Statistical Query

- STAT oracle: D is the input distribution over distribution X

For a tolerance, $\tau > 0$, $STAT(\tau)$ returns

$$v \in [\mathbb{E}_{x\sim D}[h(x)] - \tau, \mathbb{E}_{x\sim D}[h(x)] + \tau]$$

for query function $h: X \to [-1, 1]$

VSTAT oracle returns a value $v \in [P-\tau, P+\tau]$, $P = \mathbb{E}_{x\sim D}[h(x)]$ $\tau = \max\{\frac{1}{t}, \sqrt{\frac{P(1-P)}{t}}\}$

- Searching Problem over Distribution.

X domain, $\mathcal{D}$: set of distribution over X

for example: ϵ-optimal functions

$\mathcal{F}$: set of "solution" : $\mathcal{Z}: \mathcal{D} \to 2^{\mathcal{F}}$ map from distribution to solutions.

Statistical Dimension

$\frac{D'-D}{D} = \frac{D'}{D} - 1 =: \hat{D}'$

Motivation: $\mathbb{E}_{x\sim D'}[f(x)] - \mathbb{E}_{x\sim D}[f(x)] = \langle \frac{D'-D}{D}, f \rangle_D$

Pairwise Correlation: $\chi_D(D_1, D_2) = \left| \langle \frac{D_1}{D} - 1, \frac{D_2}{D} - 1 \rangle_D \right|$ D is a base measure.

set of distributions

Average Correlation: $\rho(\mathcal{D}', D) = \frac{1}{|\mathcal{D}'|^2} \cdot \sum_{D_1, D_2 \in \mathcal{D}'} \chi_D(D_1, D_2)$

- Statistical Dimension: (γ, β)

$\mathcal{F} \subseteq \{f: X \to \mathbb{R}\}$. for every $f \in \mathcal{F}$ there exists a set of m distributions, $D_1 \cdots D_m$.

- $f \notin \mathcal{Z}(D_i)$ $i = 1, \cdots, m$
- $\langle \frac{D_i}{D} - 1, \frac{D_j}{D} - 1 \rangle_D \leq \begin{cases} \beta & \text{for } i = j \in [m] \\ \gamma & \text{for } i \neq j \in [m] \end{cases}$

dimension tolerance correlation

Thm. $m = SD(r, \beta)$. Then at least $m\left(\frac{\tau^2 - r}{\beta - r}\right)$ calls of τ-STAT is needed.

β always big as $O(1)$

$\exists D_1 \cdots D_m$ s.t. $f \notin Z(D_i)$

$\|\frac{D_i}{D} - 1\|_D^2 \le \beta$ if $i \in [m]$ and $\langle \frac{D_i}{D} - 1, \frac{D_j}{D} - 1\rangle_D \le r$ for $i \ne j \in [m]$

$h_1 \cdots h_q$ are query. A_k: set of distribution D_i such that $|\mathbb{E}_D[h_k(x)] - \mathbb{E}_{D_i}[h_k(x)]| > \tau$.

Claim 1 $\quad \sum_{k \le q} |A_k| \ge m$. $\Rightarrow q \ge m(\tau^2 - r)/(\beta - r)$

Claim 2 $\quad \forall k,\ |A_k| \le \frac{\beta - r}{\tau^2 - r}$

[Proof of Claim 2]: Consider $\langle h_k, \sum_{i \in A_k} \hat{D}_i \cdot \mathrm{sign}\langle h_k, \hat{D}_i\rangle \rangle \quad (\Delta)$

① By Cauchy-Schwarz.

$(\Delta)^2 \le \|h_k\|^2 \cdot \|\sum_{i \in A_k} \hat{D}_i \cdot \mathrm{sign}\langle h_k, \hat{D}_i\rangle\|^2$

$\le 1 + \left(\sum_{i \in A_k} \|\hat{D}_i\|^2 + \sum_{i \ne j} |\langle \hat{D}_i, \hat{D}_j\rangle|\right) \le \beta |A_k| + r(|A_k|^2 - |A_k|)$

② At the same time.

$(\Delta)^2 = \left(\sum_{i \in A_k} \langle h_k, \hat{D}_i\rangle \,\mathrm{sign}\langle h_k, \hat{D}_i\rangle\right)^2 \ge \tau^2 |A_k|^2$

$\Rightarrow |A_k| \le \frac{\beta - r}{\tau^2 - r}$

Example

- parity. $x \in \{0,1\}^n$ and $c \in \{0,1\}^n$. let $\chi_c : \{0,1\}^n \to \{-1,1\}$

$$\chi_c(x) \doteq -(-1)^{c \cdot x}$$

distribution. D_c: uniform over $\{x \mid \chi_c(x) = 1\}$

[Lemma] $\mathbb{E}_{x \sim D_c}[\chi_{c'}(x)] = \begin{cases} 1 & \text{if } c = c' \\ 0 & \text{otherwise} \end{cases}$ $\Rightarrow$ set $r = 0$, $B = 1$

$$\mathbb{E}_{x \sim u}[\chi_c(x)\, \chi_{c'}(x)] = \begin{cases} 1 & \text{if } c = c' \\ 0 & \text{otherwise.} \end{cases}$$

$\Rightarrow$ (0, 1)-Statistical Dimension of MAX-XOR-SAT is $2^n - 1$

Remark, k-parity is the example have Statistical-Computation Gap.

