10/23/24

Lecture 14: Statistical Query Model (cont'd)

<u>SQ dimension</u> (recipe for supervised problems)

<u>Def</u>: Class of functions $\mathcal{F} = \{f : \mathbb{R} \to [-1,1]\}$ has <u>SQ dimension</u> $\geq D$ w.r.t q if $\exists$ $f_1, \ldots, f_D \in \mathcal{F}$ s.t. $\forall$ $i \neq j$

$$\left| \mathbb{E}_{x \sim q}[f_i(x) \, f_j(x)] \right| \leq \frac{1}{D}$$

<u>Thm</u>: If $\mathcal{F}$ has SQ dimension D, then CSQ learning requires tolerance τ or $\Omega(D\tau^2)$ queries

<u>Pf</u>: For convenience, define $\langle f, g \rangle \triangleq \mathbb{E}[f(x) g(x)]$.

WLOG let $\mathcal{F} = \{f_1, \ldots, f_D\}$.

Given query $\phi : \mathbb{R}^d \to [-1, 1]$, let

$$A^+ \triangleq \{ f \in \mathcal{F} : \langle f, \phi \rangle \geq \tau \}$$

By Cauchy-Schwartz:

$$\left\langle \phi, \sum_{f \in A^+} f \right\rangle^2 \leq \underbrace{\|\phi\|^2}_{\leq 1} \cdot \Big\| \sum_{f \in A^+} f \Big\|^2$$

$$\leq \sum_{f \in A^+} \|f\|^2 + \frac{|A^+|(|A^+| - 1)}{D}$$

$$\leq \frac{|A^+|^2}{D} + |A^+|$$

But $\langle \phi, \sum_{f \in A^+} f \rangle \geq \tau |A^+|$ by defn., so

$$\tau^2 |A^+|^2 \leq \frac{|A^+|^2}{D} + |A^+|$$

$$\Rightarrow |A^+| \leq \frac{D}{D\tau^2 - 1} \leq O\left(\frac{1}{\tau^2}\right)$$

Similarly, for $A^- \triangleq \{ f \in \mathcal{F} : \langle f, \phi \rangle \leq -\tau \}$, $|A^-| \leq O\left(\frac{1}{\tau^2}\right)$.

So regardless of ϕ, all but $O\left(\frac{1}{\tau^2}\right)$ many f's consistent w/ the answer 0, so need $\Omega(D\tau^2)$ queries. $\square$

SQ dimension bound for 1-hidden-layer MLPs:

[Goel - Gollakota - Jin - Karmalkar - Klivans '20]:

Let $S \subseteq [d]$ be of size $m = \lg_2 k$.

Given $w \in \{\pm 1\}^m$, define $w^{(S)} \in \mathbb{S}^{d-1}$ by

$$w_i^{(S)} = \begin{cases} w_i/\sqrt{m} & \text{if } i \in S \\ 0 & \text{o.w} \end{cases}$$

Define $f_S : \mathbb{R}^d \to \mathbb{R}$ by

$$f_S(x) \triangleq \sum_{w \in \{\pm 1\}^m} \left(\prod_{i=1}^{m} w_i \right) \cdot \mathrm{relu}\left(\langle w^{(S)}, x \rangle \right)$$

(<u>Note</u>: this is one-hidden-layer MLP w/ size $2^m = k$)

<u>Claim:</u> If q is sign-symmetric dist. over $\mathbb{R}$ (i.e. for any x, x and $-x$ are equally likely under q), then SQ dim of $\{f_S\}$ is $\geq \binom{d}{m}$.

<u>Pf</u>: For any distinct S, T of size m,

w.t.s $\langle f_S, f_T \rangle_q = 0$

For any $z \in \{\pm 1\}^d$, note that

$$f_S(x \odot z) = \sum_w \prod_i w_i \cdot \underbrace{\mathrm{relu}(\langle w^{[S]}, x \odot z \rangle)}_{\mathrm{relu}(\langle w^{[S]} \odot z,\, x\rangle)}$$

$$= \sum_{w'} \prod_i w'_i \cdot z_S \cdot \mathrm{relu}(\langle w'^{[S]}, x \rangle)$$

$$= z_S \cdot f_S(x).$$

by sign symmetry

$$\text{so } \langle f_S, f_T \rangle_q \overset{\downarrow}{=} \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_z [f_S(x \odot z) f_T(x \odot z)]$$

$$= \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_z [f_S(x) f_T(x) z_S z_T]$$

$$= \mathop{\mathbb{E}}_x [f_S(x) f_T(x)] \cdot \underbrace{\mathop{\mathbb{E}}_z [z_S \cdot z_T]}_{\substack{= \\ 0 \\ \text{if } S \neq T}} . \quad \square$$

NB: Technically, also have to make sure $\{f_S\}$ are nonzero functions! See paper for this calculation (Hermite analysis).

