Characterizing the Sample Complexity of Private Learners*

Amos Beimel, Kobbi Nissim, Uri Stemmer
CS BGU

*Will be presented in ITCS2013
What is Private Learning?

Kasiviswanathan, Lee, Nissim, Raskhodnikova, Smith 08

- PAC learner (Probably Approximately Correct).
- $\varepsilon$-differentially private.

This work: characterizing the sample complexity of private learners.

- Analogous to VC dim. for non-private learning.
Previous Work

- [KLNRS08] every finite concept class $\mathcal{C}$ can be learned privately using $\ln |\mathcal{C}|$ samples.

- Beimel, Kasiviswanathan, Nissim (2010): sample complexity $\ln |\mathcal{H}|$
  - $\mathcal{H}$ = minimal (deterministic) representation of $\mathcal{C}$.

- [BKN10] $\ln |\mathcal{H}|$ samples are NOT necessary.
Probabilistic Representation

[BKN10] Representation for class $\mathcal{C}$:
A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$, there exists a hypothesis $h \in \mathcal{H}$ that is close to $c$.

Our Contribution - Probabilistic Rep. for $\mathcal{C}$:
A list of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$, w.h.p., a randomly chosen $\mathcal{H}_i$ contains a hypothesis close to $c$. 
The “size” of a list \( \{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\} \) is defined as \( \max \ln |\mathcal{H}_i| \).

**RepDim:** The Representation Dimension of \( \mathcal{C} \) is its minimal probabilistic representation.

**Characterization:** \( \Theta(RepDim(\mathcal{C})) \) samples are necessary and sufficient for privately learning \( \mathcal{C} \).
Representation ⇒ PPAC

**Learning Algorithm:**

**Input:** a labeled sample.

1. Randomly choose a hypothesis class $\mathcal{H}_i$ from the list.
2. Use the exponential mechanism to choose a hypothesis from $\mathcal{H}_i$. 
A PAC learner has to output (w.h.p) a good hypothesis \( h \) for a correctly sampled data.

- Differential privacy guarantees a non-zero probability for outputting \( h \) on an arbitrary data.

- For \( i=1 \) to \( r \): Construct \( \mathcal{H}_i \) by executing the learner several times.
Open Problem

• For class POINT:
  \[ \text{RepDim} = \text{VC} = O(1). \]

• Open:
  A concept class for which \( \text{RepDim} > \text{VC} \).
The End.