EN.601.774 Theory of Replicable ML	Spring 2025
Lecture 21	
Instructor: *Jess Sorrell*	Scribe: *Jess Sorrell*

Last time we discussed a replicable algorithm for heavy-hitters, relying on randomized thresholding.

1 Replicable Heavy Hitters

Theorem 1.1. *Impagliazzo et al. [2022] The sample complexity is*

$$m_1 + m_2 \in \tilde{\Omega}\left(\frac{1}{(\rho\epsilon(v-\epsilon))^2}\right).$$

Corollary 1.2. *If v and ε are constants, then* $\mathsf{rHeavyHitters}_{\rho,v,\varepsilon}$ *has sample complexity* $\widetilde{O}\left(1/\rho^2\right)$*.*

Learning Heavy-hitters using Statistical Queries. Next, we show that any statistical query algorithm for the v-heavy-hitters problem requires $\Omega(\log|\mathcal{X}|/\log(1/\tau))$ calls to the SQ oracle. Since our heavy hitter algorithm has a sample complexity independent of the domain size, this implies a separation between reproducible problems and problems solvable using only SQ queries.

Consider the ensemble $\{D_x\}_{x\in\mathcal{X}}$ on $\mathcal{X}$, where distribution D_x is supported entirely on a single $x \in \mathcal{X}$.

Claim 1.3 (Learning Heavy-hitters using Statistical Queries)**.** *Any statistical query algorithm for the v-heavy-hitters problem on ensemble $\{D_x\}_{x\in\mathcal{X}}$ requires $\Omega(\log|\mathcal{X}|/\log(1/\tau))$ calls to the SQ oracle.*

Proof. An SQ algorithm for the v-heavy-hitters problem must, for each distribution D_x, output set $\{x\}$ with high probability. An SQ oracle is allowed tolerance τ in its response to statistical query ϕ. So, for any ϕ, there must be some distribution D_x for which the following holds: at least a τ-fraction of the distributions $D_{x'}$ in the ensemble satisfy $|\phi(x')-\phi(x)| \leq \tau$. Thus, in the worst case, any correct SQ algorithm can rule out at most a $(1-\tau)$-fraction of the distributions in the ensemble with one query. If $\mathcal{X}$ is finite, then an SQ algorithm needs at least $\log_{1/\tau}(|\mathcal{X}|)$ queries. ∎

2 Replicable learning of finite hypothesis classes

Definition 2.1 (Probably Approximately Correct (PAC) Learning [Valiant, 1984])**.** Fix a data domain $\mathcal{X}$ and let $\mathcal{Y} = \{0, 1\}$. A model class $\mathcal{H}$ is PAC learnable if there exists an algorithm $\mathcal{L}$ and a function $m_0 : (0,1)^2 \to \mathbb{N}$ such that for all distributions D over $\mathcal{X} \times \mathcal{Y}$, any $\varepsilon, \delta \in (0,1)$, and any $m \geq m_0(\varepsilon, \delta)$, letting $S \sim_{i.i.d.} D^m$ and $h \leftarrow \mathcal{L}(S)$,

$$\Pr_S[\ell_D(h) \geq \min_{h^* \in \mathcal{H}} \ell_D(h^*) + \varepsilon] \leq \delta$$

Algorithm 1 Replicable PAC learner for finite classes $\mathcal{L}(S, OPT)$
Input: Sample $S \sim D$, $OPT = \min_{h' \in H} \mathrm{err}_D(h')$
Parameters: Finite Class H
Replicability, Accuracy, Confidence $\rho, \alpha, \beta > 0$
Sample Complexity $m = m(\rho, \alpha, \beta) \leq O\left(\frac{\log^2 |H| \log \frac{1}{\rho} + \rho^2 \log \frac{1}{\beta}}{\alpha^2 \rho^4}\right)$
Replicability bucket size $\tau \leq O(\frac{\alpha\rho}{\ln |H|})$

---

  for $h \in H$ **do**
    compute $\mathrm{err}_S(h) = \frac{1}{|S|} \sum_{(x,y) \in S} \mathbb{1}[h(x) \neq y]$
  end for
  Select random $v_{\text{init}} \sim Unif[OPT, OPT + \alpha/2]$
  Select random threshold $v \sim Unif(\{v_{\text{init}} + \frac{3}{2}\tau, v_{\text{init}} + \frac{5}{2}\tau, \ldots, v_{\text{init}} + \alpha/4 - \tau/2\})$
  Note that $v \in [OPT + \frac{3\tau}{2}, OPT + \frac{3\alpha}{4} - \frac{\tau}{2}]$
  Randomly order all $h \in H$
  return first hypothesis h in the order s.t. $\mathrm{err}_S(h) \leq v$.

Theorem 2.2 (Intermediate Learnability of Finite Classes)**.** *Let H be any finite concept class. Then Algorithm $\mathcal{L}$ is an agnostic replicable PAC learning algorithm for H with sample complexity:*

$$m(\rho, \alpha, \beta) \leq O\left(\frac{\log^2 |H| \log(\frac{1}{\rho}) + \rho^2 \log \frac{1}{\beta}}{\alpha^2 \rho^4}\right).$$

We note that we'll be able to improve the dependence on ρ to ρ^2 using a replicability amplification procedure. We may or may not get to this, depending on time and interest.

Claim 2.3. *Let $\alpha, \beta, \rho > 0$ be the target accuracy, failure probability, and replicability failure respectively. Let $|H|$ be the size of the hypothesis class. Then letting*

$$m(\rho, \alpha, \beta) \in O\left(\frac{\log^2 |H| \log \frac{1}{\rho} + \rho^2 \log \frac{1}{\beta}}{\alpha^2 \rho^4}\right),$$

algorithm $\mathcal{L}$ is a ρ-replicable PAC learner.

Proof. To show that $\mathcal{L}$ is a PAC learner, we need to show that it returns a hypothesis $h \in H$ such that

$$\mathrm{err}_D(h) \leq OPT + \alpha$$

This follows from uniform convergence arguments. $\mathcal{L}$ returns a hypothesis with empirical error $\mathrm{err}_S(h) \leq v \leq OPT + \frac{3\alpha}{4} - \tau/2$, so as long as we empirically estimate the error of h to within $\frac{\alpha}{4} + \frac{\tau}{2}$, we should be set. Recall that $\tau \in O(\frac{\alpha\rho}{\ln |H|})$, so $\frac{\alpha}{4} + \frac{\tau}{2} \in O(\alpha)$. Our sample complexity is

$$O\left(\frac{\log^2 |H| \log \frac{1}{\rho} + \rho^2 \log \frac{1}{\beta}}{\alpha^2 \rho^4}\right) \in \Omega\left(\frac{\log(1/\beta)}{\alpha^2}\right)$$

and so our algorithm is at least α-accurate, except with probability β, and is therefore a PAC learner. □

References

Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. Reproducibility in learning. In *Proceedings of the 54th annual ACM SIGACT symposium on theory of computing*, pages 818–831, 2022.

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.