- kernel Method is SQ-optimal exp(d) compute.
- Mean-Field NN is Statistical-optimal

- k-Clique. D is a distribution over $X = \{0,1\}^{\binom{n}{2}}$ (Graph)

$$I_S(G) := \begin{cases} 1 & \text{if } S \text{ include a Clique in } G. \\ 0 & \text{otherwise.} \end{cases}$$

Find $S \subseteq V$ of size k to maximize. $\mathbb{E}_{G \sim D}[I_S(G)]$

(0.1)-Statistical Dimension. is $\binom{n}{k} - 1$

Example. Moment - Maximization.

find unit vector u that maximize. $\mathbb{E}_{x\sim D}[(u\cdot x)^r]$

[lemma] if r is odd. $c\in\{0,1\}^n$

D_c: uniform over. $x\in\{-1,1\}^n$. for $\chi_c(x)=-1$

Then $\mathbb{E}_{x\sim D_c}[(x\cdot u)^r] = r!\prod_{i:c_i=1} u_i$

1) $\mathbb{E}_{x\in\{\pm1\}}[(x\cdot u)^r] = \frac{1}{2}\mathbb{E}_{\chi_c(x)=1}(x\cdot u)^r + \frac{1}{2}\mathbb{E}_{\chi_c(x)=1}(x\cdot u)^r$

2) $-\mathbb{E}_{x\sim\{\pm1\}}[\chi_c(x)(x\cdot u)^r] = \frac{1}{2}\mathbb{E}_{\chi_c(x)=-1}(x\cdot u)^r - \frac{1}{2}\mathbb{E}_{\chi_c(x)=1}(x\cdot u)^r$

3). $\mathbb{E}_{x\sim\{\pm1\}}[\chi_c(x)(x\cdot u)^r] = -r!\prod_{i:c_i=1} u_i$ (by induction)

$\rightarrow$ find exact parity

Thm. $r!/2(r+1)^{r/2}$ - optimal of Moment - Maximization.

(0.1)- Statistical dimension is $\binom{n}{r}-1$

Example. Gaussian - Single - Index Problem

a.k.a. learn a single. Neuron.

Information Exponent and Generative. Exponent $\to k^*$

k^*: $\chi_0^2(\mathbb{P}_w, \mathbb{P}_{w'}) \leq \lambda_k m^{k^*} + \frac{m^{k^*+1}}{1-m}$ ①

- For any $m \lesssim d^{d^{1/4}}$ (dimension). $\exists\, w_1 \cdots w_m$ with correlation.

$\max_{i \neq j} |w_i \cdot w_j| \lesssim \varepsilon = \sqrt{\frac{c \log_d(m)^2}{d}}$ ②

① ② $\Rightarrow$ SQ lower Bound.

$\Rightarrow m \left(\frac{\frac{3}{n} - 2\lambda_{k^*}^2 \varepsilon^{k^*}}{1 - 2\lambda_{k^*}^2 \varepsilon^{k^*}} \right) \geq \frac{1}{2} m \left(\frac{3}{n} - 2\lambda_{k^*}^2 \varepsilon^{k^*} \right)$ STAT query → #query

at same scale.

$\delta = m \cdot \frac{1}{n}$

$n \simeq \frac{1}{\lambda_{k^*}^2} \left(\frac{1}{\log_d^2(m)} \right)^{k^*/2}$, $m = \delta n$

$n \simeq \frac{C_k}{\lambda_{k^*}} \left(\frac{d}{\log_d^2(\delta n)} \right)^{k^*/2}$.

Any poly(d) Alogrithm needs. $n \gtrsim d^{k^*/2}$ queries.

Average Correlation:

$$\rho_D(T) = \frac{1}{|T|^2} \sum_{D_1, D_2 \in T} \langle D_1, D_2 \rangle_D$$

average correlation γ, dimension Δ

$\Rightarrow$ tol $(\sqrt{\gamma})$ or $\Omega(\Delta)$ queries.

<u>Example</u> Gaussian Mixture.

idea: Moment Matching. $\mathbb{E}_{x \sim A}[x^i] = \mathbb{E}_{x \sim N(0,1)}[x^i]$

for all $1 \le i \le m$.

moment of u respect to A

Then. for any u, v. $\langle P_u, P_v \rangle_{N(0, I_d)} < |\langle u, v \rangle|^{m+1} \langle A, A \rangle_{N(0,1)}$