Machine Learning - Assignment 1

Due Thursday, September 20, 2001

1. [30 points] Mitchell, pg. 50, Problem 2.8.

2. [30 points] Mitchell, pg. 50, Problem 2.9.

3. (a) [10 points] Mitchell, pg. 227, Problem 7.2(a).

   (b) [10 points] Consider the more general case in which the rectangle boundaries, a, b, c, d, are real numbers. Propose a learning algorithm for this case.

   (c) [10 points] Provide a lower bound on the number of training examples needed for your algorithm to achieve error at most ϵ with probability at least $(1 - \delta)$. Hint: the error in this case is related to the area difference between the concept rectangle and the learned rectangle; you may consider using some of the inequalities mentioned in lecture 3

   (d) [10 points] Implement your rectangle learning algorithm. Generate 30 random concepts (rectangles) to be learned, with coordinates between 0 and 100. For each one of these, generate training examples by drawing points with coordinates between 0 and 100 uniformly randomly. Compute the error between the "learned" rectangle and the true one for $m = 0, 10, 50, 100, 500, 1000, 5000, 10000$, training examples. Plot, on the same graph:

   - the error averaged over the 30 random concepts,
   - the minimum error over the 30 concepts
   - the maximum error over the 30 concepts
   - the ϵ values for the number of examples m specified above, for $\delta = 0.95$ (you can get this from the formula you provide for the previous part of the problem).

   Does the theory fit the data? Explain your results.