Theoretical Foundations of Machine Learning - Homework #2

Jacob Abernethy and Chansoo Lee   Due: 10/16/2015

Homework Policy: Working in groups is fine, but *every student* must submit their own writeup. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

1) **Efficient PAC Learning.** Let $\mathcal{X} := \{0,1\}^n$ and let $\mathcal{C}$ be the class of *monotone disjunctions*; that is, $\mathcal{C} := \{h(\mathbf{x}) = \bigvee_{i \in S} x_i \; : \; \forall S \subseteq [n]\}$ where $\vee$ is the logical OR operator. Prove that $\mathcal{C}$ is *efficiently* PAC-learnable. You need to construct an algorithm and prove that it accomplishes what you need.

2) **Learning Circles.** (Chapter 2.6 exercise 2.3) Let $\mathcal{X} := \mathbb{R}^2$ and consider the set of concepts $\mathcal{C}$ of the form $c = \{(x,y) : x^2 + y^2 \leq r^2\}$ for some real number r. Show that this class can be (ϵ, δ)-PAC-learned from training data of size $m \geq (1/\epsilon) \log(1/\delta)$.

3) **Reducing to PAC Learning.** (Chapter 2.6 exercise 2.9) We showed that for a finite hypothesis set $\mathcal{C}$, a consistent learning algorithm A is a PAC-learning algorithm. Here, we consider a converse question. Let S be a finite set of m labeled points. Suppose that you are given a PAC-learning algorithm A. Show that you can use A (that will take in finitely many points) to find in polynomial time a hypothesis $h \in \mathcal{C}$ that is consistent with S, with high probability. (Hint: you can select an appropriate distribution D over S and give a condition on $R(h)$ for h to be consistent.).

4) **Union of Intervals.** (Chapter 3.6 exercise 3.6) What is the VC-dimension of subsets of the real line formed by the union of k intervals?

5) **Ellipsoids.** (Chapter 3.6 exercise 3.10) An n-dimensional *ellipsoid* E is defined by a center point $\mathbf{x}_0 \in \mathbb{R}^n$ and a symmetric positive semidefinite matrix $M \in \mathbb{R}^{n \times n}$, such that $\mathbf{x} \in E \iff (\mathbf{x} - \mathbf{x}_0)^\top M (\mathbf{x} - \mathbf{x}_0) \leq 1$.

   (a) Give a quadratic upper bound on the VC dimension of concept class of ellipsoids in $\mathbb{R}^n$. That is, show that the VCD $= O(n^2)$.

   (b) (**Challenge**) Give an exact characterization of the VC dimension of ellipsoids.

6) **More Chernoff.** (**Challenge**) Let's say the random variables $X_1, \ldots, X_n$ are iid Bernoulli with $\Pr(X_i = 1) = \epsilon$. We want to bound the probability that $\sum_{i=1}^n X_i < \frac{n\epsilon}{2}$. The standard Hoeffding inequality gives us that
$$\Pr\left(\sum_{i=1}^n X_i < \tfrac{n\epsilon}{2}\right) \leq \Pr\left(\left|\sum_{i=1}^n X_i - \epsilon n\right| > \tfrac{n\epsilon}{2}\right) \leq 2\exp(-n\epsilon^2/2),$$

which suggests we need $n > 1/\epsilon^2$ before we can get any decay in probability. But you can get something tighter. Find a way to obtain the following bound, using the Chernoff technique or otherwise:

$$\Pr\left(\sum_{i=1}^{n} X_i < \tfrac{n\epsilon}{2}\right) \le \exp(-O(n\epsilon)).$$