COS 511: Foundations of Machine Learning

Homework #2 | Due in class: February 27, 2003

Problem 1

This problem explores another general method for bounding the error when the hypothesis space is infinite.

Some algorithms output hypotheses that can be represented by a small number of examples from the training set. For instance, suppose the domain is $\mathbb{R}$ and we are learning a half-line of the form $x \geq a$ where a defines the half-line. A simple algorithm chooses the left most positive training example a and outputs the corresponding half-line, which is clearly consistent with the data. Thus, in this case, the hypothesis can be represented by a single training example.

More formally, let F be a function mapping labeled examples to concepts, and assume that algorithm A, when given training examples $(x_1, c(x_1)), \ldots, (x_m, c(x_m))$ labeled by some unknown $c \in \mathcal{C}$, chooses some $i_1, \ldots, i_k \in \{1, \ldots, m\}$ and outputs the consistent hypothesis $F((x_{i_1}, c(x_{i_1})), \ldots, (x_{i_k}, c(x_{i_k})))$. In a sense, the algorithm has "compressed" the sample down to a sequence of just k of the m training examples.

a. Give such an algorithm for axis-aligned hyper-rectangles in $\mathbb{R}^n$ with $k = O(n)$. (An axis-aligned hyper-rectangle is a set of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$. For $n = 2$, this is the class of rectangles used repeatedly as an example in class.) Your algorithm should run in time polynomial in m and n.

b. As usual, assume that the examples are chosen at random from some distribution D. Also assume that the size k is fixed. Argue carefully that the error of the output hypothesis h, with probability at least $1 - \delta$ satisfies the bound:

$$\mathrm{err}_D(h) \leq O\left(\frac{\ln(1/\delta) + k \ln m}{m - k}\right).$$

(Hint: Fix $i_1, \ldots, i_k$ above, and bound the probability of the hypothesis defined by these k examples being consistent but ϵ-bad. Then apply the union bound.)

Problem 2

Let the domain be $\mathbb{R}^d$, and consider the class $\mathcal{C}$ of linear threshold functions passing through the origin. That is, each such function is defined by a vector $\mathbf{w} \in \mathbb{R}^d$ and is equal to 1 on points $\mathbf{x}$ for which $\mathbf{w} \cdot \mathbf{x} \geq 0$, and 0 on all other points. Show that the VC-dimension of $\mathcal{C}$ is exactly equal to d. (Although you can use whatever technique you wish, you might find it helpful to think about the fact that any $d + 1$ points $\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}$ in $\mathbb{R}^d$ must be linearly dependent, i.e., there exists scalars $\lambda_1, \ldots, \lambda_{d+1}$, not all equal to zero, such that $\sum_{i=1}^{d+1} \lambda_i \mathbf{x}_i = \mathbf{0}$.)

Problem 3

For each $d = 0, 1, 2, \ldots$, give an example of a class $\mathcal{C}$ for which Sauer's Lemma is tight, i.e., for which the VC-dimension of $\mathcal{C}$ is d, and, for each m, $\Pi_{\mathcal{C}}(m) = \sum_{i=0}^{d} \binom{m}{i}$.