CS7545: Machine Learning Theory **Fall 2021**

Homework 2

Lecturer: Santosh Vempala *Due Date: 25 Oct 2021*

Notes:

- You can discuss and collaborate, but please write your own solutions, and clearly mention everyone you discussed with.
- Start on a new page for each problem.
- Submit on Canvas via Gradescope

1. [PAC Learning]

Give a PAC algorithm for learning the following concept class defined by 3 halfspaces in $\mathbb{R}^d$: A point $x \in \mathbb{R}^d$ is labeled positive if it lies in exactly one of 3 unknown halfspaces (but not the other two), $w_i^\top x \geq 0, i = 1, 2, 3$, or if it lies in all 3 of them; otherwise it is labeled negative. [Hint: try to write the labeling function as a polynomial.]

2. [VC Dimension]

Bound the VC dimension of the following concept classes:

1. Simplices in $\mathbb{R}^d$.
2. Parities of subsets of k variables out of n.

3. [Large Margin Classifiers]

1. For a decision list of length k, give a bound on the margin of the corresponding halfspace, and thereby bound the number of mistakes made by Perceptron and by Winnow in the worst case.
2. Let $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a labeled sample of n points in $\mathbb{R}^n$ with

$$x_i = (\underbrace{(-1)^i, \ldots, (-1)^i, (-1)^{i+1}}_{i \text{ first components}}, 0, \ldots, 0) \text{ and } y_i = (-1)^{i+1}.$$

   Show that the Perceptron algorithm makes $\Omega(2^n)$ updates before finding a separating hyperplane, regardless of the order in which it receives the points.
3. Let $w^\top x \geq 0$ be a halfspace in $\mathbb{R}^n$ with margin $\gamma > 0$ for $\|w\|_2 = 1$ and $\|x\|_2 \leq 1$. Consider the following algorithm for learning such a halfspace: project examples $x \in \mathbb{R}^n$ randomly to dimension k as $y = Rx$ where R is a random $k \times n$ matrix with iid entries from $N(0, 1)$; learn a halfspace $u^\top x \geq 0$ in $\mathbb{R}^k$ using $\widetilde{O}(k/\epsilon)$ examples; ouput $R^\top u$ as the hypothesis in $\mathbb{R}^n$. How small can we make k and guarantee that the algorithm works with probability at least $1 - \delta$?