# LAST TIME

Noisy example oracle $EX^{\eta}(c, D)$:

- Draw $x \sim D$; Output $(x, c(x))$ with probability $1-\eta$;
  $(x, 1-c(x))$ with probability $\eta$.

PAC + RCN: ( Replace $EX(c,D)$ with $\underline{EX^{\eta}(c,D)}$; allow dependence on $\frac{1}{1-2\eta_0}$, where $0 \le \eta \le \eta_0 < 1/2$ )

---

Statistical Query Oracle: $STAT(c,D)$:
On query $(\chi, \tau)$, where $\chi : X \times \{0,1\} \to \{0,1\}$ and $\underset{\text{tolerance}}{\tau} \in [0,1]$,
outputs $\hat{v}$, s.t. $\left| \hat{v} - \mathbb{E}_{x\sim D}[\chi(x, c(x))] \right| \le \tau$

[ SQ Learning: Access to $STAT(c,D)$ rather than $EX(c,D)$.
(Efficient: $\chi$ polytime evaluatable, $1/\tau$ polynomially bounded)

---

**Theorem**: If $C$ is efficiently SQ learnable, then $C$ is PAC-learnable in the presence of RCN.

**Proof**: Access to $EX^{\eta}(c,D)$ is sufficient to simulate $STAT(c,D)$ with low failure probability.

Let us treat both the target $c : X \to \{-1, 1\}$ and also $\chi : X \times \{-1,1\} \to \{-1,1\}$.

$$\boxed{\begin{array}{ll} 1 \to -1 & x \mapsto (-1)^x \\ 0 \to 1 & x \mapsto 2x - 1. \end{array}}$$

$$\mathbb{E}_{x\sim D}[\chi(x, c(x))] = \mathbb{E}_{x\sim D}[\chi(x,1)\cdot \mathbb{1}(c(x)=1)] + \mathbb{E}_{x\sim D}[\chi(x,-1)\cdot \mathbb{1}(c(x)=-1)]$$

$\mathbb{1}(A) = 1$ if $A$ is true
$= 0$ o/w

$$= \mathbb{E}_{x\sim D}\left[\chi(x,1)\cdot\left(\frac{1+c(x)}{2}\right)\right] + \mathbb{E}_{x\sim D}\left[\chi(x,-1)\cdot\left(\frac{1-c(x)}{2}\right)\right]$$

$$= \underbrace{\frac{1}{2}\mathbb{E}_{x\sim D}[\chi(x,1)] + \frac{1}{2}\mathbb{E}_{x\sim D}[\chi(x,-1)]}_{\text{target-independent}} + \underbrace{\frac{1}{2}\mathbb{E}[\chi(x,1)\,c(x)] - \frac{1}{2}\mathbb{E}[\chi(x,-1)\,c(x)]}_{\text{correlational}}$$

Look at a model which allows two types of queries

- distribution-related (target-independent): $\psi : X \to \{-1,1\}$, outputs: $\hat{v} \in \left(\mathbb{E}_{x\sim D}[\psi(x)] \pm \tau\right)$
- correlational: $\varphi : X \to \{-1,1\}$ outputs $\hat{v} \in \left(\mathbb{E}[\varphi(x)\,c(x)] \pm \tau\right)$

Distribution-related / or target-independent queries can be easily simulated.

$(x_1, y_1), \dots, (x_s, y_s) \sim EX^{\eta}(c,D)$, output $\left(\frac{1}{s}\sum_{i=1}^{s}\psi(x_i)\right)$; $s = O\left(\frac{1}{\tau^2}\log\frac{1}{\delta}\right)$

To simulate correlational queries.

$EX^\eta$ can be thought of as follows. | $c(x) \in \{-1, 1\}$.

$(x, c(x)) \sim EX(c, D)$

$Z \sim \{-1, 1\}$, s.t. $P(Z=-1) = \eta$ | $\mathbb{E}[Z] = 1 - 2\eta$.

Outputs $(x, c(x) Z)$.

$$\underbrace{\mathbb{E}_{(x,y) \sim EX^\eta(c,D)} [\varphi(x) y]}_{\text{can estimate}} = \mathbb{E}_{(x, c(x)) \sim EX(c,D)} \mathbb{E}_{Z} [\varphi(x) \cdot c(x) Z]$$

$$= \mathbb{E}_{x \sim D} \cdot \mathbb{E}_{Z} [\varphi(x) \cdot c(x) \cdot Z] = \underbrace{\mathbb{E}_{x \sim D} [\varphi(x) c(x)]}_{\text{wanted to estimate}} \cdot (1 - 2\eta)$$

- Suppose we know $\hat{\eta}$, s.t. $|\hat{\eta} - \eta| < \Delta$.
- Let $\hat{v}$, be such that $\left| \hat{v} - \mathbb{E}_{(x,y) \sim EX^\eta(c,D)} [\varphi(x) y] \right| \leq \tau'$
- $v^* = \mathbb{E}_{x \sim D} [\varphi(x) c(x)]$

$$\left| \frac{\hat{v}}{1 - 2\hat{\eta}} - v^* \right| \leq \frac{1}{1 - 2\hat{\eta}} \cdot |\hat{v} - (1 - 2\hat{\eta}) v^*|$$

$$\leq \frac{c \cdot \Delta}{1 - 2\eta} \cdot |\hat{v} - \underline{(1 - 2\eta) v^*} + c \cdot \Delta|$$

$$\leq \frac{c \cdot \Delta}{1 - 2\eta} \tau' + \frac{c \Delta^2}{1 - 2\eta} \leq \tau$$ by choosing $\Delta$ and $\tau'$ appropriately.

