Homework 2, ECE 6254, Spring 2014

Due: Wednesday Feb 5, at the beginning of class

Problems:

1. Consider the hypothesis set consisting of positive rectangles in $\mathbb{R}^2$, i.e., h such that

$$h(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x}) = +1 & \text{if } a \le x(1) \le b \text{ and } c \le x(2) \le d \\ -1 & \text{otherwise,} \end{cases}$$

   for some $a, b, c, d \in \mathbb{R}$. Show that for this $\mathcal{H}$, $m_{\mathcal{H}}(4) = 2^4$ and $m_{\mathcal{H}}(5) < 2^5$. Use this to give an upper bound for $m_{\mathcal{H}}(n)$.

2. Prove by induction that $\sum_{i=0}^{d} \binom{n}{i} \le n^d + 1$, and hence $m_{\mathcal{H}}(n) \le n^{d_{\mathrm{VC}}} + 1$.

3. The VC dimension implicitly depends on the input space $\mathcal{X}$ as well as $\mathcal{H}$, since $\mathcal{H}$ is defined as a set of functions that map $\mathcal{X}$ to $\{+1, -1\}$. For a fixed $\mathcal{H}$, consider two input spaces $\mathcal{X}_1 \subseteq \mathcal{X}_2$. Let $d_{\mathrm{VC}}(\mathcal{X}_1)$ and $d_{\mathrm{VC}}(\mathcal{X}_2)$ denote the VC dimension of $\mathcal{H}$ with respect to these two input spaces. Show that $d_{\mathrm{VC}}(\mathcal{X}_1) \le d_{\mathrm{VC}}(\mathcal{X}_2)$.

4. Suppose that our input space $\mathcal{X} = \mathbb{R}$ and consider the hypothesis set

$$\mathcal{H} = \left\{ h : h(x) = \mathrm{sign}\left(\sum_{i=0}^{d} c_i x^i \right) \quad \text{for some } c_0, \dots c_d \in \mathbb{R} \right\}$$

   In words, $\mathcal{H}$ is the set of classifiers obtained by evaluating some polynomial of degree d and comparing the result to a threshold. Prove that the VC dimension of $\mathcal{H}$ is exactly $d+1$ by showing that

   (a) There are $d+1$ points which are shattered by $\mathcal{H}$.

   (b) There are no $d+2$ points which are shattered by $\mathcal{H}$. [Hint: Try relating this to a linear classifier in d dimensions and use the result of problem 3.]

5. Suppose that the VC dimension of our hypothesis set $\mathcal{H}$ is $d_{\mathrm{VC}} = 3$ (e.g., linear classifiers in $\mathbb{R}^2$) and that we have an algorithm for selecting some $h^* \in \mathcal{H}$ based on a training sample of size n (i.e., we have n example input-output pairs to train on).

   (a) Using the generalization bound given in class and derived in the handout, give a precise upper bound on $R(h^*)$ that holds with probability at least 0.95 in the case where $n = 100$. Repeat for $n = 1,000$ and $n = 10,000$.

   (b) Again using the generalization bound given in class, how large does n need to be to obtain a generalization bound of the form

$$R(h^*) = \widehat{R}_n(h^*) + 0.01$$

   that holds with probability at least 0.95? How does this compare to the "rule of thumb" given in class?

6. (Optional) The VC dimension "usually" corresponds to the number of parameters or degrees of freedom in the hypothesis set. However, this is not *always* true. In this problem you will show a counter-example. Consider the hypothesis set for $x \in \mathbb{R}$ with

$$\mathcal{H} = \left\{ h : h(x) = (-1)^{\lfloor \alpha x \rfloor} \quad \text{for some } \alpha \in \mathbb{R} \right\},$$

where $\lfloor \cdot \rfloor$ is the floor function (i.e., the $\lfloor \alpha x \rfloor$ is the largest integer less than αx). Thus, even though this hypothesis set has a single parameter, it still has an infinite VC dimension! [Hint: Consider the inputs of $x_i = 10^i$ for $i = 1, \ldots, n$ and show how to choose α to realize an arbitrary dichotomy.]