Machine Learning Theory (CSC 482A/581A)

Problem Set 1 **Due on Friday, October 4th, 7pm**

Instructions:

- You must write up your solutions individually.
- You may have high-level discussions with 1 other student registered in the course. If you discuss problems with another student, include at the top of your submission: their name, V#, and the problems discussed.
- Your must type up your solutions and are encouraged to use LaTeX to do this. For any problems where you only have a partial solution, be clear about any parts of your solution for which you have low confidence.
- Please be sure to submit your solutions via conneX by the due date/time indicated above. This is a hard deadline.

Questions:

1. Let $\mathcal{X} = \mathbb{R}^2$ and take $\mathcal{C}$ to be the class of concentric circles $\mathcal{C} = \{c_r : r \geq 0\}$, where, for each nonnegative real number $r \geq 0$, we have $c_r(x) = \mathbf{1}\left[\|x\|_2 \leq r\right]$. Prove that $\mathcal{C}$ is PAC learnable. In particular, show a PAC learning algorithm which, given a training sample of size $n \geq \frac{\log \frac{1}{\delta}}{\varepsilon}$, finds with probability at least $1 - \delta$ a hypothesis $\hat{f} \in \mathcal{C}$ for which $R(\hat{f}) \leq \varepsilon$.

2. Devise an efficient mistake bound learner for the concept class k-term DNF over $\mathcal{X} = \{0, 1\}^d$. The runtime and mistake bound of your algorithm both should be polynomial in d; you may treat k as a constant.

3. Let $\mathcal{X} = \{0, 1\}^d$ and consider PAC learning a finite concept class $\mathcal{C}$. Assume that the inputs are drawn i.i.d. from an unknown distribution P over $\mathcal{X}$, and the labels are generated via the rule $Y = c(X)$ for some $c \in \mathcal{C}$.

   Let's call this problem the "clean" problem; so, in the clean problem, the training sample consists of random examples of the form (X, Y) for $X \sim P$ and $Y = c(X)$.

   Next, consider the following "corrupted" problem: Each time we request a random example (X, Y), with probability $\alpha(X) \in [0, 1]$ the value of the label Y is flipped. Call the resulting label $\widetilde{Y}$. Thus,

$$\widetilde{Y} = \begin{cases} -Y & \text{with probability } \alpha(X) \\ Y & \text{with probability } 1 - \alpha(X) \end{cases}$$

   In the corrupted problem, the examples are of the form $(X, \widetilde{Y})$, and so the labels are noisy.

(a) Using c and α, derive an expression for the Bayes classifier for the corrupted problem.

(b) For the remaining questions, assume that $\alpha(x) = \frac{1}{4}$ for all $x \in \mathcal{X}$. What is the Bayes classifier for the corrupted problem?

(c) What is the Bayes risk for the corrupted problem?

(d) Let $c_\varepsilon \in \mathcal{C}$ be a hypothesis for which $\Pr(c_\varepsilon(X) \neq c(X)) = \varepsilon > 0$. What is the risk (expected zero-one loss) of c_ε for the corrupted problem?

(e) Design an algorithm for PAC learning $\mathcal{C}$ given access only to corrupted labeled examples $(X_1, \widetilde{Y}_1), \ldots, (X_n, \widetilde{Y}_n)$. That is, your algorithm should, with probability at least $1 - \delta$, output a concept $\hat{f} \in \mathcal{C}$ for which $\mathsf{E}_{X \sim P}[\hat{f}(X) \neq c(X)] \leq \varepsilon$. Your algorithm should be statistically efficient (you should mention the sample size n required, and n should be polynomial in $\frac{1}{\varepsilon}$ and $\frac{1}{\delta}$), but it need not be computationally efficient. Please explain why your algorithm is correct.