Machine Learning - Assignment 4

Extended to Monday, November 4, 2002

1. [10 points] Mitchell, pg. 152 Problem 5.2

2. [30 points] Compute the VC dimension for the following classes of hypotheses:

   (a) [10 points] Triangles in the x-y plane. Points inside the triangle are classified as positive examples.

   (b) [20 points] Convex polygons in the x-y plane. Points inside the polygon are classified as positive examples.

3. [30 points] Mitchell, pg. 228 Problem 7.8.

4. [30 points] **Nearest-neighbor algorithm**

   This question requires you to do a small PAC-style analysis for a nearest-neighbor classifier. Suppose the instances are drawn from $[0,1]^n$. A hypothesis $h : [0,1]^n \rightarrow [0,1]$ is defined as a set of labeled prototype points $\langle y_1, c_1 \rangle, \ldots \langle y_N, c_N \rangle$. New examples are classified according to the nearest prototype in the list. That is, we classify a new instance $x \in [0,1]^n$ using label c_k, where $k = \arg\min_{i=1,N} \max_{j=1}^{n} |y_{ij} - x_j|$. This is called the L_∞ norm (different from the Euclidean norm we used in class).

   (a) Assume that the test instances are drawn from a uniform distribution over $[0,1]^n$. For any point $y \in [0,1]^n$, what is the probability that a random test instances x falls within ϵ of y? I.e., what is the probability that $\max_{j=1}^{n} |y_j - x_j| \leq \epsilon$?. (You can assume that ϵ is less than the smallest distance from y to any boundary of $[0,1]^n$.

   Assuming that the training points are also drawn uniformly randomly, what is a lower bound on the number of training points (prototypes) needed to guarantee that any instance falls within ϵ of a prototype with probability at least $(1 - \delta$?

   (b) Let H_n be the class of hypotheses generated by this algorithm. What is the VC dimension of H_n?

(c) Now consider a restricted class of hypotheses, H_c, defined by prototypes at the corners of the unit cube $[0, 1]^n$ only. How many corners are there? What is the cardinality of H_c? What is the VC dimension of H_c?

(d) Now consider a different hypothesis class, H_2, in which we use only two prototypes, and we use Euclidean distance to evaluate closeness. What is the cardinality and VC dimension of H_2?

(e) For each of the hypotheses classes H_n, H_2, H_c, given an arbitrary target function f from the hypotheses class, and an arbitrary distribution on $[0, 1]^n$, what is a sufficient number of training examples to guarantee that f can be learned within ϵ accuracy with probability at least $(1 - \delta)$? Express the answer as a function of n, ϵ and δ.

(f) **Extra credit (5 points)** Can you identify a target function f in H_c that would cause the nearest neighbor algorithm described above to perform poorly? Give your intuition (a formal proof is not needed).