Exercise 5.1

Consider hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_{100}$ that contain Boolean functions on 10 Boolean variables, so $\mathcal{X} = \{-1, +1\}^{10}$. $\mathcal{H}_1$ contains all Boolean functions which evaluate to $+1$ on exactly one input point, and to -1 elsewhere; $\mathcal{H}_{100}$ contains all Boolean functions which evaluate to $+1$ on exactly 100 input points, and to -1 elsewhere.

(a) How big (number of hypotheses) are $\mathcal{H}_1$ and $\mathcal{H}_{100}$?

(b) How many bits are needed to specify one of the hypotheses in $\mathcal{H}_1$?

(c) How many bits are needed to specify one of the hypotheses in $\mathcal{H}_{100}$?

Exercise 5.2

Suppose that for 5 weeks in a row, a letter arrives in the mail that predicts the outcome of the upcoming Monday night football game. You keenly watch each Monday and to your surprise, the prediction is correct each time. On the day after the fifth game, a letter arrives, stating that if you wish to see next week's prediction, a payment of $50.00 is required. Should you pay?

(a) How many possible predictions of win-lose are there for 5 games?

(b) If the sender wants to make sure that at least one person receives correct predictions on all 5 games from him, how many people should he target to begin with?

(c) After the first letter 'predicting' the outcome of the first game, how many of the original recipients does he target with the second letter?

(d) How many letters altogether will have been sent at the end of the 5 weeks?

(e) If the cost of printing and mailing out each letter is $0.50, how much would the sender make if the recipient of 5 correct predictions sent in the $50.00?

(f) Can you relate this situation to the growth function and the credibility of fitting the data?

Exercise 5.3

In an experiment to determine the distribution of sizes of fish in a lake, a net might be used to catch a representative sample of fish. The sample is

then analyzed to find out the fractions of fish of different sizes. If the sample is big enough, statistical conclusions may be drawn about the actual distribution in the entire lake. Can you smell ☺ sampling bias?

Exercise 5.4

Consider the following approach to learning. By looking at the data, it appears that the data is linearly separable, so we go ahead and use a simple perceptron, and get a training error of zero after determining the optimal set of weights. We now wish to make some generalization conclusions, so we look up the d_{vc} for our learning model and see that it is $d+1$. Therefore, we use this value of d_{vc} to get a bound on the test error.

(a) What is the problem with this bound   is it correct?

(b) Do we know the d_{vc} for the learning model that we actually used? It is this d_{vc} that we need to use in the bound.

Exercise 5.5

Assume we set aside 100 examples from $\mathcal{D}$ that will not be used in training, but will be used to select one of three final hypotheses g_1, g_2, g_3 produced by three different learning algorithms that train on the rest on the data. Each algorithm works with a different $\mathcal{H}$ of size 500. We would like to characterize the accuracy of estimating $E_{\text{out}}(g)$ on the selected final hypothesis if we use the same 100 examples to make that estimate.

(a) What is the value of M that should be used in (1.6) in this situation?

(b) How does the level of contamination of these 100 examples compare to the case where they would be used in training rather than in the final selection?

Problem 5.1 The idea of *falsifiability* – that a claim can be rendered false by observed data – is an important principle in experimental science.

> **Axiom of Non-Falsifiability.** *If the outcome of an experiment has no chance of falsifying a particular proposition, then the result of that experiment does not provide evidence one way or another toward the truth of the proposition.*

Consider the proposition "There is $h \in \mathcal{H}$ that approximates f as would be evidenced by finding such an h with in sample error zero on $\mathbf{x}_1, \cdots, \mathbf{x}_N$." We say that the proposition is falsified if no hypothesis in $\mathcal{H}$ can fit the data perfectly.

(a) Suppose that $\mathcal{H}$ shatters $\mathbf{x}_1, \cdots, \mathbf{x}_N$. Show that this proposition is not falsifiable *for any* f.

(b) Suppose that f is random ($f(\mathbf{x}) = \pm 1$ with probability $\frac{1}{2}$, independently on every $\mathbf{x}$), so $E_{\text{out}}(h) = \frac{1}{2}$ for every $h \in \mathcal{H}$. Show that

$$\mathbb{P}[\text{falsification}] \geq 1 - \frac{m_{\mathcal{H}}(N)}{2^N} .$$

(c) Suppose $d_{\text{vc}} = 10$ and $N = 100$. If you obtain a hypothesis h with zero E_{in} on your data, what can you 'conclude' from the result in part (b)?