Computational and Statistical Learning Theory

Problem set 4

Due: October 31th

Please send your solutions to `learning-submissions@ttic.edu`

1. For any family of hypothesis classes $\mathcal{H}_n \subseteq \{\pm 1\}^{\mathcal{X}_n}$, where $\mathcal{X}_n = \{0,1\}^n$, define the following decision problem:

   $$\textsc{Agreement}_{\mathcal{H}} = \{(S,k) \mid S \subseteq (\mathcal{X}_n \times \{\pm 1\})^n, k \in \mathbb{Z}, \exists_{h \in \mathcal{H}_n} |\{(x,y) \in S \,|\, h(x) = y\}| \geq k\}$$

   Prove that if H_n is efficiently agnostically properly PAC learnable, and every $h \in H_n$ is computable in time poly(n), then $\textsc{Agreement}_{\mathcal{H}} \in$ **RP**.

2. Prove that for the class $\mathcal{H}_n$ of half-spaces (linear predictors) over $\{0,1\}^n$, the problem $\textsc{Agreement}_{\mathcal{H}}$ is NP-hard.

   Hint: Consider the decision problem $\textsc{HittingSet}$:

   $$\textsc{HittingSet} = \left\{(C,k) \;\middle|\; C \subseteq 2^{[n]}, \exists_{R,|R|=k} \forall_{A \in C} A \cap R \neq \emptyset\right\}$$

   That is, the input is a collection C of subset of the integers $1..n$, and an integer k, and the problem is to decide whether there exists a set of cardinality at most k that "hits" (has non-empty intersection) with all sets in C. The problem $\textsc{HittingSet}$ is a classic NP-hard problem, and you may base your proof on this fact.

   First, show that a restricted version of $\textsc{HittingSet}$ where all sets in C are required to be the same size is also NP-hard (e.g. show a simple reduction from $\textsc{HittingSet}$). Then, consider the following mapping from inputs (C,k), where all sets in C are of cardinality exactly t, to a labeled sample in $\mathbb{R}^{sn}$ (for convenience, we will index vectors in $\mathbb{R}^{sn}$ as $v_{i,j}$ where $1 \leq i \leq s$ and $1 \leq j \leq n$, and denote $e_{i,j}$ the vector of all-zeros except a single one at (i,j)):

   - Positive points at $\sum_{i=1}^{s} e_{i,j}$ for each $j = 1..n$.
   - Negative points at $\sum_{j \in A} e_{i,j}$ for each $i = 1..s$ and each $A \in C$.

   Use the above mapping to construct a reduction from the restricted version of $\textsc{HittingSet}$ to $\textsc{Agreement}_{\mathcal{H}}$.

Challenge Problems :

- For a family of hypothesis classes $\mathcal{H}_n$ with VCdim$(\mathcal{H}_n) \leq$ poly(n), we already know that if we have a polynomial-time algorithm that given a sample, returns a hypothesis from $\mathcal{H}_n$ consistent with the sample (if one exists), then $\mathcal{H}_n$ is efficiently properly PAC learnable. Is solving the decision problem enough? Prove or disprove the following: if VCdim$(\mathcal{H}_n) \leq$ poly(n) and $\textsc{Consistent}_{\mathcal{H}} \in$ **RP**, then $\mathcal{H}_n$ is efficiently PAC learnable.

- Prove that for any polynomial $p(n)$, there exists a family $\mathcal{H}_n$ of hypothesis, such that $\mathcal{H}_n$ is (not necessarily efficiently) PAC learnable with poly$(\log n, 1/\epsilon, \log 1/\delta)$ examples, but that any polynomial-time learning algorithm for $\mathcal{H}_n$ needs at least $p(n)$ examples in order to get error less then 0.1 with probability at least $1/2$. You may rely on any standard cryptographic assumption, or just on the existence of a language which is not commutable.

- Based on the result that it is hard to efficiently learn intersections of n^ϵ halfspaces over $\mathbb{R}^n$, prove that it is hard to efficiently *agnostically* learn halfspaces.