(Diakonikolas - Kane - Kontonis - Zarifis '20) showed stronger lbd ($d^{-\Omega(k)}$ queries or $2^{-\text{poly}(d)}$ tolerance) using different instance of approximately orthogonal functions: given 2D subspace $U \subseteq \mathbb{R}^d$,

$$f_U(x) = h(\langle u_1, x\rangle, \langle u_2, x\rangle)$$

where $h(a,b) = \sum_{i=1}^{k} (-1)^i \left\langle \left(\cos\left(\frac{2\pi}{k}\right), \sin\left(\frac{2\pi}{k}\right)\right), (a,b)\right\rangle$.

and u_1, u_2 are basis for U.

idea: $\deg \le \frac{k+1}{2}$ Hermite coeffs of f_u are 0, so

$$\langle f_U, f_V\rangle_{N(0,Id)} = \langle f_U^{>\frac{k+1}{2}}, f_V^{>\frac{k+1}{2}}\rangle$$

$$\leq \left(\text{"Correlation}(U,V)\text{"}\right)^{\Omega(k)},$$

So if we take a packing of the space of 2D subspaces, we get the desired lbd on the SQ dimension.

Statistical dimension:

Def: Set of distributions $\mathcal{T}^*$ has <u>statistical dimension</u> $\geq \Delta$ w.r.t. reference distribution D and avg. correlation γ if $\forall$ $\mathcal{T}$ of size $\geq |\mathcal{T}^*|/\Delta$,

$$\rho_D(\mathcal{T}) \triangleq \frac{1}{|\mathcal{T}|^2} \sum_{D_i, D_j \in \mathcal{T}} \langle D_i, D_j \rangle_D \leq \gamma,$$

where $\langle D_i, D_j \rangle_D = \mathbb{E}_{x \sim D}\left[\left(\frac{D_i(x)}{D(x)} - 1\right)\left(\frac{D_j(x)}{D(x)} - 1\right)\right]$

Thm: If $\mathcal{T}^*$ has statistical dimension $\geq \Delta$, then learning dist's in $\mathcal{T}^*$ w/ SQ alg. requires either $\Omega(\Delta)$ queries or $\sqrt{\gamma}$ tolerance.

Pf: Given query $\phi : \mathbb{R}^d \to \mathbb{R}$, define

$$A^+ \triangleq \left\{ D' \in T^\circ \text{ s.t. } \mathop{\mathbb{E}}_{D'}[\phi] \geq \mathop{\mathbb{E}}_{D}[\phi] + \sqrt{\gamma} \right\}$$

$$A^- \triangleq \left\{ D' \in T^\circ \text{ s.t. } \mathop{\mathbb{E}}_{D'}[\phi] \leq \mathop{\mathbb{E}}_{D}[\phi] - \sqrt{\gamma} \right\},$$

i.e. set of dist's in T° s.t. answering ϕ w/ $\mathop{\mathbb{E}}_{D}[\phi]$ is incorrect. W.t.s $|A^+|$, $|A^-|$ small.

$$\mathop{\mathbb{E}}_{x\sim D}\left[\phi(x) \cdot \sum_{D' \in A^+} \left\{ \frac{D'(x)}{D(x)} - 1 \right\}\right]^2 \leq \underbrace{\mathop{\mathbb{E}}_{x}\left[\phi(x)^2\right]}_{\leq 1} \cdot \underbrace{\mathop{\mathbb{E}}_{x}\left[\left(\sum_{D' \in A^+} \left(\frac{D'(x)}{D(x)} - 1\right)\right)^2\right]}_{= \sum_{D', D'' \in A^+} \langle D', D'' \rangle_D}$$

$$= \sum_{D', D'' \in A^+} \langle D', D'' \rangle_D$$

$$\leq |A^+|^2 \cdot \rho_D(A^+)$$

Remains to lower bound

$$\mathop{\mathbb{E}}_{x\sim D}\left[\phi(x) \cdot \sum_{D' \in A^+} \left\{ \frac{D'(x)}{D(x)} - 1 \right\}\right]^2. \qquad (*)$$

Note: $\mathop{\mathbb{E}}_{x\sim D}\left[\phi(x) \cdot \left\{ \frac{D'(x)}{D(x)} - 1 \right\}\right] = \mathop{\mathbb{E}}_{x\sim D'}[\phi(x)] - \mathop{\mathbb{E}}_{x\sim D}[\phi(x)]$

$$\geq \sqrt{\gamma},$$

so $(\ast) \geq |A^+|^2 \cdot \gamma$, so

$$\rho_D(A^+) \geq \gamma,$$

implying $|A^+| / |T^0| \leq 1/\Delta$.

Similarly, $|A^-| / |T^0| \leq 1/\Delta$.

