EN.601.774 Theory of Replicable ML Spring 2025

Lecture 8

Instructor: *Jess Sorrell* Scribe: *Jess Sorrell*

Last time we defined

$$H(p) = \Pr_{S \sim D_p^m}[\mathcal{A}(S; r^*) = 1]$$

and showed that $H'(p) \in O(\sqrt{m})$ in $\frac{1}{4} \leq p \leq \frac{3}{4}$. Since the derivative is at most $O(\sqrt{m})$ everywhere in $(1/4, 3/4)$, there is an interval I of length $\Omega(1/\sqrt{m})$ around p^* so that $1/3 < \mathrm{H}(p) < 2/3$ for all p in this interval. Since $\mathrm{H}(p) \notin (1/3, 2/3)$ at $p = 1/2 - \tau$ and $p = 1/2 + \tau$, interval I is entirely contained in $(1/2 - \tau, 1/2 + \tau)$. So, there is an $\Omega(1/\tau\sqrt{m})$ chance that a random $p \sim \mathcal{U}([1/2 - \tau, 1/2 + \tau])$ falls in interval I.

Step 3. For $p \in I$, the probability of replicability failure is significant for this r^* then!

$$\begin{aligned}
\Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*)] &= 2 \Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) = 1 \wedge \mathcal{A}(S_2; r^*) = 0] \\
&= 2 \Pr_{S_1 \sim D_p}[\mathcal{A}(S_1; r^*) = 1] \cdot \Pr_{S_2 \sim D_p}[\mathcal{A}(S_2; r^*) = 0] \\
&= 2\mathrm{H}(p)(1 - \mathrm{H}(p)) \\
&> 4/9
\end{aligned}$$

We said early on that when $p \sim \mathcal{U}([1/2 - \tau,\ 1/2 + \tau])$ uniformly, and then $S_1, S_2 \sim D_p$,

$$\Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*)] < 4\rho.$$

Therefore

$$\begin{aligned}
4\rho &> \Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*)] \\
&= \Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*) \mid p \in I] \cdot \Pr[p \in I] \\
&\qquad + \Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*) \mid p \notin I] \cdot \Pr[p \notin I] \\
&\geq \Pr_{S_1, S_2 \sim D_p}[\mathcal{A}(S_1; r^*) \neq \mathcal{A}(S_2; r^*) \mid p \in I] \cdot \Pr[p \in I] \\
&\in \tfrac{4}{9} \cdot \Omega\left(\tfrac{1}{\tau\sqrt{m}}\right) \in \Omega(\tfrac{1}{\tau\sqrt{m}})
\end{aligned}$$

Therefore, $\rho \in \Omega(1/\tau\sqrt{m})$ and $m \in \Omega(\frac{1}{\tau^2\rho^2})$.

Adaptive Statistical Queries

Previously, we proved the following statement about answering statistical queries with empirical estimates:

Claim 0.1. *Let $\phi_1, \cdots, \phi_K$ be arbitrary statistical queries. Then with probability at least $1 - \delta$ over $S \sim_{i.i.d.} D^m$,*

$$\max_{k \in [K]} |\mathbb{E}_S[\phi_k] - \mathbb{E}_D[\phi_k]| \leq \sqrt{\frac{\log(2K/\delta)}{2m}}$$

Note that once we fix a sample $S \sim_{i.i.d.} D^m$, we can no longer meaningfully bound the deviation $|\mathbb{E}_S[\phi_k] - \mathbb{E}_D[\phi_k]|$ for arbitrary statistical queries. For instance, let D be the uniform distribution over $[N]$. Let $S \sim_{i.i.d.} D^m$. Then if we take $\phi_S(x) = \begin{cases} 1, & x \in S \\ 0, & o/w \end{cases}$ we have $\mathbb{E}_S[\phi_S] = 1$ and $\mathbb{E}_D[\phi_S] \leq \frac{m}{N}$. Therefore

$$|\mathbb{E}_S[\phi_S] - \mathbb{E}_D[\phi_S]| \geq 1 - \frac{m}{N}$$

with probability 1.

Even in our SQ model in which the learner does not get direct access to the sample, we can still run into issues with generalization when answering adaptive SQs with empirical estimates. Suppose our data domain is the integers $\mathbb{Z}$ and our distribution is uniform over some large subset of $\mathbb{Z}$. Let $\phi_1(x) = 2^{-x}$. Then $\mathbb{E}_S[\phi(x)] = \frac{1}{m}\sum_{i=1}^{m} 2^{-x}$ and the learner can determine the sample S completely by inspecting the binary representation of $m\,\mathbb{E}_S[\phi]$ (so long as there aren't duplicate elements in S, which we can assume whp as long as D is supported on sufficiently many integers). So if we answer SQs with the empirical average on our sample, we reduce to the case where the learner has access to the sample itself, and can select its next query to overfit badly.