3.1 Information theoretic attack on statistical databases

Let $x \in \{0,1\}^n$. A query to x is a binary string $q \in \{-1,1\}^n$ and the exact answer to a query q is

$$a(q) = \langle x, q \rangle = \sum_{i=1}^{n} x_i q_i.$$

A computationally unbounded attacker is given access to a mechanism A answering queries noisily, with the guarantee that issued on queries $q_1, \ldots, q_m$, the mechanism A answers at least 51% of them within accuracy E (i.e., $|A(q) - a(q)| \leq E$). All other queries are answered arbitrarily (the attacker is not informed which queries are answered within bound E, and which are not).

Analyze the following algorithm and show that it allows the attacker to recover (with high probability) a database $x' \in \{0,1\}^n$ that agrees with x on all but $O(E^2)$ entries. The algorithm is query efficient (issues $m = \text{poly}(n)$ queries to A), but runs in exponential time:

1. Issue m random queries $q_1, \ldots, q_m$ and get answers $\tilde{a}_i = A(q_i)$.
2. Output $x' \in \{0,1\}$ such that for at least 51% of the queries $|\langle q_i, x' \rangle - \tilde{a}_i| \leq E$.

- Let $x' \in \{0,1\}^n$ and define $z = x - x'$. Note that $z \in \{-1,0,1\}^n$ where the non-zero entries of z correspond to entries where x, x' differ. We will call z *heavy* if it contains more than $(400E)^2$ non-zero entries. If z is heavy and $|\langle q, z \rangle| < 2E$ for 2% of the queries $q_1, \ldots, q_m$, then we call z *bad*.

  Explain why if no *bad* z exist then the attacker algorithm succeeds.

- Prove the following claim: Let $Y = \sum_{i=1}^{k} X_i$ where each X_i is distributed uniformly in $\{-1,1\}$, then for any y,
$$\Pr[Y = y] \leq 1/\sqrt{k}.$$

- Use the claim to show that for a *heavy* z
$$\Pr[|\langle q, z \rangle| < 2E] \leq 0.01.$$

- Use the Chernoff Bound to bound the probability that a *heavy* z becomes *bad*.

- Show that taking $m = O(n)$ suffices for making the algorithm succeed, except for an exponentially small probability.

3.2 Universal learner for computable functions

Let $\mathcal{F}$ be the class of Turing-computable functions $f : \{0,1\}^* \to \{0,1\}$; in other words,

$$f^{-1}(\{1\}) = \{x \in \{0,1\}^* : f(x) = 1\}$$

is a Turing-decidable language for each $f \in \mathcal{F}$. Since $\mathcal{F}$ is countable, there exists a representation (encoding) scheme $\Re : \{0,1\}^* \to \mathcal{F}$. As usual, define the size of $f \in \mathcal{F}$ by

$$|f| = \min \{|z| : \Re(z) = f\}.$$

We are in the usual supervised learning setting: a teacher fixes some distribution P on $\{0,1\}^*$ and a target $f \in \mathcal{F}$; then an iid $\sim P$ sample is drawn and labeled correctly by f. Consider the following "universal learner" for $\mathcal{F}$: for a given labeled sample $(X_i, Y_i)_{i=1,\ldots,m}$, let $h_m \in \mathcal{F}$ be the shortest (in terms of $|\cdot|$) consistent hypothesis.

(a) Show that $\mathcal{F}$ is not a PAC-learnable class (and thus the Universal Learner is not a PAC learner)

(b) However, argue that the Universal Learner is superior to the Notebook Learner. In particular, give an explicit sample bound for the UL of the type

$$m = m(\varepsilon, \delta, |f|)$$

that depends on the target function f via its size but **not** on P. Explain why no such bound is possible for the Notebook Learner.

3.3 Proper and Improper PAC

Let's define an *information-theoretic* PAC that ignores all issues of computational efficiency. Fix an instance space $\mathcal{X}$, a concept class $\mathcal{C} \subseteq \{0,1\}^{\mathcal{X}}$ and a hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$. We say that L is an iPAC learner for $\mathcal{C}$ by $\mathcal{H}$ if for all ε, δ there is an $m = m(\varepsilon, \delta)$ such that for all $f \in \mathcal{C}$

$$L : (X_i, f(X_i))_{i=1,\ldots,m} \mapsto h_m \in \mathcal{H}$$

with

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

The learning is *proper* if $\mathcal{H} = \mathcal{C}$ and *improper* otherwise. Show that in iPAC, proper and improper learning are actually the same notion (that is, $\mathcal{C}$ is iPAC-learnable improperly iff is is iPAC-learnable properly).