Problem Set 2

Instructions:

- The topics for this problem set are:
  Unit 2 – PAC Learning.
  Unit 3 – Concentration of Measure.
  For this problem set, you may only cite results that appeared in Units 2 and 3. Do not cite any results from other units.
- Before you start, make sure you are familiar with the course's Homework Policy.

1. (a) Consider the class of intervals over the real line,

$$\mathcal{I} = \left\{ f : \mathbb{R} \to \{0,1\} \;\middle|\; \exists a, b \in \mathbb{R} : \; a \le b \;\wedge\; f(x) = \mathbb{1}(x \in [a,b]) \right\}.$$

Present an algorithm that PAC learns $\mathcal{I}$ with sample complexity $O\left(\frac{\log(1/\delta)}{\varepsilon}\right)$, and prove that it does so correctly. You may assume that the unknown distribution $\mathcal{D}$ over the domain $\mathbb{R}$ has a continuous CDF. [5 pts]

Hint: You may use a technique similar to what we saw in the proof of Lemma 13 in Unit 4. However, do not cite any results from Unit 4.

(b) Let $d \in \mathbb{N}$ and consider the class of unions of d intervals over the real line,

$$\mathcal{I}_d = \left\{ f : \mathbb{R} \to \{0,1\} \;\middle|\; \exists a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{R} : \begin{array}{l} (\forall i \in [d] : \; a_i \le b_i) \;\wedge \\ f(x) = \mathbb{1}\left(x \in \cup_{i \in [d]} [a_i, b_i]\right) \end{array} \right\}.$$

Present an algorithm that PAC learns $\mathcal{I}_d$ with sample complexity $O\left(\frac{d \log(d/\delta)}{\varepsilon}\right)$, and prove that it does so correctly. Do not assume that the CDF is continuous as in part (a). [10 pts]

2. Let $\mathcal{H}$ be the class of circles in $\mathbb{R}^2$ centered at the origin,

$$\mathcal{H} = \left\{ f : \mathbb{R}^2 \to \{0,1\} \;\middle|\; \exists r \in \mathbb{R} \; \forall z \in \mathbb{R}^2 : \; f(z) = \mathbb{1}\left(\|z\|_2 \le r\right) \right\}.$$

Prove that $\mathcal{H}$ is PAC learnable with sample complexity $m(\varepsilon, \delta) \le \left\lceil \frac{\ln(1/\delta)}{\varepsilon} \right\rceil$. [10 pts]

Hint: consider a ring $R = \{z \in \mathbb{R}^2 : \; r < \|z\|_2 < r*\}$ such that $\mathcal{D}(R) = \varepsilon$, where r^* is the radius of the target function. Bound the probability that the sample S satisfies $S \cap R = \varnothing$. Also, keep in mind that such a ring might not exists.

3. Fix $d \in \mathbb{N}$ and consider boolean vectors of the form $x = (x_1, \ldots, x_d) \in \{0,1\}^d$. A *literal of* x is a boolean value $x_i \in \{0,1\}$ or its negation $\overline{x_i} = 1 - x_i$ where $i \in [d]$. A *conjunction of literals of* x is a function $f : \{0,1\}^d \to \{0,1\}$ such that $f(x) = \wedge_{i=1}^k \ell_i$, where $0 \le k \le d$ and for all $i \in [k]$, ℓ_i is a literal for x. Let $\mathcal{C}$ be the class of conjunctions,

$$\mathcal{C} = \left\{ f : \{0,1\}^d \to \{0,1\} \;\middle|\; f(x) \text{ is a conjunction of literals of } x \right\}$$

Present an algorithm that PAC learns $\mathcal{C}$ with sample complexity $m = O\left(\frac{d+\log(1/\delta)}{\varepsilon}\right)$ and time complexity $O(md)$. [15 pts]

Hint: Start with the set of all possible literals, and proceed by elimination, using only the examples with label 1.

