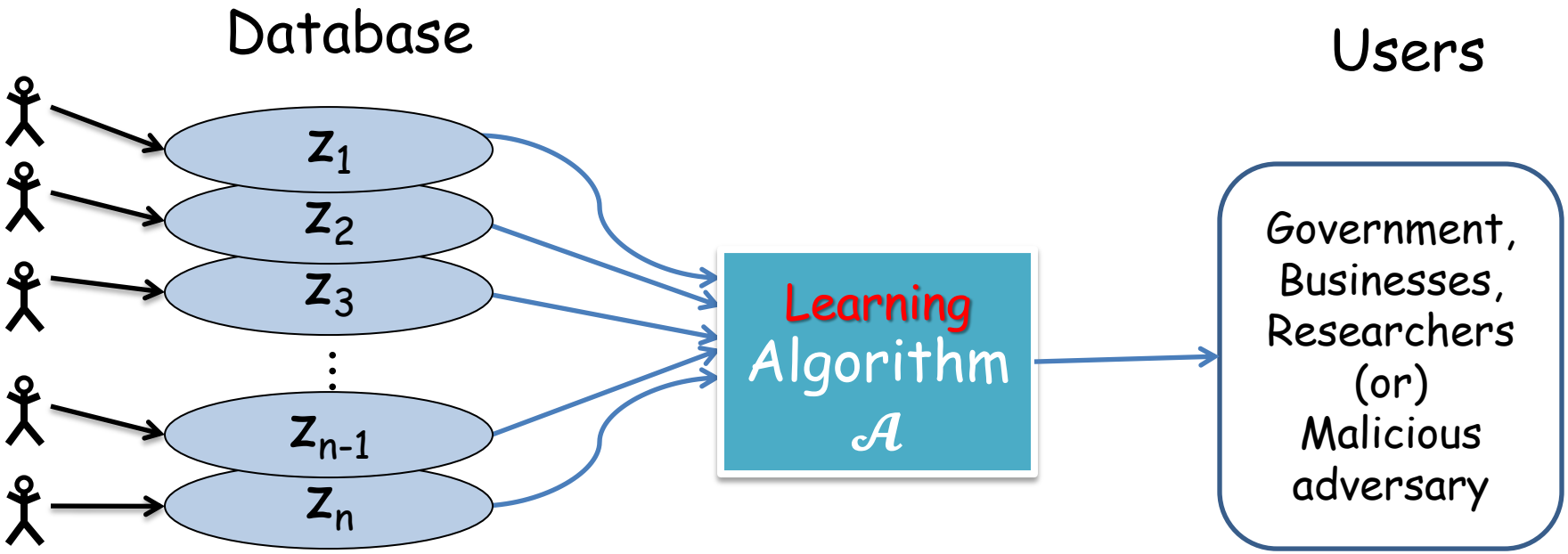
Characterizing the Sample Complexity of Private Learners

ITCS 2013

Uri Stemmer (BGU)

With Amos Beimel and Kobbi Nissim

Why Private Learners?



Often, this algorithmic task can be abstracted as a learning problem:

- Bank is interested in predicting (based on past customers) whether new customers are good/bad credit

Our Results:

- Introducing a new measure of concept classes: **Representation Dimension**
- A combinatorial characterization for the **sample complexity** of private learners in terms of the representation dimension.
- Implications to **sanitization**, private optimization
  - Not in this talk

What is Private Learning?

Kasiviswanathan, Lee, Nissim, Raskhodnikova, Smith 08

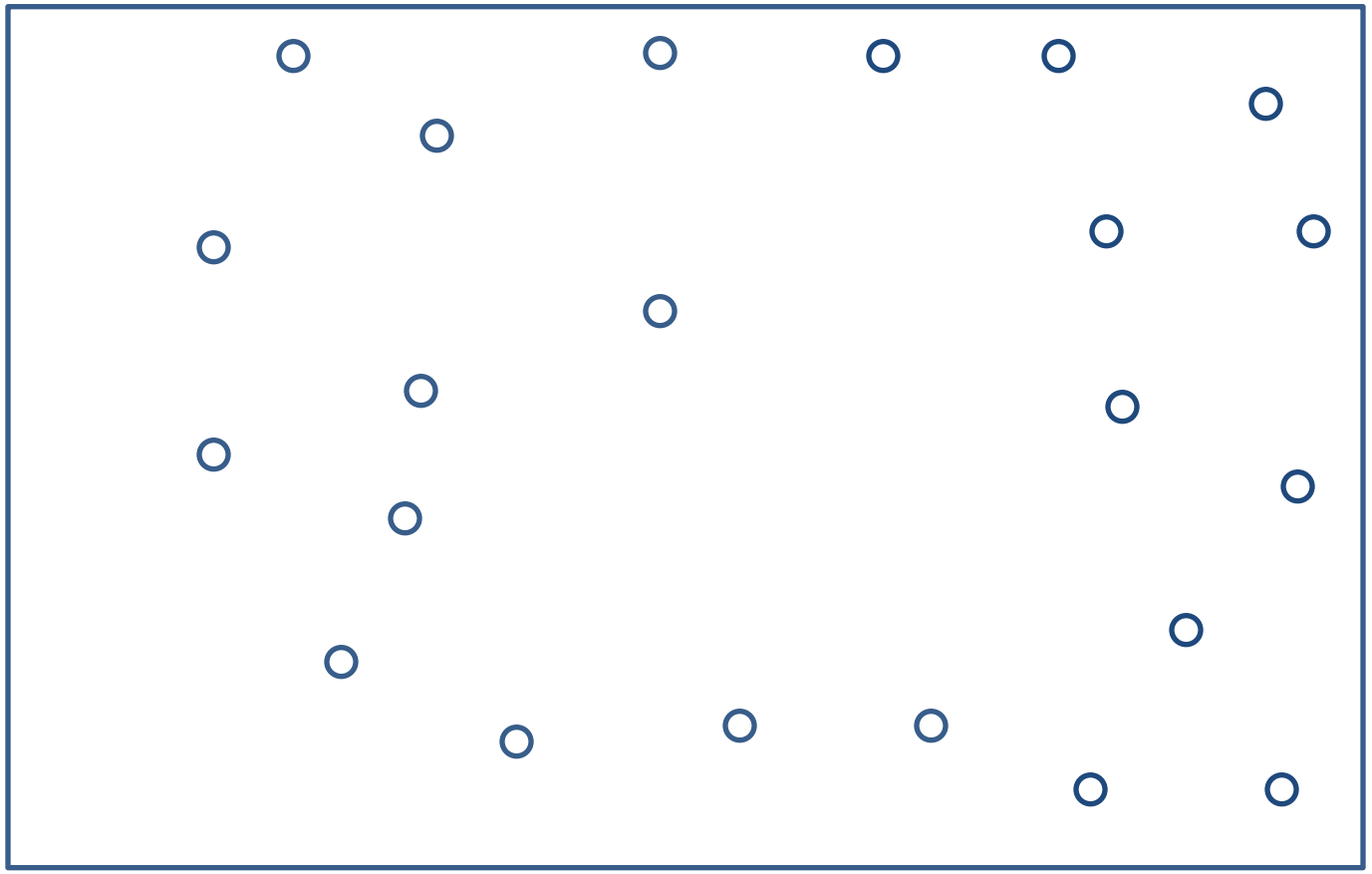
Definition:

PAC Learning
\+
Differential Privacy

Private Learning

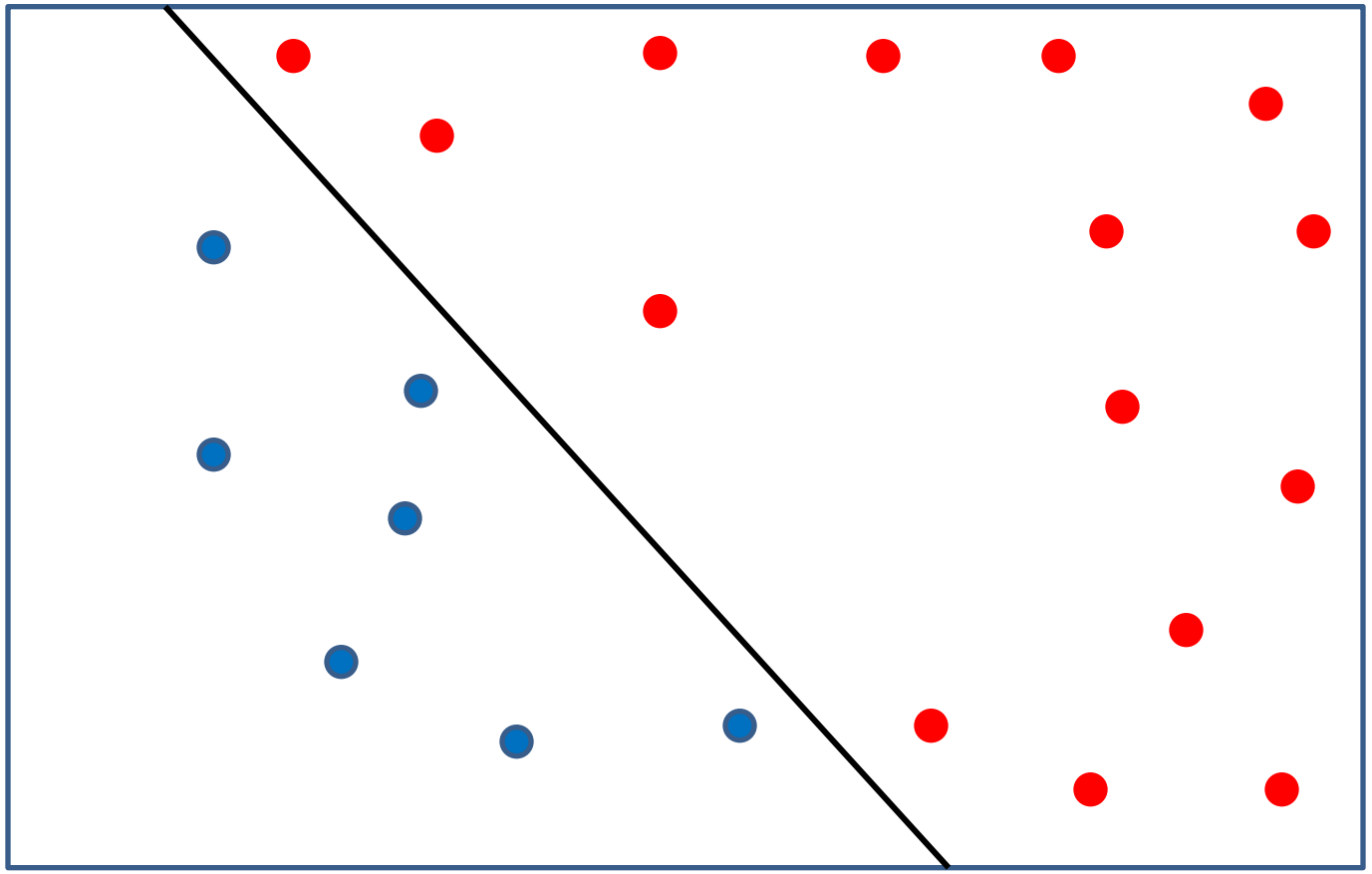
PAC Model [Valiant 84]

