1.1 Learning points

For $n \in \mathbb{N}$, let $\mathcal{X} = \{0,1\}^n$ and define $\mathcal{C}$ to be the collection of all *point* functions on $\mathcal{X}$, i.e., $\{c_x\}_{x \in \mathcal{X}}$ where

$$c_x(y) = \begin{cases} 1 & y = x \\ 0 & \text{otherwise} \end{cases}$$

Prove that the point functions are PAC-learnable. Give an efficient learning algorithm and the best sample complexity $m = m(\varepsilon, \delta, n)$ that you can.

1.2 Learning parities

For $n \in \mathbb{N}$, let $\mathcal{X} = \{0,1\}^n$ and define $\mathcal{C}$ to be the collection of all *parity* functions on $\mathcal{X}$. Examples of parity functions include $f(x) = x_1 \oplus \bar{x}_2 \oplus x_4$ and $g(x) = x_2 \oplus x_3 \oplus \bar{x}_5$, where $\oplus$ is addition modulo 2. (Thus, $f(11111) = 1 + 0 + 1 \pmod 2 = 0$ and $g(00000) = 0 + 0 + 1 \pmod 2 = 1$.)

Prove that the parity functions are PAC-learnable. Give an efficient learning algorithm and the best sample complexity $m = m(\varepsilon, \delta, n)$ that you can.

1.3 Analysis of notebook classifier

Let $\mathcal{X} = \{1, 2, \ldots, N\}$ for some large N and P be the uniform distribution on $\mathcal{X}$. The teacher picks some function $f : \mathcal{X} \to \{0,1\}$, draws m points $X_i \in \mathcal{X}$ independently according to P, and provides the labeled sequence

$$S = (X_1, Y_1), \ldots, (X_m, Y_m)$$

where $Y_i = f(X_i)$. The notebook classifier constructs a hypothesis $h_m : \mathcal{X} \to \{0,1\}$ as follows:

$$h_m(x) = \begin{cases} y, & \text{if the example } (x,y) \text{ appears in } S \\ 1, & \text{otherwise} \end{cases},$$

where we use the subscript m to emphasize the dependence of the hypothesis on the sample size.

(a) Define the random variable U_m to be fraction of $\mathcal{X}$ that has **not** been observed after m draws:

$$U_m = P(\{x \in \mathcal{X} : x \neq X_i, 1 \leq i \leq m\}) = \frac{|\{x \in \mathcal{X} : x \neq X_i, 1 \leq i \leq m\}|}{N}.$$

Show that

$$\mathbf{E}[U_m] = \left(1 - \frac{1}{N}\right)^m \leq e^{-m/N}$$

and that for $N \geq 100$,

$$\mathbf{E}[U_m] \geq e^{-1.0051m/N}.$$

Use Markov's inequality to show that for any $0 < \delta < 1$, with probability at least $1 - \delta$ we have

$$U_m \leq \delta^{-1} e^{-m/N}.$$

Conclude that for any $\varepsilon, \delta > 0$ and any $f : \mathcal{X} \to \{0, 1\}$, after observing

$$m > N \ln\left(\frac{1}{\varepsilon\delta}\right)$$

uniformly drawn labeled examples, the notebook classifier will achieve

$$\mathrm{err}(h_m) < \varepsilon$$

with probability at least $1 - \delta$. Thus, if we take $m > 2N \ln N$, we can guarantee $\mathrm{err}(h_m) = 0$ (why?) with probability at least $1 - 1/N$.

(b) Note that for $m < N$, we trivially have $U_m \geq 1 - m/N > 0$ with probability 1. We will use Markov's inequality again to get a nontrivial lower bound on U_m. Assuming $N \geq 100$, show that with probability at least $1 - \delta$, we have

$$U_m \geq 1 - \delta^{-1}(1 - e^{-1.0051m/N}).$$

Conclude that for $m \leq (N \ln(25/9))/1.0051 \approx 1.017N$, we have

$$U_m \geq \frac{1}{5}$$

with probability at least $\frac{1}{5}$. Give an explicit function $f : \mathcal{X} \to \{0, 1\}$ such that with probability at least $\frac{1}{5}$, the notebook classifier will have a generalization error of $\frac{1}{5}$ or more, whenever he sees fewer than $1.016N$ examples (for $N \geq 100$).

(c) Now take $\mathcal{X} = \mathbb{N} = \{1, 2, 3, \ldots\}$ and let P be *any* probability distribution on $\mathcal{X}$. Prove that the notebook classifier's generalization error converges to zero *in probability*: for all $\varepsilon, \delta > 0$ there is an $m = m(\varepsilon, \delta)$ such that whenever the sample size is at least m, we have

$$\mathbf{P}\{\mathrm{err}(h_m) > \varepsilon\} < \delta.$$

Prove that the *rate* of convergence may be arbitrarily slow: we can tweak P to make m grow as quickly in $\varepsilon^{-1}, \delta^{-1}$ as we like. [Extra credit: prove the stronger claim that $\mathrm{err}(h_m)$ converges to zero *almost surely*, meaning that

$$\mathbf{P}\Big\{\lim_{m\to\infty} \mathrm{err}(h_m) > 0\Big\} = 0.$$

]

1.4 Attacks on privacy and anonymity

Recall our exploration of the Social Survey Table Generator of the Israeli Central Bureau of Statistics (`http://surveys.cbs.gov.il/Survey/survey.htm`). Suggest ways in which this information can be re-identified and/or exploited.

1.5 Randomized Response

Define the randomized operator (on a single bit)

$$\text{flip}_\alpha(z) = \begin{cases} z & \text{with probability } 1/2+\alpha \\ 1-z & \text{otherwise} \end{cases}$$

Consider the algorithm that on input a database $x = (x_1, \ldots, x_n)$ outputs $s = \sum_{i=1}^{n} y_i$ where $y_i = \text{flip}_\alpha(x_i)$ (fresh independent randomness is used in each invocation of flip).

1. Is the above algorithm is ϵ-differentially private? If *yes* for which values of ϵ?
2. Show how to use s to estimate $\sum_{i=1}^{n} x_i$. Analyze how good an approximation you get.