18.465 PS8 due Thursday, May 12, 2005

The problems relate to "A tutorial on support vector machines for pattern recognition" by C. J. C. Burges.

1. In $\mathbb{R}^2$ suppose given a training sample with $l = 6$. A set A of three X_i with $y_i = +1$ consists of $(-1, 2)$, $(-3, 1)$, and $(2, -1)$. The set B of the other three X_i, with $y_i = -1$, consists of $(0, 3)$, $(1, 1)$, and $(4, 0)$. Let $co(C)$ be the convex hull of a set C, namely the smallest convex set including C, specifically for $C = A$ or B.
(a) Show that $co(A)$ and $co(B)$ do not intersect. (Parts (b) and (c) will make this more specific.)
(b) Find two points u and v such that $u \in co(A)$, $v \in co(B)$, and $|u - v|$ is as small as possible. Are such u and v unique?
(c) Find two parallel lines L_1 and L_3 such that L_1 intersects A, L_3 intersects B, there are no points of A or B between L_1 and L_3, and the (perpendicular) distance from L_1 to L_3 is as large as possible. *Hint*: the lines are perpendicular to the line through u and v.

2. (Continuation) (a) Support vectors are points of $L_1 \cap A$ or $L_3 \cap B$. What are they?
(b) Find a vector w and number b such that inequalities (10) and (11) on p. 129 of Burges hold (with $x_i \equiv X_i$) and the length $\|w\|$ is as small as possible. *Hint*: the inequalities should become equalities at each support vector, and the equalities should be equations defining the lines L_3 and L_1.
(c) Let L_2 be a line parallel to L_1 and L_3 and halfway between them. If a new X is observed with unknown y, the y will be predicted to be $+1$ or -1 depending on which side of L_2 X is on, the A side or the B side respectively. If $X = (0.31, 0.17)$, what is the predicted y?
(d) Burgers, p. 130, in the second line after (13), says that at a solution, the partial derivatives of L_P with respect to all the α_i vanish. Show, however, that this is only true (in the general case, not only the special one we've been treating) for those i such that x_i is a support vector. As mentioned in the last paragraph of p. 130, if x_i is not a support vector then $\alpha_i = 0$. So, write L_D in (16) in our special case, where only the non-zero α_i need to be included. Find those α_i that are > 0 by maximizing L_D in this case over such α_i.

3. This is about structural risk minimization as in §2.6, p. 128, of Burgers, where an inequality $h_1 < h_2 < \cdots$ is displayed. One VC class may be effectively included in another with the same dimension. For example, in $\mathbb{R}^d$, let $\mathcal{C}_1$ be the class $\mathcal{H}_d$ of all open half-spaces, and let $\mathcal{C}_2$ be the class of all balls.
(a) Show that for any finite set F and any $H \in \mathcal{C}_1$ there is a ball $B \in \mathcal{C}_2$ with $B \cap F = H \cap F$ (consider balls with large enough radius and distant enough centers to approximate H).
(b) Show that for each d, for some F there is a ball B such that $B \cap F \neq H \cap F$ for any half-space H.
(c) Conclude that the empirical risk as in (2) is always at least as small for $\mathcal{C}_2$ as it is for $\mathcal{C}_1$.
(d) Show (or find a proof already given) that $S(\mathcal{C}_1) = S(\mathcal{C}_2)$ for each d.

(d) Show that the upper bound for $R(\alpha)$ given by Vapnik's inequality (3), p. 123, is always as small or smaller for $\mathcal{C}_2$ than it is for $\mathcal{C}_1$. So there is no need to compute both and compare them.