EXERCISES

7.1. Consider training a two-input perceptron. Give an upper bound on the number of training examples sufficient to assure with 90% confidence that the learned perceptron will have true error of at most 5%. Does this bound seem realistic?

7.2. Consider the class C of concepts of the form $(a \le x \le b) \wedge (c \le y \le d)$, where a, b, c, and d are integers in the interval (0, 99). Note each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the x, y plane. Hint: Given a region in the plane bounded by the points $(0, 0)$ and $(n-1, n-1)$, the number of distinct rectangles with integer-valued boundaries within this region is $(\frac{n(n+1)}{2})^2$.

(*a*) Give an upper bound on the number of randomly drawn training examples sufficient to assure that for any target concept c in C, any consistent learner using $H = C$ will, with probability 95%, output a hypothesis with error at most .15.

(*b*) Now suppose the rectangle boundaries a, b, c, and d take on *real* values instead of integer values. Update your answer to the first part of this question.

7.3. In this chapter we derived an expression for the number of training examples sufficient to ensure that every hypothesis will have true error no worse than ϵ plus its observed training error $error_D(h)$. In particular, we used Hoeffding bounds to derive Equation (7.3). Derive an alternative expression for the number of training examples sufficient to ensure that every hypothesis will have true error no worse than $(1+\gamma)error_D(h)$. You can use the general Chernoff bounds to derive such a result.

Chernoff bounds: Suppose $X_1, \ldots, X_m$ are the outcomes of m independent coin flips (Bernoulli trials), where the probability of heads on any single trial is $\Pr[X_i = 1] = p$ and the probability of tails is $\Pr[X_i = 0] = 1 - p$. Define $S = X_1 + X_2 + \cdots + X_m$ to be the sum of the outcomes of these m trials. The expected value of S/m is $E[S/m] = p$. The Chernoff bounds govern the probability that S/m will differ from p by some factor $0 \le \gamma \le 1$.

$$\Pr[S/m > (1+\gamma)p] \le e^{-mp\gamma^2/3}$$

$$\Pr[S/m < (1-\gamma)p] \le e^{-mp\gamma^2/2}$$

4. Consider a learning problem in which $X = \Re$ is the set of real numbers, and $C = H$ is the set of intervals over the reals, $H = \{(a < x < b) \mid a, b \in \Re\}$. What is the probability that a hypothesis consistent with m examples of this target concept will have error at least ϵ? Solve this using the VC dimension. Can you find a second way to solve this, based on first principles and ignoring the VC dimension?

7.5. Consider the space of instances X corresponding to all points in the x, y plane. Give the VC dimension of the following hypothesis spaces:

(*a*) H_r = the set of all rectangles in the x, y plane. That is, $H = \{((a < x < b) \wedge (c < y < d)) | a, b, c, d \in \Re\}$.

(*b*) H_c = circles in the x, y plane. Points inside the circle are classified as positive examples

(*c*) H_t =triangles in the x, y plane. Points inside the triangle are classified as positive examples

7.6. Write a consistent learner for H_r from Exercise 7.5. Generate a variety of target concept rectangles at random, corresponding to different rectangles in the plane. Generate random examples of each of these target concepts, based on a uniform distribution of instances within the rectangle from $\langle 0, 0 \rangle$ to $\langle 100, 100 \rangle$. Plot the generalization error as a function of the number of training examples, m. On the same graph, plot the theoretical relationship between ϵ and m, for $\delta = .95$. Does theory fit experiment?

7.7. Consider the hypothesis class H_{rd2} of "regular, depth-2 decision trees" over n Boolean variables. A "regular, depth-2 decision tree" is a depth-2 decision tree (a tree with four leaves, all distance 2 from the root) in which the left and right child of the root are *required to contain the same variable.* For instance, the following tree is in H_{rd2}.

```
          x3
         /   \
       x1     x1
      / \    / \
     +   -  -   +
```

(*a*) As a function of n, how many syntactically distinct trees are there in H_{rd2}?

(*b*) Give an upper bound for the number of examples needed in the PAC model to learn H_{rd2} with error ϵ and confidence δ.

(*c*) Consider the following WEIGHTED-MAJORITY algorithm, for the class H_{rd2}. You begin with all hypotheses in H_{rd2} assigned an initial weight equal to 1. Every time you see a new example, you predict based on a weighted majority vote over all hypotheses in H_{rd2}. Then, instead of eliminating the inconsistent trees, you cut down their weight by a factor of 2. How many mistakes will this procedure make at most, as a function of n and the number of mistakes of the best tree in H_{rd2}?

7.8. This question considers the relationship between the PAC analysis considered in this chapter and the evaluation of hypotheses discussed in Chapter 5. Consider a learning task in which instances are described by n boolean variables (e.g., $x_1 \wedge \bar{x_2} \wedge x_3 \ldots \bar{x_n}$) and are drawn according to a fixed but unknown probability distribution $\mathcal{D}$. The target concept is known to be describable by a conjunction of boolean attributes and their negations (e.g., $x_2 \wedge \bar{x_5}$), and the learning algorit