2}||\mathbf{w}||^2 - \sum_i \alpha_i z_i(\mathbf{w}^T\mathbf{x} + b) + \sum_i \alpha_i$$

w.r.t. $\mathbf{w}, b$. (Never mind the conditions on the αs in this case.)

Problem 4. Prove that L_P is indeed convex w.r.t. $\mathbf{w}, b$. (This is easier than one might think.)

3 An example

We are given four points in $\mathbb{R}^2$, $\{(0,0),(1,1),(2,2),(1,3)\}$ with labels $\{1,1,-1,-1\}$.

Problem 5. Evaluate graphically the optimal separating hyperplane. How much is the margin? What samples are support vectors? What are the $\mathbf{w}, b$ in the equation of the hyperplane, $\mathbf{w}^T\mathbf{x} + b$?

Problem 6. Write down the dual form of f (assume the αs and b are still unknown).

4 Another example

Now you are given a data set with data from a single feature x_1 in $\mathbb{R}$ and corresponding labels $y \in \{+1,-1\}$. Data points for $+1$ are at $-3,-2,3$ and data points for -1 are at $-1,0,1$.

Problem 7. Can this data set in its current feature space be separated using a linear separator? Why/why not?

Let's define a simple feature map $\Phi(u) = (u, u^2)$ that transforms points in $\mathbb{R}$ to points in $\mathbb{R}^2$.

Problem 8. After applying Φ to the data, can it now be separated using a linear separator? Why/why not (plotting the data may help you with your answer ...)?

Problem 9. *Construct* a maximum-margin separating hyperplane (i.e. you do not need to solve a quadratic program). Clearly mark the support vectors. Also draw the resulting decision boundary in the original feature space. Is it possible to add another point to the training set in such a way, that the hyperplane *does not* change? Why/why not?