- A domain X
- $\mathcal{C}$ = a set of boolean functions over X.



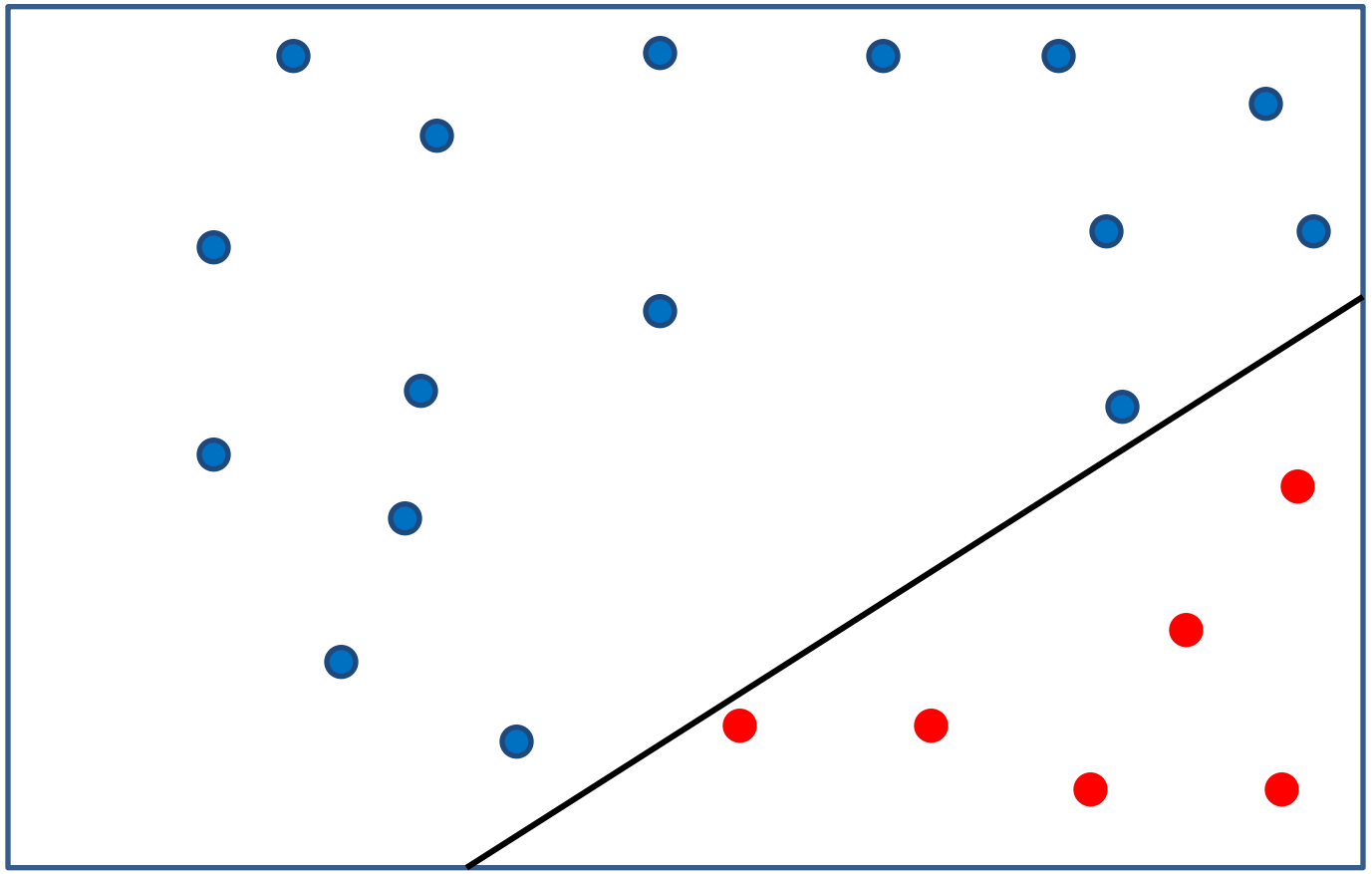
PAC Model [Valiant 84]

- A domain X
- $\mathcal{C}$ = a set of boolean functions over X.



PAC Model [Valiant 84]

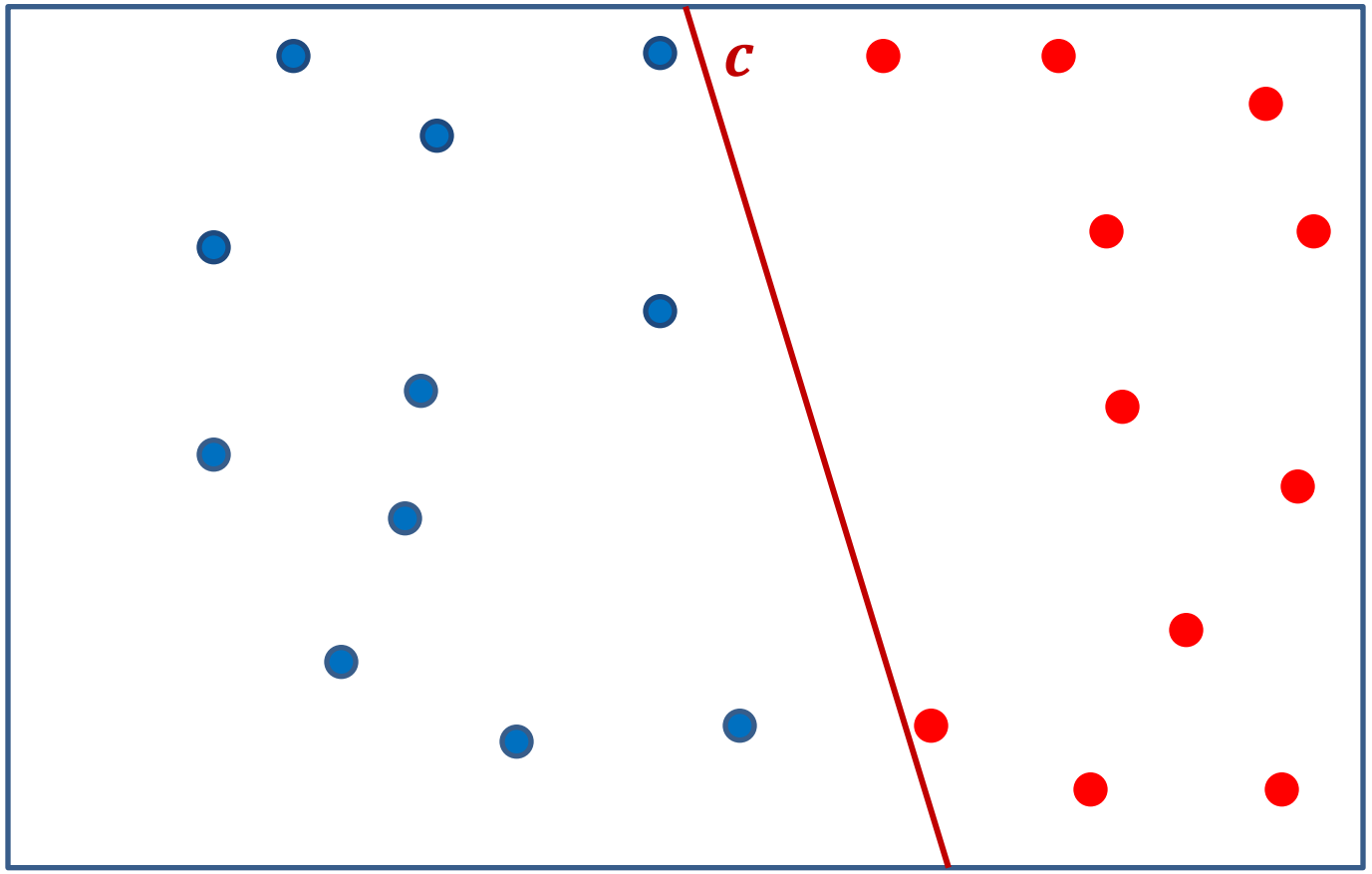
- A domain X
- $\mathcal{C}$ = a set of boolean functions over X.



