Examples drawn from a distribution and labeled by an unknown predicate \( q \in \mathbb{Q} \). How big a study do we need to conduct to answer our questions and preserve privacy?

**QUESTION FOR THIS WORK**

Can we privately release/learn thresholds over infinite domains?

If not, does the sample complexity depend on \( R \)?

**Answer:** Thresholds require \( n > \log^* R \)

### Prior Work & Our Results

- **No Privacy**
  - \( Q \) arbitrary
  - \( Q = \text{POINT}_a \)
  - \( Q = \text{THRESH}_a \)

- **(ε, δ)-diff. priv.**
  - Query release
    - \( O(\log^2 Q \log(R) \log^2 n) \)
  - \( \text{OPT} \)
  - \( \text{POINT} \)
  - \( \text{THRESH} \)

**Our Work:**

- \( \log^2 R < n < 2\log R \)

### Our Techniques

- **Releasing Thresholds**
  - Input: Database \( D \in \{1, 2, \ldots, R\}^n \)
  - Output: Any \( x \) with \( \min D \leq x \leq \max D \)

- **Learning Thresholds**

- **Approximate Medians**

- **Undominated Point Problem**

### Interior Point Lower Bound

- Hard distribution for domain size \( R(x) \)

### Interior Point Upper Bound

- Hard distribution for domain size \( R(x) = 2^{2R(x)} \)

### Differential Privacy

- \( D \) and \( D' \) are neighbors if they differ only on one user’s data

**Pr[\( M(D) \in S \leq \epsilon \) \( \text{Pr}[M(D') \in S] + \delta \)**

Think of \( \epsilon = O(1) \) and \( \delta = o(1/n) \)

### Conclusions

- Releasing/learning thresholds requires sample complexity growing with \( R \)
- Separates private release/learning from non-private cases, even for \( VC(D) = 1 \)
- Open questions: Can we characterize the difference between private/non-private sample complexity? Extend results to improper learning?

### References


### Contact

mbun@seas.harvard.edu

Harvard School of Engineering and Applied Sciences

Maxwell Dworkin 138

33 Oxford St.

Cambridge, MA 02138