15-854 Machine Learning Theory

Homework # 3 **Due: February 16, 1998**

Exercises:

1. In the months before the 1992 presidential election, reporters were sent to interview the residents of Crook County, Oregon, to find out who they planned to vote for. The reason: Crook County had gone with the winner in every one of the 27 presidential elections since it was founded. Furthermore, it was the only remaining county out of 3141 counties in the United States to have this distinction. In this problem, we will try analyze the significance of this. For instance, imagine you were going to place a bet on whether Crook County would go with the winner in next election.

   Let's think of each election as an example, and each county as a hypothesis. We can put this into the PAC framework by supposing that each county i has some fixed error probability ε_i. That is, on each election, county i goes with the winner with probability $(1 - \varepsilon_i)$ and goes with a loser with probability ε_i.

   Suppose the ε_i's are all equal (i.e., all counties have the same probability of going with the winner), and we look at 27 elections. For what value of ε_i is the expected number of counties that predict correctly in all 27 elections equal to 1?

   (This suggests that there is no reason to believe that Crook County is any better of a predictor than the value you calculated.)

2. For $\delta = 0.2$, calculate a value of ε such that given 27 examples (elections), with probability $\geq 1 - \delta$ all counties of error greater than ε are inconsistent with the data. In other words, we can be reasonably confident that Crook County is at least as good a predictor as this.

   As it turns out, Crook County went with Bush in 1992, ending its winning streak.

Problems:

3. In this problem you will prove that the VC-dimension of the class H_n of halfspaces in n dimensions is $n+1$. (H_n is the set of functions $a_1x_1 + \ldots + a_nx_n \geq a_0$, where $a_0, \ldots, a_n$ are real-valued.) Recall that VC-dim(C) is the largest d such that there exists a set of d points *shattered* by C (split in all possible ways by concepts in C).

   We will use the following definition: The *convex hull* of a set of points S is the set of all convex combinations of points in S; this is the set of all points that can be written as $\sum_{x_i \in S} \lambda_i x_i$, where each $\lambda_i \geq 0$, and $\sum_i \lambda_i = 1$. It is not hard to see that if a halfspace has all points from a set S on one side, then the entire convex hull of S must be on that side as well.

   (a) **[lower bound]** Prove that VC-dim$(H_n) \geq n + 1$ by presenting a set of $n + 1$ points in n-dimensional space such that one can partition that set with halfspaces in all possible ways. (And, show how one can partition the set in any desired way.)

(b) **[upper bound part 1]** The following is "Radon's Theorem," from the 1920's.

Theorem. *Let S be a set of $n+2$ points in n dimensions. Then S can be partitioned into two (disjoint) subsets S_1 and S_2 whose convex hulls intersect.*

Show that Radon's Theorem implies that the VC-dimension of halfspaces is *at most* $n+1$. Conclude that VC-dim$(H_n) = n+1$.

(c) **[upper bound part 2]** Now we prove Radon's Theorem. We will need the following standard fact from linear algebra. If $x_1, \ldots, x_{n+1}$ are $n+1$ points in n-dimensional space, then they are linearly dependent. That is, there exist real values $\lambda_1, \ldots, \lambda_{n+1}$ *not all zero* such that $\lambda_1 x_1 + \ldots + \lambda_{n+1} x_{n+1} = 0$.

You may now prove Radon's Theorem however you wish. However, as a suggested first step, prove the following. For any set of $n+2$ points $x_1, \ldots, x_{n+2}$ in n-dimensional space, there exist $\lambda_1, \ldots, \lambda_{n+2}$ *not all zero* such that $\sum_i \lambda_i x_i = 0$ and $\sum_i \lambda_i = 0$. (This is called *affine dependence.*) Now, think about the lambdas...