PAC Model [Valiant 84]

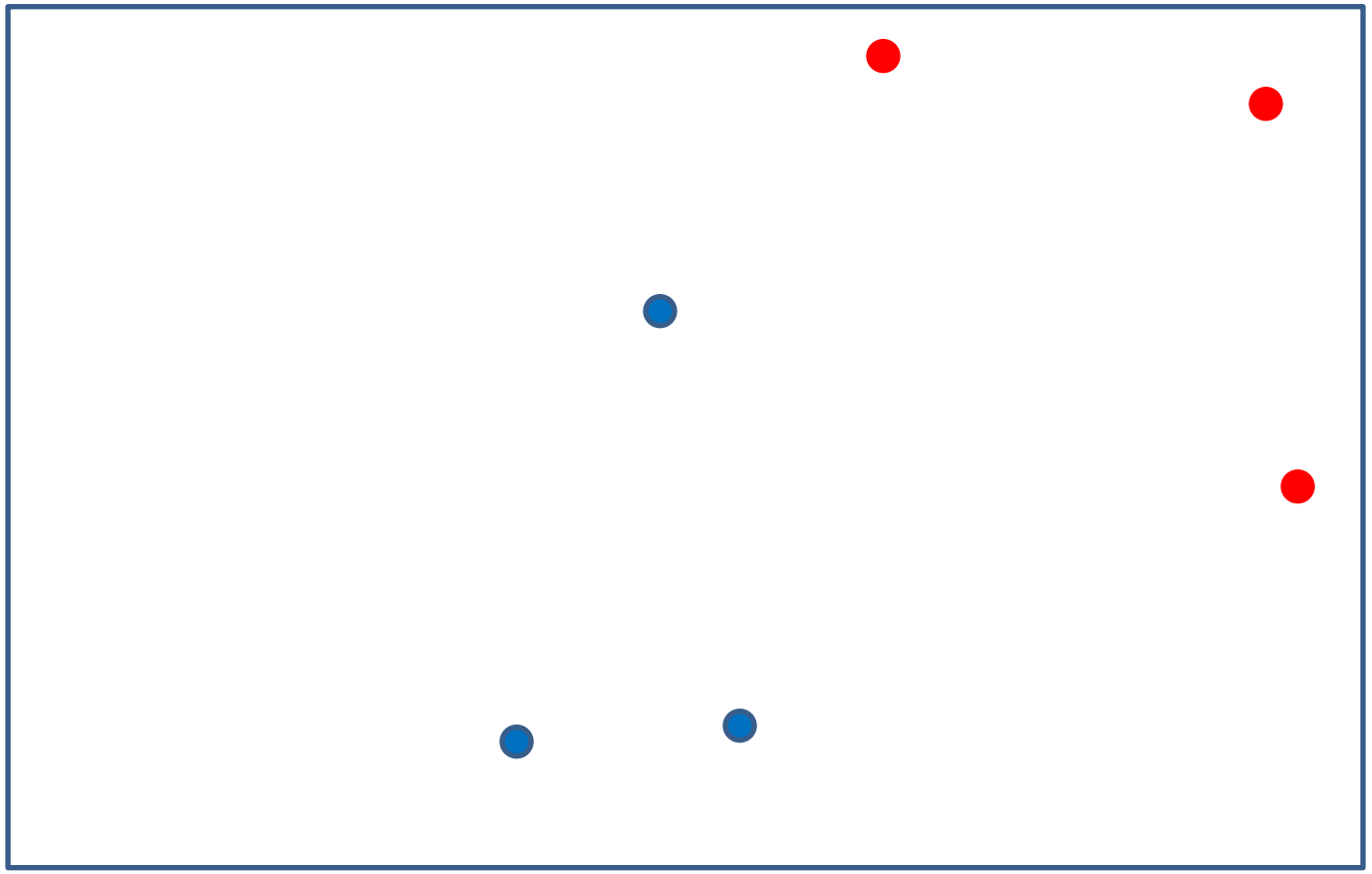
- A domain X
- $\mathcal{C}$ = a set of boolean functions over X.

Unknown: - distribution $\mathcal{D}$ over X.
- A specific target concept $c \in \mathcal{C}$.



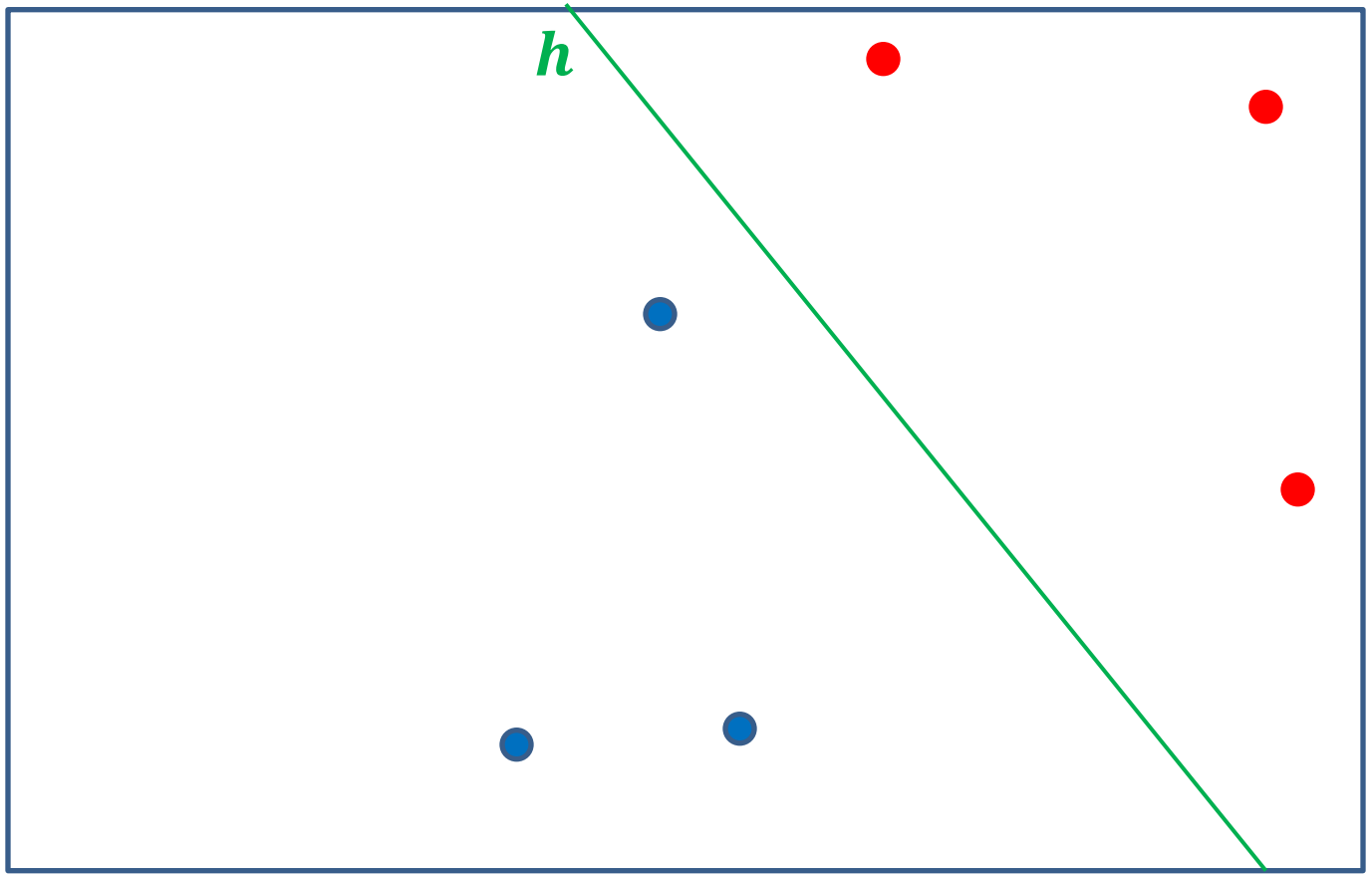
PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c