Estimating $\hat{v}$ can be done using $O\left(\frac{1}{\tau'^2} \log \frac{1}{\delta}\right)$ samples of Chernoff bounds.

---

$\left\{ i\Delta \mid 0 \leq i \leq \left\lceil \frac{\eta_0}{\Delta} \right\rceil \right\}$ ← Try all values of $\hat{\eta}$ in that range.

$h_1, h_2, \ldots, h_{\lceil \frac{\eta_0}{\Delta} \rceil}$ ← We know that at least one has error at most $\varepsilon/100$

$$\mathrm{err}(h) = \mathbb{E}_{x \sim D} (\mathbb{1}(h(x) \neq c(x)) = \frac{1}{2} - \frac{\mathbb{E}[h(x) c(x)]}{2}. \qquad \frac{1 - h(x) \cdot c(x)}{2}$$

$$\mathbb{E}_{(x,y) \sim EX^\eta(c,D)} [h(x) y] = (1 - 2\eta) \cdot \mathbb{E}_{x \sim D} [h(x) c(x)].$$

**Theorem**: If $X = \{-1, 1\}^n$ and if $\mathcal{U}$ is the uniform distribution over $X$, then any SQ learning algorithm which makes queries with $\tau \geq \tau_{min}$, must make at least $\Omega(\tau_{min}^2 2^n)$ queries to $STAT(c, D)$ when learning PARITIES.

$$S \subseteq \{1, \ldots, n\}, \qquad f_S(\underline{x}) = \prod_{i \in S} x_i \doteq \begin{cases} -1 & \text{if odd } \# x_i \text{ in } S \text{ are } -1 \\ +1 & \text{otherwise.} \end{cases}$$

$$\text{PARITIES} = \{ f_S \mid S \subseteq \{1, \ldots, n\} \}, \qquad |\text{PARITIES}| = 2^n.$$

- $$\mathbb{E}_{\underline{x} \sim \mathcal{U}} [f_S \cdot f_T] = \mathbb{E}_{\underline{x} \sim \mathcal{U}} \left[ \prod_{i \in S} x_i \cdot \prod_{i \in T} x_i \right] = \mathbb{E}_{\underline{x} \sim \mathcal{U}} \left[ \prod_{i \in S \Delta T} x_i \right]$$
$$= \begin{cases} 0 & \text{if } S \neq T \\ 1 & \text{if } S = T. \end{cases}$$

---

Functions from $\{-1, 1\}^n \to \mathbb{R}$ and use the inner product $\langle f, g \rangle = \mathbb{E}_{\underline{x} \sim \mathcal{U}} [f(x) g(x)]$

then PARITIES is an orthonormal basis of this vector space

Let $\varphi : \{-1, 1\}^n \to \{-1, 1\}$, $\mathbb{E}_{\underline{x} \sim \mathcal{U}} [\varphi(\underline{x})^2] = 1$

$$\varphi = \sum_S \alpha_S f_S \quad \text{where} \quad \alpha_S = \mathbb{E}_{\underline{x} \sim \mathcal{U}} [\varphi(\underline{x}) f_S(\underline{x})].$$

$$1 = \mathbb{E}_{\underline{x} \sim \mathcal{U}} [\varphi(x)^2] = \sum_S \alpha_S^2. \quad \text{(Parseval's Identity).}$$

There can be at most $\frac{1}{\tau^2}$ $S$, such that $|\alpha_S| \geq \tau$.

Always answer $0$ to any query made by the algorithm.

Doesn't cause a problem until you make $\tau^2 \cdot 2^n$ queries.

$\varphi_1, \varphi_2, \ldots, \varphi_k$

As long as $k/\tau^2 < 2^n - 2$

Focus only on correlational queries.

As long as there exist $S \neq S'$ s.t. 0 was a valid answer for all queries if either of them was the target, then the algorithm can't succeed.

Because $\underline{err(h; f_S)} + \underline{err(h; f_{S'})} \geq err(f_S; f_{S'}) = \frac{1}{2}$.

where $h$ is the hypothesis output by the algorithm.

$$err(f_S; f_{S'}) = \frac{1}{2} - \frac{\mathbb{E}}{2}[f_S(x) f_{S'}(x)] .$$

<u>Remark</u>: Target independent queries can be exactly computed.

$\mathbb{E}_{x \sim \mathcal{U}}[\varphi(x)]$ ← can be computed using brute force if needed.

---

<u>PARITIES with NOISE</u>: Conjectured to be "hard" (in a computational sense).

Without computational considerations picking $m = \Theta\left(\frac{n}{\varepsilon^2} \log \frac{1}{\delta}\right)$ examples from $EX^{\eta}(c, D)$ and finding the parity function that makes fewest errors is sufficient.

Current best algorithm require $\underbrace{2^{n/\log n}}_{\text{slightly better than } 2^n}$ time and examples.