4. Consider an alternative definition of PAC learning, in which there are two unknown distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ over the domain $\mathcal{X}$. The loss of a hypothesis $h : \mathcal{X} \to \{0, 1\}$ is given by

$$L_{\mathcal{D}_0,\mathcal{D}_1}(h) = \max\left\{\mathbb{P}_{x\sim\mathcal{D}_0}[h(x) \neq 0], \mathbb{P}_{x\sim\mathcal{D}_1}[h(x) \neq 1]\right\}.$$

▶ **Definition 1.** *We say that a class $\mathcal{H}$ of functions $\mathcal{X} \to \{0,1\}$ is two-distribution PAC learnable if there exists a function $m : (0,1)^2 \to \mathbb{N}$ and an algorithm A such that for every $\varepsilon, \delta \in (0,1)$, and every distributions $\mathcal{D}_0, \mathcal{D}_1$ over $\mathcal{X}$, if there exists $f \in \mathcal{H}$ such that $L_{\mathcal{D}_0,\mathcal{D}_1}(f) = 0$ then*

$$\mathbb{P}_{S_0\sim(\mathcal{D}_0)^m, S_1\sim(\mathcal{D}_1)^m}\left[L_{\mathcal{D}_0,\mathcal{D}_1}(A(S_0, S_1, \varepsilon, \delta)) > \varepsilon\right] < \delta,$$

where $m = m(\varepsilon, \delta)$.

Let $\mathcal{X}$ be a set and let $\mathcal{H}$ be a class of functions $\mathcal{X} \to \{0, 1\}$.

(a) Prove that if $\mathcal{H}$ is PAC learnable (according to the standard single-distribution definition) then it is two-distribution PAC learnable. [10 pts]

(b) Let h_0 and h_1 be constant functions that always output 0 or 1 respectively. Assume that $h_0, h_1 \in \mathcal{H}$. Prove that if $\mathcal{H}$ is two-distribution PAC learnable then it is PAC learnable. [15 pts]

5. **(a)** Let X be a random variable with distribution $\mathcal{N}(0, 1)$. Prove that for any $t > 0$,

$$\mathbb{P}[X \geq t] \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{t^2}{2}}}{t}. \qquad [5\text{ pts}]$$

Hint: compute an integral using the PDF of the Gaussian distribution, using the fact that $X \geq t$.

(b) Let X be a random variable with distribution $\mathcal{N}(0, \sigma^2)$. Prove that X is sub-Gaussian with variance factor σ^2. [5 pts]

Hint: complete the square – note that $-\frac{x^2}{2} + \lambda x = -\frac{(x-\lambda)^2}{2} + \frac{\lambda^2}{2}$ and then use the fact that

$$\int_{-\infty}^{\infty} e^{-z^2/2} \mathrm{d}z = \sqrt{2\pi}.$$

(c) Let X be a sub-Gaussian random variable with variance factor v. Show that $\mathrm{Var}[X] \leq v$. [5 pts]

Hint: Use Taylor's theorem.

6. Consider the following theorem.

▶ **Definition 2.** *Let $p \in [0, 1]$. A random variable X is Bernoulli with parameter p if $\mathbb{P}[X = 1] = p$ and $\mathbb{P}[X = 0] = 1 - p$.*

▶ **Theorem 3.** *Let $X_1, \dots, X_m$ be independent Bernoulli random variables with parameters $p_1, \dots, p_m$ respectively. Let $S = \sum_{i=1}^m X_i$ and $p = \frac{1}{m}\sum_{i=1}^m p_i$. Then*

$$\forall \varepsilon \in [0,1]: \; \mathbb{P}\left[S - mp \geq mp\varepsilon\right] \leq \exp\left(-\frac{mp\varepsilon^2}{3}\right).$$

(a) Prove that $\psi_S(\lambda) \leq mp(e^\lambda - 1)$. [10 pts]

Hint: Use the inequality $\forall x \in \mathbb{R}: \; 1 + x \leq e^x$.

(b) Prove Theorem 3. You may use the following fact without proof:

$$\forall t \in [0,1]: \; t - (1+t)\ln(1+t) \leq -\frac{t^2}{3}.$$ [10 pts]