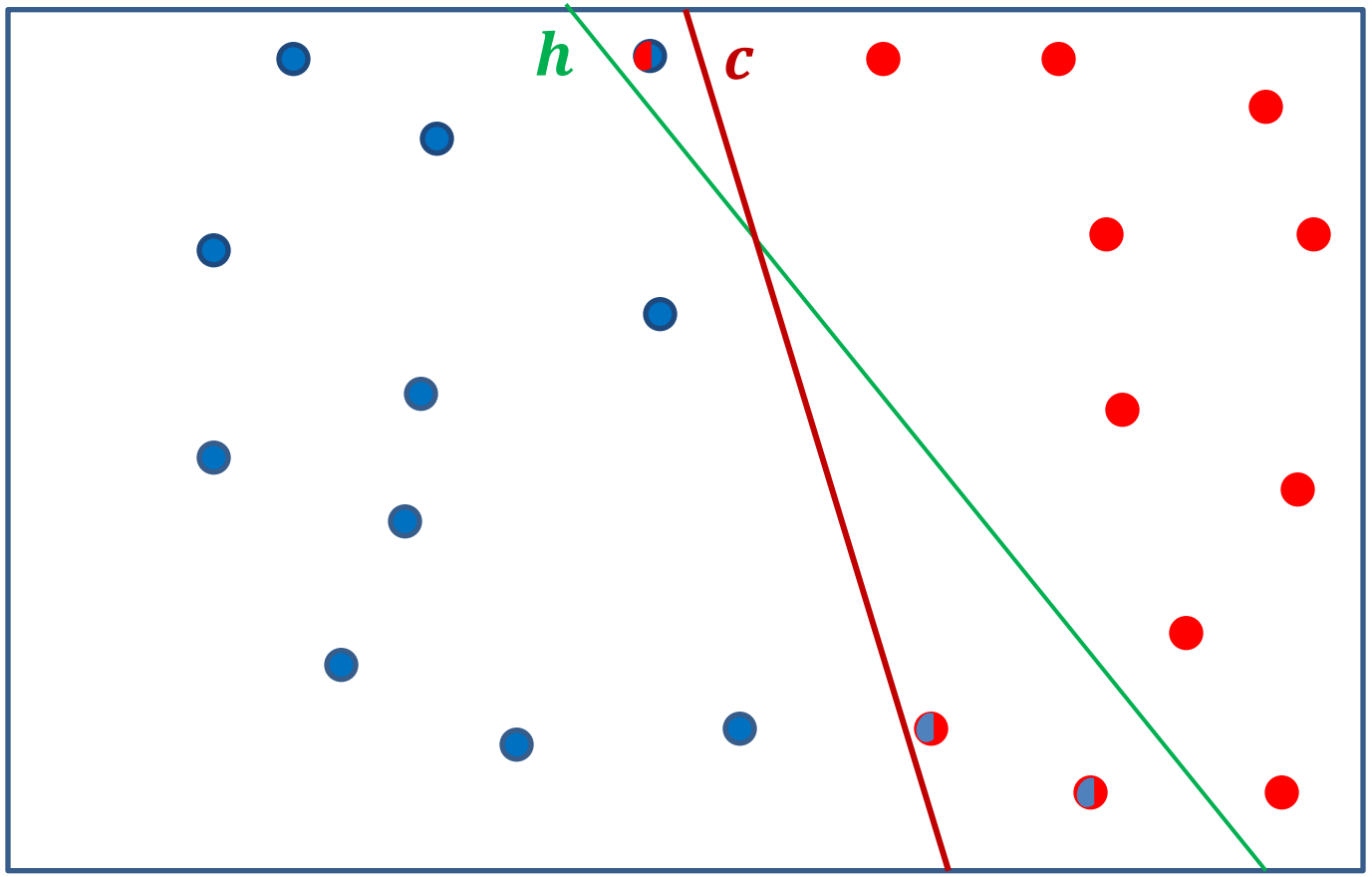
PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c
- A hypothesis h – a "guess" for c



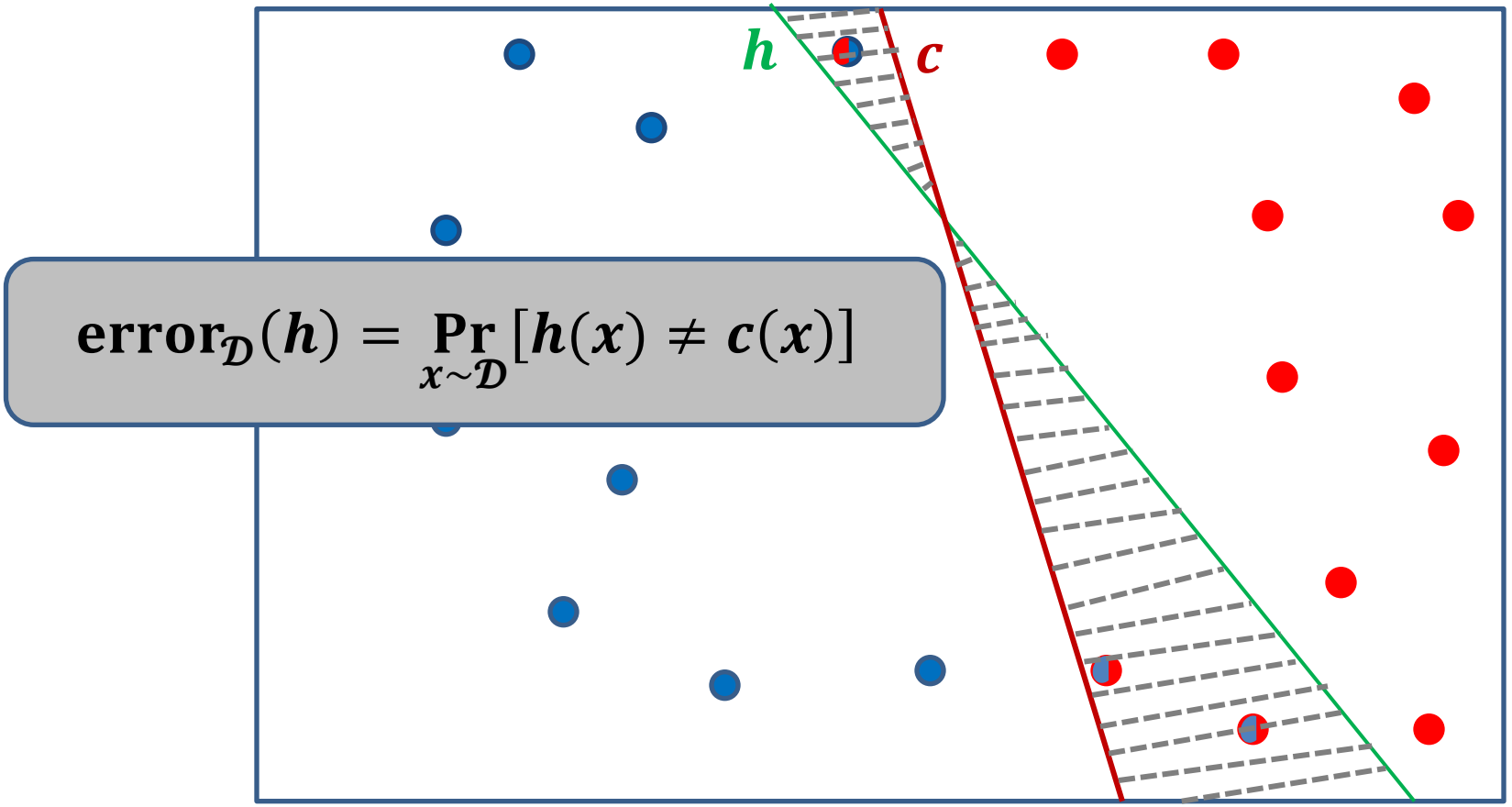
PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c
- A hypothesis h – a "guess" for c



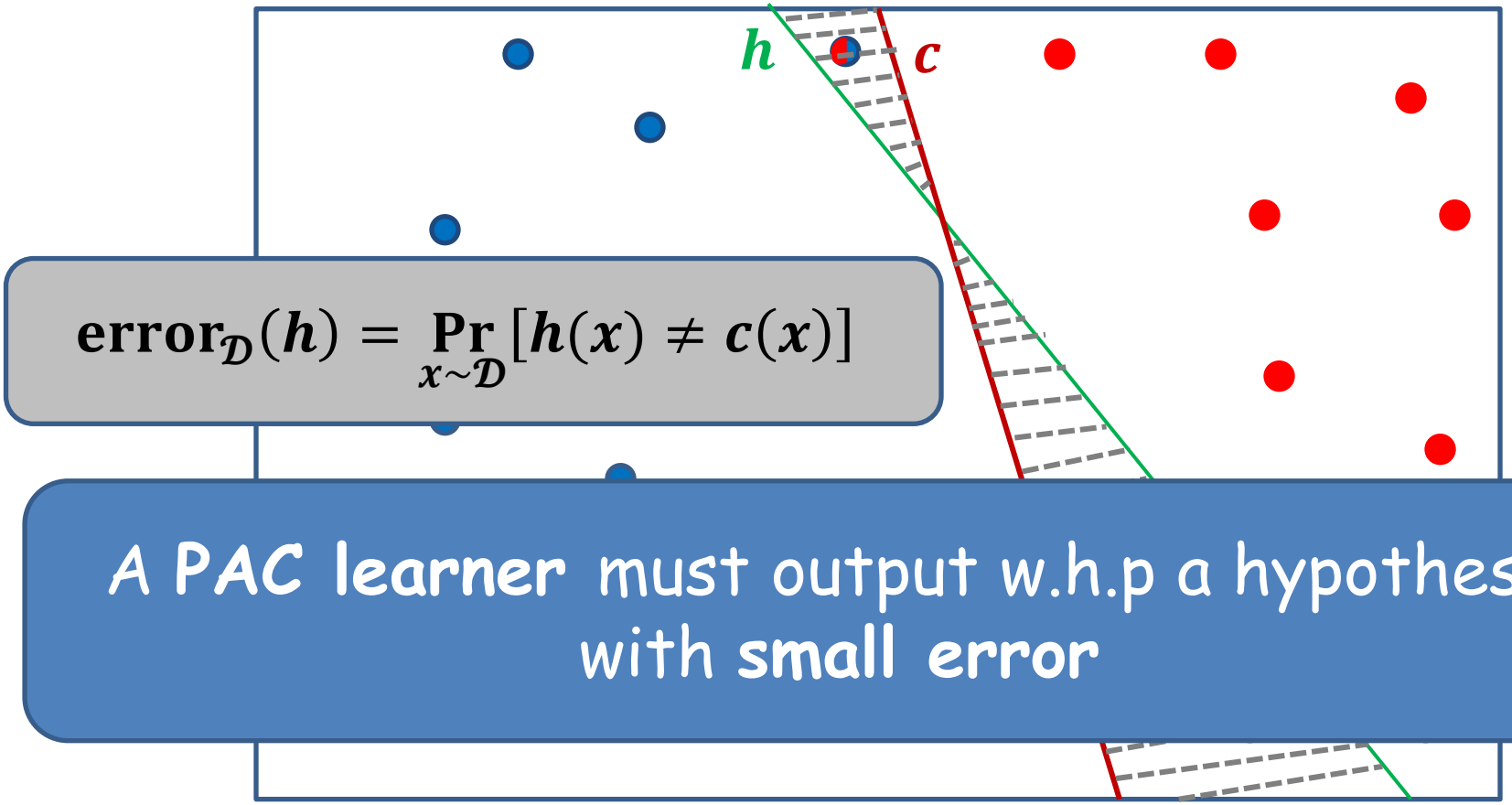
PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c
- A hypothesis h – a "guess" for c



PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c
- A hypothesis h – a "guess" for c

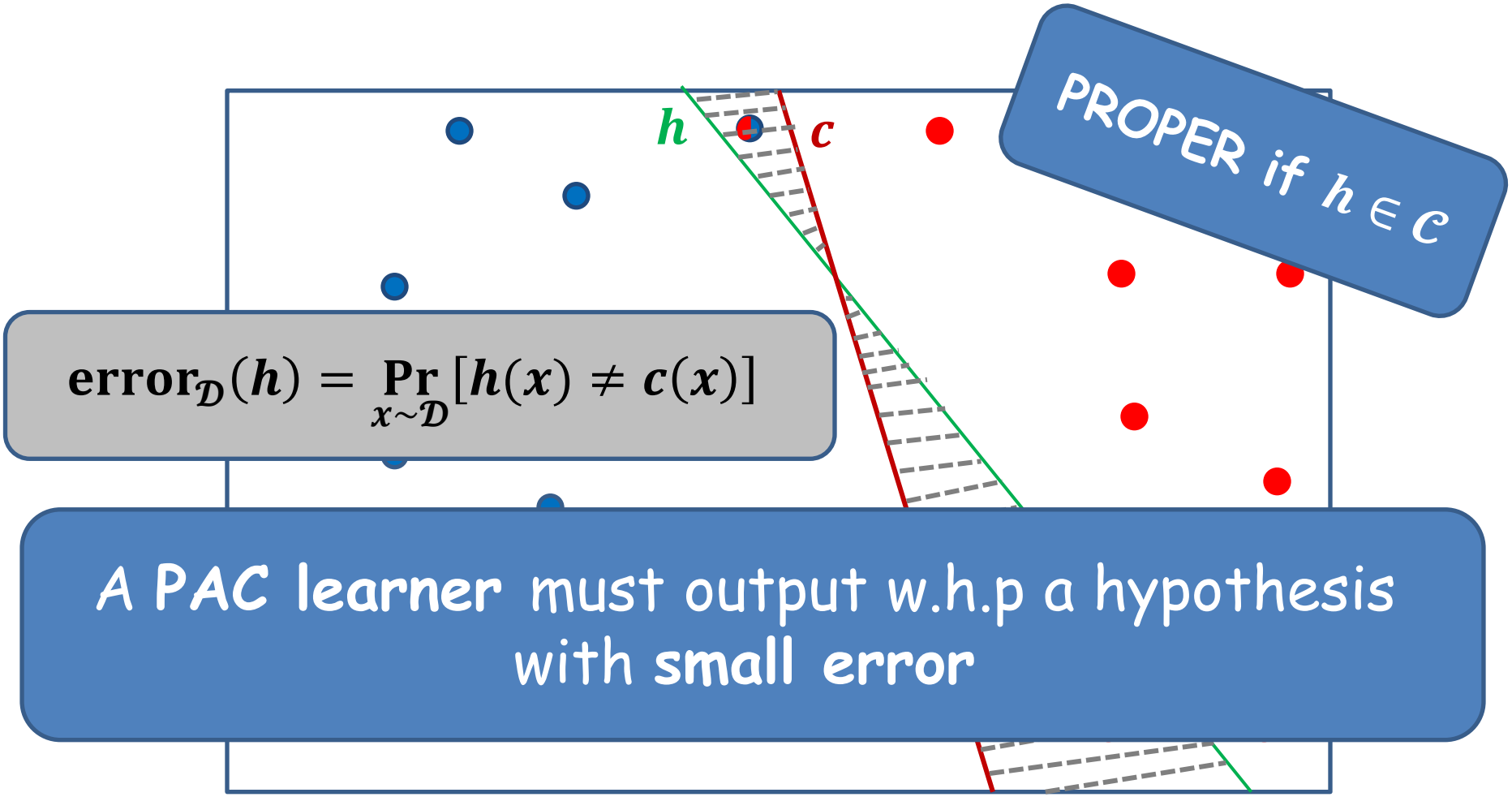


$$\mathbf{error}_{\mathcal{D}}(h) = \Pr_{x \sim \mathcal{D}}[h(x) \neq c(x)]$$

A PAC learner must output w.h.p a hypothesis with small error

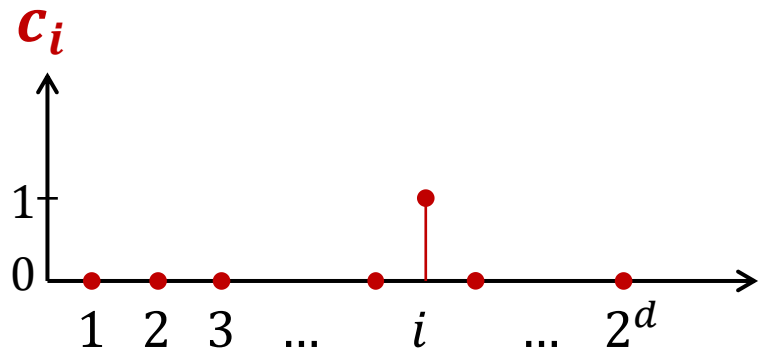
PAC Model [Valiant 84]

- Samples drawn according to $\mathcal{D}$ and labeled by c
- A hypothesis h – a "guess" for c



Example: $\mathbf{POINT_d}$

Concept Class: $\mathcal{C} = \mathbf{POINT_d} = \{c_1, \ldots, c_{2^d}\}$



$$c_i(x) = 1 \iff x = i$$

A (non-private) learner for $\mathbf{POINT_d}$ with $O(1)$ samples:

- If there exists i s.t. $(i, 1)$ in the sample, return c_i.
- Otherwise (all labels are zero), return $h \equiv 0$.

What is Private Learning?

Kasiviswanathan, Lee, Nissim, Raskhodnikova, Smith 08

Definition:

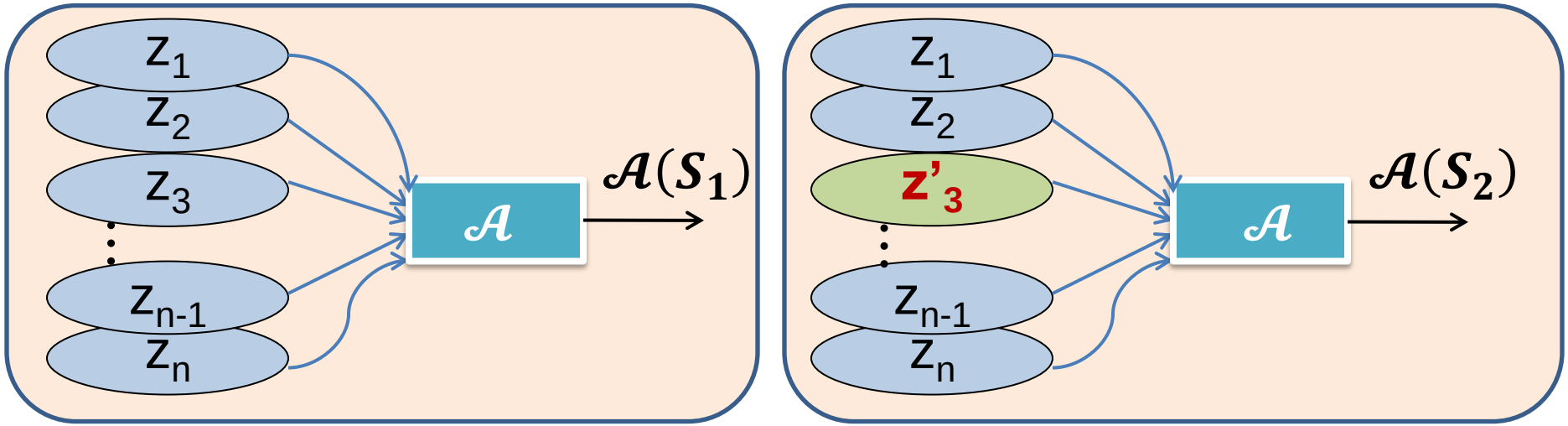
PAC Learning
\+
Differential Privacy

Private Learning

Differential Privacy

Dwork, McSherry, Nissim, Smith 2006

Changing one record does not change the output distribution **"too much"**



Differential Privacy

Dwork, McSherry, Nissim, Smith 2006

Changing one record does not change the output distribution **"too much"**

A (rand) algorithm $\mathcal{A}$ is differentially private if for all neighboring databases S_1, S_2 and for all sets of outputs F:

