Groundrules

- Homeworks will generally consist of *exercises*, easier problems designed to give you practice, and *problems*, that may be harder, trickier, and/or somewhat open-ended. You should do the exercises by yourself, but you may work with a friend on the harder problems if you want. One exception: no fair working with someone who has already figured out (or already knows) the answer. If you work with a friend, then write down who you are working with.
- If you've seen a problem before (sometimes we'll give problems that are "famous"), then say that in your solution (it won't affect your score, we just want to know). Also, if you use any sources other than the textbook, write that down too (it's fine to look up a complicated sum or inequality or whatever, but don't look up an entire solution).

Exercises

1. What is the VC-dimension d of axis-parallel rectangles in R^3? Specifically, a legal target function is specified by three intervals $[x_{min}, x_{max}]$, $[y_{min}, y_{max}]$, and $[z_{min}, z_{max}]$, and classifies an example (x, y, z) as positive iff $x \in [x_{min}, x_{max}]$, $y \in [y_{min}, y_{max}]$, and $z \in [z_{min}, z_{max}]$. Recall that VC-dimension is the size of the largest set S of points that can be labeled in all possible ways (shattered) using functions in the class. Be sure to argue why no set of $d + 1$ points can be shattered.

Problems

1. Below, 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.) 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...

2. **(These go to eleven.)** In class we saw (most of) the construction of amplification for BPP algorithms (see the notes and/or book for the full argument). For this problem, let us abstract out a useful lemma using essentially the same techniques we used there, and use that to prove a weaker result, that of amplifying success probability for RP algorithms.

   (a) Given a regular graph $G = (V, E)$, let M be the transition matrix of the natural random walk on G. Recall that the top eigenvalue of this matrix is 1; suppose $\max(|\lambda_2|, |\lambda_n|) \le \epsilon$. Fix some subset $S \subseteq V$ with size $\delta|V|$, with $\delta < 1$.

   Show that if we pick a uniformly random vertex in G, and take a random walk of length k starting at this vertex, the probability that we only see vertices in set S is at most $(\epsilon + \sqrt{\delta})^k$.

   (b) Consider an RP algorithm A that uses r bits of randomness; i.e., one with $\Pr_{R \in \{0,1\}^r}[A(x, R) = 1] = 0$ for x not in the language, and $\Pr_{R \in \{0,1\}^r}[A(x, R) = 1] \ge 1/2$ for x in the language. (Note that R is the randomness used by the algorithm.)

   Here is a (slightly) simpler amplification procedure for such an RP algorithm.

   > *Build a constant-degree expander graph on r-bit strings. Pick a random initial vertex $v_0 \in \{0,1\}^r$, and do a random walk $v_0, v_1, \ldots, v_k$ of length k. Run the algorithm $A(x, \cdot)$ using each of the $k+1$ strings $v_0, v_1, \ldots, v_k$ as the "random" strings—if the algorithm accepts on any of these strings, then output "$x \in L$", else reject and output "$x \notin L$".*

   This algorithm clearly uses $r + O(k)$ bits of randomness, and if $x \notin L$, we always reject. Use the first part to show that if $x \in L$, then we reject (and hence make a mistake) with probability at most $2^{-\Omega(k)}$.