So need $\Omega(\Delta)$ queries b/c each one only rules out a $1/\Delta$ fraction. $\square$

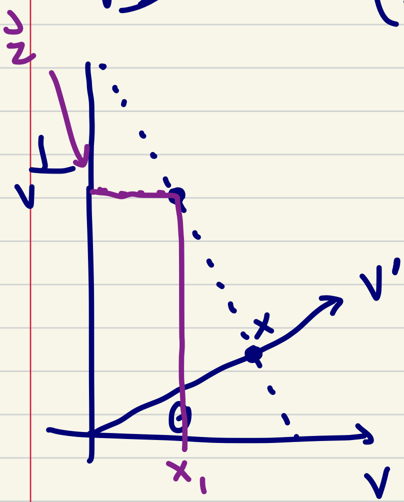
(Bonus material)

Moment-matching and SQ:

Consider P_v and $P_{v'}$ (defined in slides) for moment-matching distribution A. WLOG suppose $v = (1, 0, \ldots, 0)$, $v' = (\cos\theta, \sin\theta, 0, \ldots, 0)$.

Claim: Let Q be dist over $\mathbb{R}^d$ given by taking $X \sim P_v$ and outputting $v' \cdot X$.

Then $\langle Q, Q\rangle_{N(0,1)} \le (\cos\theta)^{2(n+1)} \langle A, A\rangle_{N(0,1)}$



Pf: $$Q(x) = \int_{\vec{x}\cdot(\vec{x},v')=x} P_v(\vec{x})\, d\vec{x}$$

Gaussian

$$= \int_{\substack{(x_1,\vec z):\\ \underbrace{\langle (x_1,\vec z), v'\rangle = x}_{\Updownarrow}\\ x_1\cos\theta + z_1\sin\theta = x}} A(x_1)\cdot G(\vec z)\, dx_1\, d\vec z$$

$\cancel{\frac{x+y}{\sqrt{2}}}$

$x = \frac{x_1+y}{\sqrt{2}}$

scaled convolution of A and Gaussian

$$= \int_{\substack{(x_1,z_1):\\ x_1\cos\theta + z_1\sin\theta = x}} A(x_1)\, G(z_1)\, dx_1\, dz_1$$

$$\int A(x)\gamma(x) \triangleq \underbrace{U_\theta}_{\text{"Ornstein-Uhlenbeck noise operator"}} A(x)$$

(note: to sample from $U_\theta A$, draw $x \sim A$ and output $\cos\theta \cdot x + \sin\theta\, g$ for $g \sim N(0,1)$)

Write $\frac{A(x)}{G(x)} = \sum_{i=0}^{\infty} a_i \underbrace{He_i(x)}_{\text{probabilist's Hermite polynomial}} \cdot \frac{1}{\sqrt{i!}}$

note: $\forall k \leq i \leq m$ $\quad a_i = \mathbb{E}_{x \sim N(0,1)}\left(\frac{A(x)}{G(x)} He_i(x) \cdot \frac{1}{\sqrt{i!}} \right)$

$$= \int A(x)\, He_i(x) \cdot \frac{1}{\sqrt{i!}}\, dx$$

$$(\star\star) \quad \overset{\text{b/c moment matching}}{=} \int G(x)\, He_i(x) \cdot \frac{1}{\sqrt{i!}}\, dx \overset{\text{orthogonality}}{=} 0$$

Mehler's lemma: $U_\theta (He_i \cdot G)(x) = \cos^i(\theta)\, He_i(x)\, G(x)$,

Intuition: noise "smears out" high-degree info

i.e. $\boxed{\text{(Hermite} \times \text{Gaussian density) is } \underline{\text{eigenfunction}} \text{ of noise operator}}$

So $$U_\theta A(x) = \sum_{i=0}^{\infty} a_i \cos^i(\theta)\, He_i(x)\, G(x) \frac{1}{\sqrt{i!}}$$

$$\overset{\text{(by } (\star\star))}{=} \sum_{i=m+1}^{\infty} a_i \cos^i(\theta)\, He_i(x)\, G(x) \frac{1}{\sqrt{i!}}$$

$$\langle Q, Q\rangle_{N(0,1)} = \langle U_\theta A, U_\theta A\rangle_{N(0,1)}$$

$$= \sum_{i=m+1}^{\infty} a_i^2 \cos^{2i}(\theta)$$

$$\leq \cos^{2m+2}(\theta) \cdot \underbrace{\sum_{i=m+1}^{\infty} a_i^2}_{\substack{\shortparallel \\ \langle A, A\rangle_{N(0,1)}}} . \square$$

We conclude that

$$\langle P_v, P_{v'}\rangle_{N(0,Id)} = \langle A, U_\theta A\rangle_{N(0,1)}$$

by the above

$$\leq \langle A, A\rangle_{N(0,1)}^{1/2} \cdot \langle U_\theta A, U_\theta A\rangle_{N(0,1)}^{1/2}$$

$$\leq \underbrace{\cos^{m+1}(\theta)}_{\text{exp. decaying in } m} \langle A, A\rangle_{N(0,1)} . \quad \square$$