$$\Pr[\mathcal{A}(S_1) \in F] \approx \Pr[\mathcal{A}(S_2) \in F]$$

Differential Privacy

Dwork, McSherry, Nissim, Smith 2006

Changing one record does not change the output distribution **"too much"**

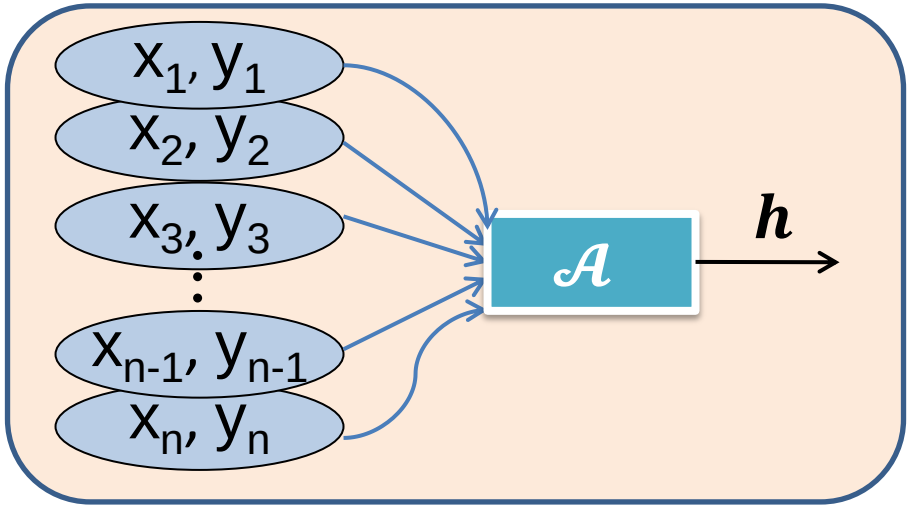
A (rand) algorithm $\mathcal{A}$ is ϵ differentially private if for all neighboring databases S_1, S_2 and for all sets of outputs F:

$$\Pr[\mathcal{A}(S_1) \in F] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(S_2) \in F]$$

Private PAC (PPAC) [KLNRS 08]

Algorithm $\mathcal{A}$ is a PPAC learner for a concept class $\mathcal{C}$ if

- Learning: $\mathcal{A}$ is a PAC learner for $\mathcal{C}$.
- Privacy: $\mathcal{A}$ preserves differential privacy.



- Need to preserve privacy:
  - Label, sample, presence in database may all be sensitive!
- Note: privacy not preserved when we publish h(x) for some x.

The Sample Complexity of Private Learners

Previous Results

Previous related work (partial list):

- [KLNRS08] Define private learning
- [BBKN10] Sample complexity of private learning
- [CH11] Learning in continuous domain, label privacy
- [CM08, CMS11, KST12] Machine learning
- [BLR08, DNRRV09, …] Synthetic Data
- [Shapire90, …] Boosting
- [DRV10] Private Boosting

Previous Work

Kasiviswanathan, Lee, Nissim, Raskhodnikova, Smith 08

Generic Construction (based on the Exp. Mechanism of [MT 07]):

Every finite concept class $\mathcal{C}$ can be learned privately using $\log|\mathcal{C}|$ samples.

Previous Work

Kasiviswanathan, Lee, Nissim, Raskhodnikova, Smith 08

Generic Construction (based on the Exp. Mechanism of [MT 07]):

Every finite concept class $\mathcal{C}$ can be learned privately using $\log|\mathcal{C}|$ samples.

- Recall: Learner for POINT_d with $O(1)$ samples (non-private).
- Generic construction of private learners results in $O(\log|\mathcal{C}|) = O(d)$ samples.

Is this gap essential?

Doom and Gloom?

[KLNRS 08]: A proper private PAC learner of POINT_d using $O(d)$ samples.

[BKN 10]: Proper private PAC learner of POINT_d must use $\Omega(d)$ samples.

[CH 11]: No proper PPAC learner of points in continuous domain.

Doom and Gloom?

[KLNRS 08]: A proper private PAC learner of POINT_d using $O(d)$ samples.

[BKN 10]: Proper private PAC learner of POINT_d must use $\Omega(d)$ samples.

[CH 11]: No proper PPAC learner of points in continuous domain.

Previous Work

Beimel, Brenner, Kasiviswanathan, Nissim 2010

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Previous Work

Beimel, Brenner, Kasiviswanathan, Nissim 2010

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Adversary		DRep $\mathcal{H}$
1. Choose $c \in \mathcal{C}$ and $\mathcal{D}$		
	$c, \mathcal{D}$ →	
		2. $\exists h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$

Previous Work

Beimel, Brenner, Kasiviswanathan, Nissim 2010

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq 1/4$.

Adversary		DRep $\mathcal{H}$
1. Choose $c \in \mathcal{C}$ and $\mathcal{D}$		
	~~$c, \mathcal{D}$~~ labeled sample →	
		2. $\exists h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq 1/4$
		3. Choose $h \in \mathcal{H}$ using the Exp. Mechanism.

Reduces the sample complexity from $\log|\mathcal{C}|$ to $\log|\mathcal{H}|$

[BBKN 10]: $\log|\mathcal{H}|$ samples are NOT necessary.

Our Contribution:

Probabilistic Representation

A list of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, a randomly chosen $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Probabilistic Representation

A list of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, a randomly chosen $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Adversary	Prob. Rep.
1. Choose $c \in \mathcal{C}$ and $\mathcal{D}$	1. Choose $\mathcal{H}_i$ at random
$c, \mathcal{D}$ →	
	2. w.p. ¾, $\exists h_0 \in \mathcal{H}_i$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$

Probabilistic Representation

A **list** of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, **a randomly chosen** $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Adversary	**Prob. Rep.**
1. Choose $c \in \mathcal{C}$ and $\mathcal{D}$	1. Choose $\mathcal{H}_i$ at random
~~$c, \mathcal{D}$~~ labeled sample →	
	2. **w.p. ¾**, $\exists h_0 \in \mathcal{H}_i$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$
	3. Choose $h \in \mathcal{H}_i$ using the Exp. Mechanism.

Reduces the sample complexity to $\log|\mathcal{H}_i|$.

RepDim

Probabilistic Representation:

A **list** of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:

w.p. ¾, **a randomly chosen** $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

RepDim

Probabilistic Representation:

A list of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:

w.p. ¾, a randomly chosen $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

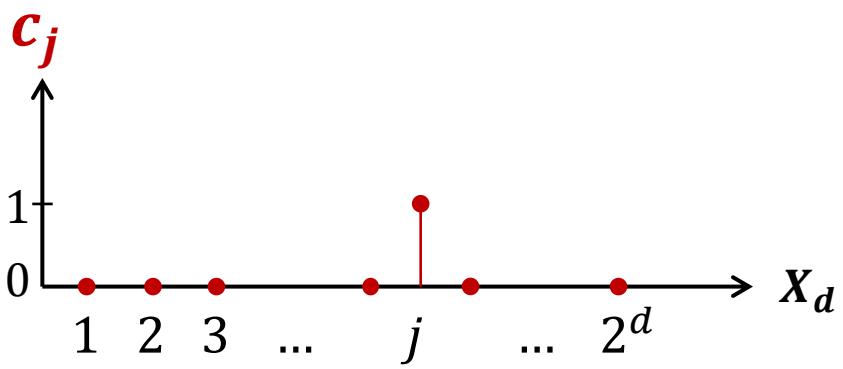
- The **size** of a prob. rep. is defined as $\max \log|\mathcal{H}_i|$.
- Define $\mathbf{RepDim}(\mathcal{C})$ as the size of $\mathcal{C}$'s minimal probabilistic representation.

Example: Prob. Rep. for $\mathbf{POINT_d}$

Probabilistic Representation:

A list of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, a randomly chosen $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

$\mathbf{POINT_d} =$



$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

Size(Rep) = $O(1)$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:
  - $\mathbb{E}[\mathbf{error}_{\mathcal{D}}(h_i) | h_i(j) = 1] \leq \frac{1}{8}$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:
  - $\mathbb{E}[\mathbf{error}_{\mathcal{D}}(h_i) | h_i(j) = 1] \leq \frac{1}{8}$.
  - By Markov's inequality $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) > \frac{1}{4} \middle| h_i(j) = 1\right] \leq \frac{1}{2}$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:
  - $\mathbb{E}[\mathbf{error}_{\mathcal{D}}(h_i) | h_i(j) = 1] \leq \frac{1}{8}$.
  - By Markov's inequality $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) > \frac{1}{4} \middle| h_i(j) = 1\right] \leq \frac{1}{2}$.
  - $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}\right] \geq \mathbf{Pr}[h_i(j) = 1] \, \mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4} \middle| h_i(j) = 1\right] \geq \frac{1}{16}$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$\boldsymbol{Rep} = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:
  - $\mathbb{E}[\mathbf{error}_{\mathcal{D}}(h_i) | h_i(j) = 1] \leq \frac{1}{8}$.
  - By Markov's inequality $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) > \frac{1}{4} \middle| h_i(j) = 1\right] \leq \frac{1}{2}$.
  - $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}\right] \geq \mathbf{Pr}[h_i(j) = 1] \, \mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4} \middle| h_i(j) = 1\right] \geq \frac{1}{16}$.
  - $\mathcal{H}$ fails to contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ w.p. at most $\left(1 - \frac{1}{16}\right)^{22} < \frac{1}{4}$.

