Homework 3

(due Friday, October 14, 2004)

1. Let the training set comprise $(y_1, x_1), \ldots, (y_l, x_l)$, $x \in R^n$, $y \in \{-1, 1\}$. Define a unit vector ϕ and a constant c such that

$$(x_i * \phi) > c, \text{ if } x_i \in C_1 \text{ and } (x_j * \phi) < c, \text{ if } x_j \in C_2.$$

For any unit vector ϕ, we also define the two values

$$c_1(\phi) = \min_{x_i \in C_1} (x_i * \phi), \; c_2(\phi) = \max_{x_j \in C_2} (x_j * \phi).$$

Based on the above, the *margin* $\rho(\phi) = \frac{c_1(\phi) - c_2(\phi)}{2}$.

Now define a vector ψ and a threshold b such that

$$(x_i * \psi) + b \geq 1, \text{ if } y_i = 1, \text{ and } (x_j * \psi) + b \leq -1, \text{ if } y_i = -1.$$

The above equation is valid if the data are linearly separable which you may assume. Now let $\psi = \frac{\phi}{\rho(\phi)}$.

- Find an appropriate value of b such that the patterns are linearly separable.
- For this choice of ψ, evaluate $(\psi * \psi)$.

2. This question is designed to help you understand the transition from $L(\psi, b, \alpha)$ to $W(\alpha)$. Given that the optimal $\psi_0 = \sum_{i=1}^{l} \alpha_i^0 y_i x_i$,

- Show that the optimal $(\psi_0 * \psi_0) = (\sum_{i=1}^{l} \alpha_i^0 y_i x_i) * (\sum_{j=1}^{l} \alpha_j^0 y_j x_j) = \sum_{i=1}^{l} \alpha_i^0$ by using the constraints which are satisfied by the optimal α_0, namely $\sum_{i=1}^{l} \alpha_i^0 \left[y_i \left(\{x_i * \psi\} + b_0 \right) - 1 \right] = 0$ and $\sum_{i=1}^{l} \alpha_i^0 y_i = 0$.
- Since the optimal $\rho(\phi_0) = \frac{1}{|\psi_0|}$, what is $\rho(\phi_0)$ in terms of α_0?
- Show that $W(\alpha_0) = \frac{1}{2} \sum_{i=1}^{l} \alpha_i^0$. Relate $W(\alpha_0)$ and $\rho(\phi_0)$. What does this mean to you?

- How do you determine the optimal b_0?
- Reconcile the apparent contradiction between seeking the maximum of $\rho(\phi)$ and the maximum of $W(\alpha)$.

3. This question is designed to help you better understand the geometry of the SVM objective function.

- For a standard two class problem, determine the hyperplane of the two "riverbanks" corresponding to C_1 and C_2 in terms of both ϕ and ψ.
- Given a set of linearly separable patterns x_i, can you determine an origin x_0 such that $(x_i - x_0) * \phi_0 > 0$ for $x_i \in C_1$ and $(x_j - x_0) * \phi_0 < 0$ for $x_j \in C_2$? [You only need to do this for the optimal vector ϕ_0].
- What is the relationship between x_0 and the "bias" b_0 [in the (ψ_0, b_0) language]?

4. A pattern $x^{(n)} \in R^D$, $n \in \{1, \ldots, N\}$ is defined as $\{x_1^{(n)}, \ldots, x_D^{(n)}\}$.

- For a set of two dimensional patterns, $x^{(n)} \in R^2$, show that the polynomial kernel $k(x, y) = (x \cdot y + 1)^3$ is non-negative definite.
- Also, show that the polynomial kernel $k(x, y) = (x \cdot y + 1)^d$ is non-negative definite for D dimensional patterns. You'll need to expand $(\sum_{i=1}^D x_i^{(m)} x_i^{(n)} + 1)^d = \sum_{d_0=0}^D \cdots \sum_{d_D=0}^D \frac{d}{\prod_{i=0}^D d_i!} \prod_{i=0}^D (p_i)^{d_i}$, where $p_0 = 1$, $p_i = x_i^{(m)} x_i^{(n)}$, $i > 0$ and $\sum_{i=0}^D d_i = d$. This problem is much easier than you think.