TTIC 31250 An Introduction to the Theory of Machine Learning

Homework # 3 **Due: May 6, 2020**

Groundrules: Same as before. You should work on the exercises by yourself but may work with others on the problems (just write down who you worked with). Also if you use material from outside sources, say where you got it.

Exercises:

1. Think about what you would like to do for your project and propose it. Some ideas:

   - Read a paper from a recent COLT conference (say COLT 2014 through COLT 2019) and write a 4-5 page explanation of what it does. Papers can be found at: `http://www.learningtheory.org/`
   - Read a paper on learning theory from a recent related conference (ICML, NeurIPS) and write a 4-5 page explanation of what it does.
   - Think about a theoretical question, which could be modeling some machine learning setting, trying to give sufficient conditions for some approach to succeed, looking at a different model for how examples are selected or the kind of feedback the algorithm is given, etc. Write up your thoughts in 4-5 pages.
   - Conduct an experiment to compare different approaches to some problem. (Note: "your" approach doesn't have to turn out to be the best one!). Create a 4-5 page writeup explaining your experiment and findings.

   For this homework I just want a brief description, such as "I plan to read and explain the paper X from conference Y" or "I would like to think about how to theoretically model Z".

2. Consider the class $\mathcal{H}$ of axis-parallel rectangles in R^3. Specifically, a legal target function is specified by three intervals $[x_1^{min}, x_1^{max}]$, $[x_2^{min}, x_2^{max}]$, and $[x_3^{min}, x_3^{max}]$, and classifies an example (x_1, x_2, x_3) as positive if $x_1 \in [x_1^{min}, x_1^{max}]$ and $x_2 \in [x_2^{min}, x_2^{max}]$ and $x_3 \in [x_3^{min}, x_3^{max}]$; otherwise, the example is classified as negative. Argue that $\mathcal{H}[m] = O(m^6)$.

Problems:

3. **[VC-dimension of Two-Layer Networks]** Suppose that hypothesis class $\mathcal{H}$ has VC-dimension d. Now suppose we create a 2-layer network by choosing k functions $h_1, h_2, \ldots, h_k$ from $\mathcal{H}$ and then running their output through some other fixed Boolean function f. That is, given an input x, the network outputs $f(h_1(x), ..., h_k(x))$. For a given f, call the class of all such functions TWO-LAYER$_{f,k}(\mathcal{H})$. Show that TWO-LAYER$_{f,k}(\mathcal{H})$ has VC-dimension $O(kd \log kd)$. Note that we are only asking for an upper bound here, not a lower bound.

Hint: Suppose you have a set S of m data points. By Sauer's lemma, we know there are at most $O(m^d)$ ways of labeling those points using functions in $\mathcal{H}$. Use that to get an upper bound on the number of ways of labeling those points using functions in TWO-LAYER$_{f,k}(\mathcal{H})$. Now select m so that this is less than 2^m which means the VC-dimension must be less than m.

In problems 4-6, you will prove that the VC-dimension of the class $\mathcal{H}_n$ of halfspaces in n dimensions is $n+1$. ($\mathcal{H}_n$ 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 it must have the entire convex hull of S on that side as well.

4. **[lower bound]** Prove that VC-dim$(\mathcal{H}_n) \geq n+1$ 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, explain how one can partition the set in any desired way.)

5. **[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$(\mathcal{H}_n) = n+1$.

6. **[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...