Example: Prob. Rep. for $\mathbf{POINT_d}$

$$Rep = \left\{ \mathcal{H} \subseteq 2^{X_d} \ : \ |\mathcal{H}| \leq 22 \right\}.$$

- **Choose $\mathcal{H}$ as follows:**
  - For i = 1 … 22 construct h_i by setting $h_i(x)=1$ with probability 1/8 and $h_i(x)=0$ otherwise.
  - $\mathcal{H} = \{h_i\}$.
- **Claim:** Fix a target concept $c_j \in \mathbf{POINT}_d$ and a distribution $\mathcal{D}$. The probability that $\mathcal{H}$ contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ is at least $\frac{3}{4}$.
- Proof:
  - $\mathbb{E}[\mathbf{error}_{\mathcal{D}}(h_i) | h_i(j) = 1] \leq \frac{1}{8}$.
  - By Markov's inequality $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) > \frac{1}{4} \middle| h_i(j) = 1\right] \leq \frac{1}{2}$.
  - $\mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}\right] \geq \mathbf{Pr}[h_i(j) = 1] \, \mathbf{Pr}\left[\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4} \middle| h_i(j) = 1\right] \geq \frac{1}{16}$.
  - $\mathcal{H}$ fails to contains h_i s.t. $\mathbf{error}_{\mathcal{D}}(h_i) \leq \frac{1}{4}$ w.p. at most $\left(1 - \frac{1}{16}\right)^{22} < \frac{1}{4}$.
- Efficiency: enough if entries of h_i are pairwise ind.

Until Now

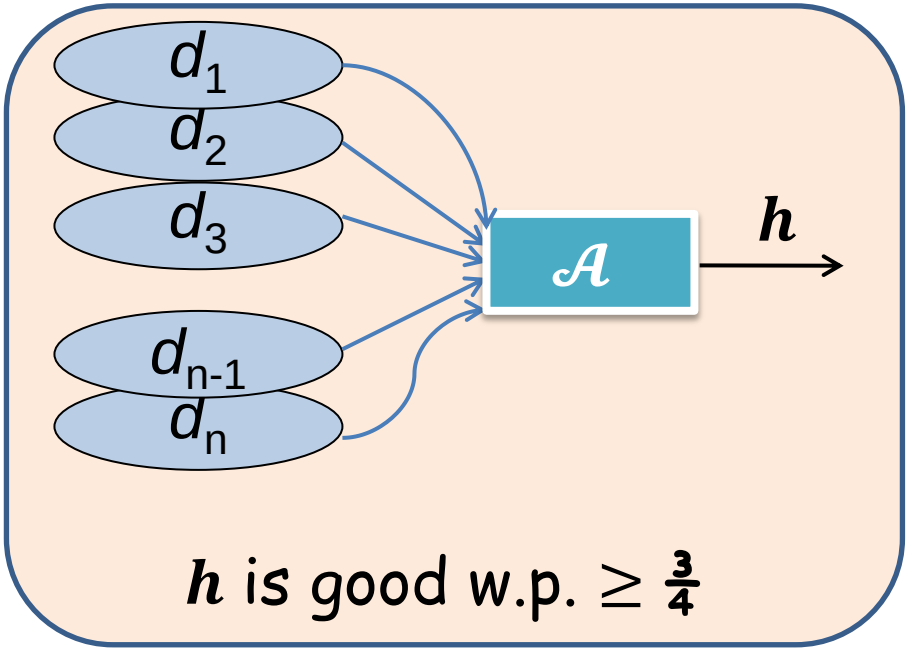
- Defined **Probabilistic Representation**.

- We showed:
  Probabilistic Representation $\Longrightarrow$ Private Learner.

Private Learner $\Longrightarrow$ Prob. Rep.

$\mathcal{A}$: PPAC learner for a concept class C using n examples.

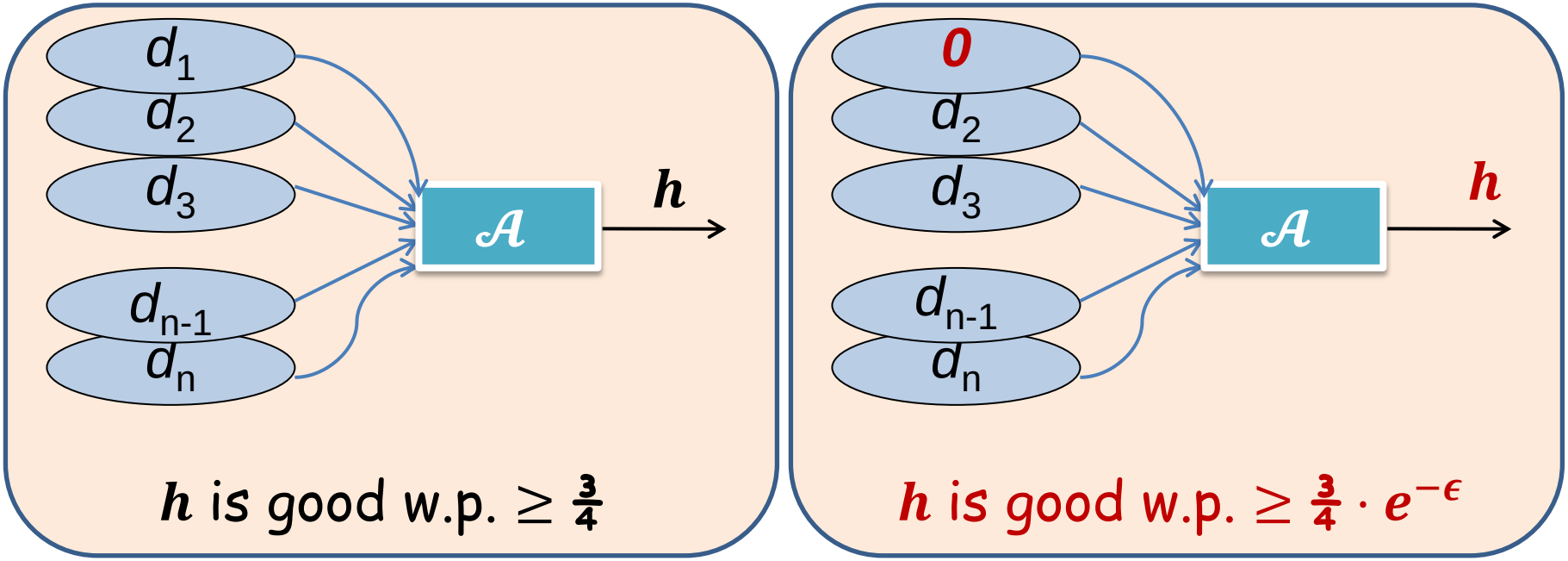
Correctly sampled data:



Private Learner $\Longrightarrow$ Prob. Rep.

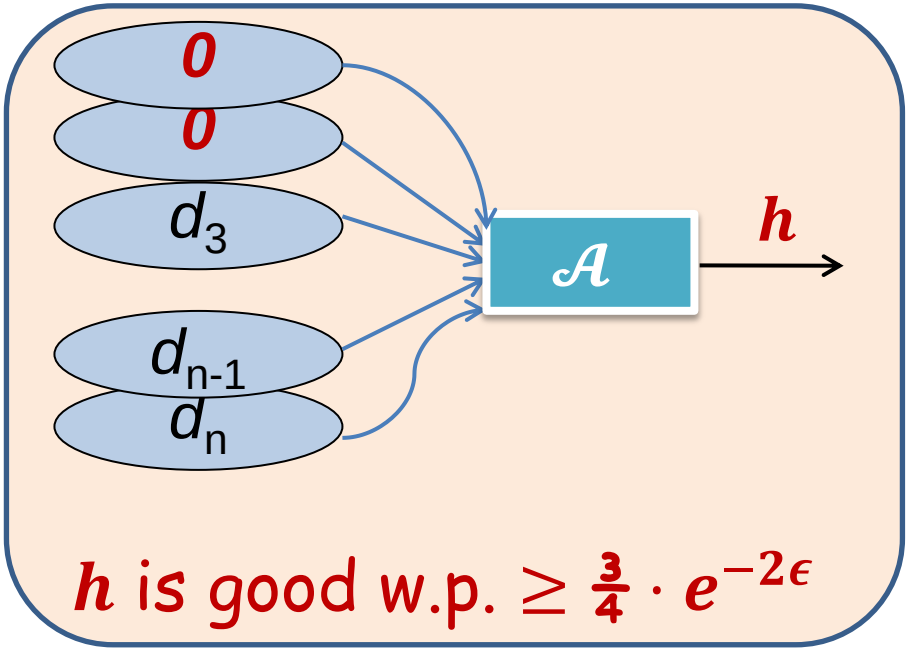
$\mathcal{A}$: PPAC learner for a concept class C using n examples.

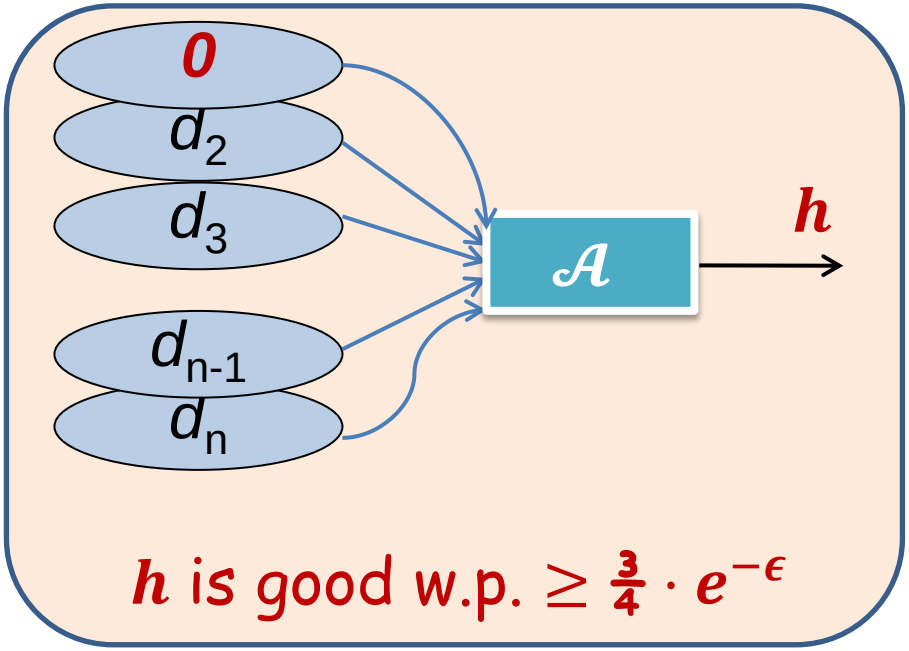
Correctly sampled data:



Private Learner $\Longrightarrow$ Prob. Rep.

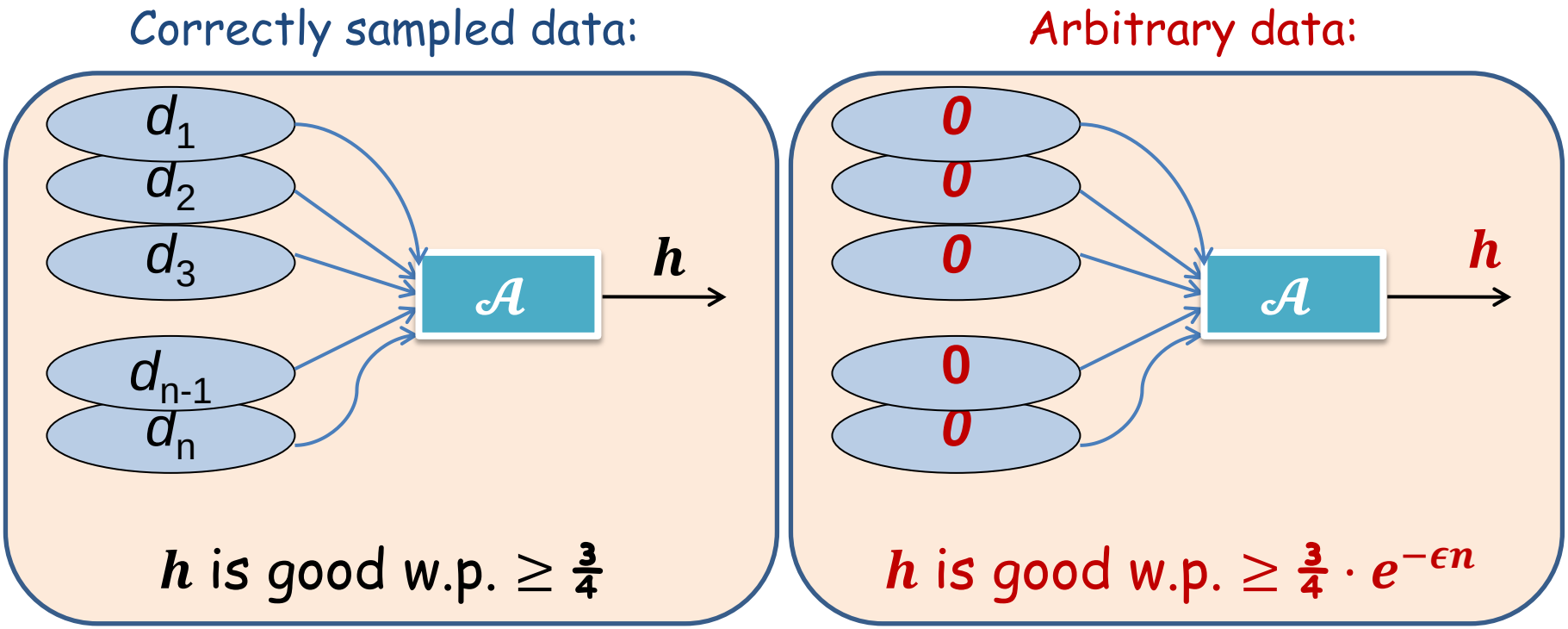
$\mathcal{A}$: PPAC learner for a concept class C using n examples.





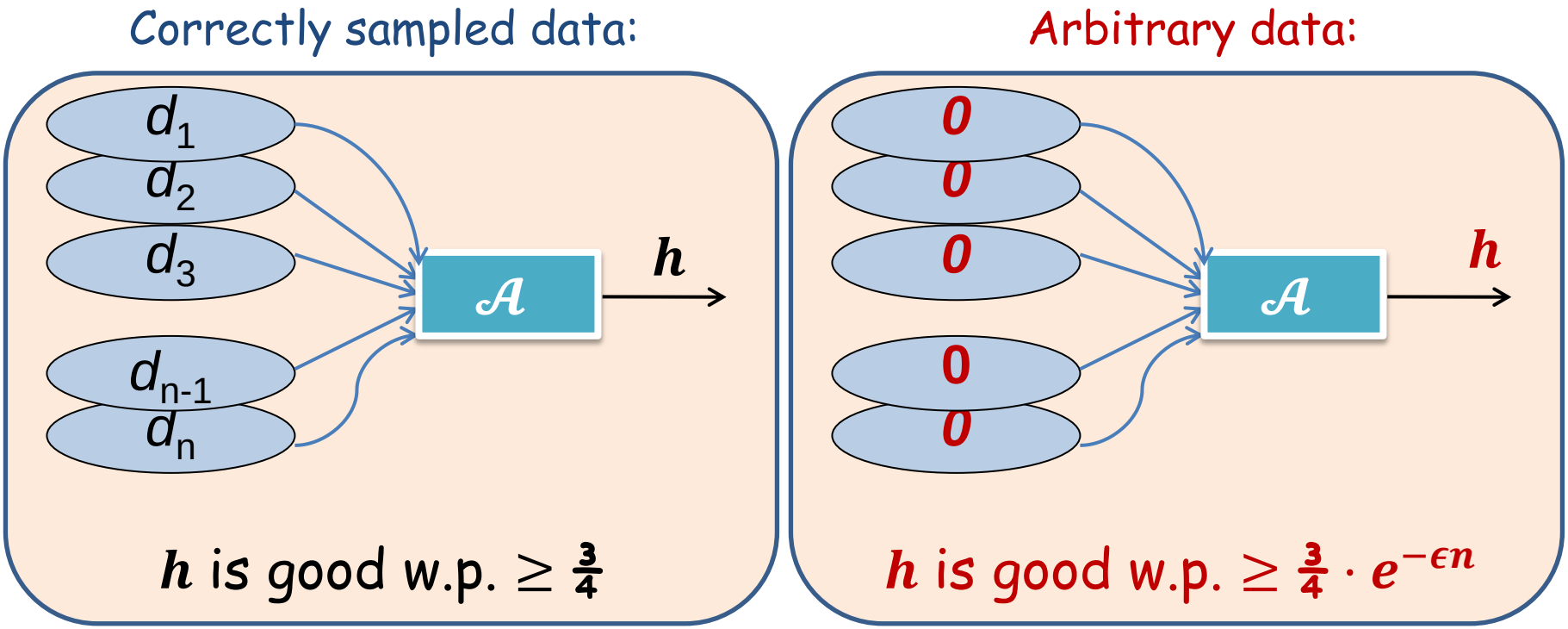
Private Learner $\Longrightarrow$ Prob. Rep.

$\mathcal{A}$: PPAC learner for a concept class C using n examples.



Private Learner $\Longrightarrow$ Prob. Rep.

$\mathcal{A}$: PPAC learner for a concept class C using n examples.



If we run $\mathcal{A}(\mathbf{0})$ for $O(e^{\epsilon n})$ times, we get a list of hypotheses s.t. with probability > $\frac{3}{4}$ contains a good hypothesis.

- Repeat to get enough lists to create a probabilistic representation

A Tight Characterization

$\Theta(\mathbf{RepDim}(\mathcal{C}))$ samples are necessary and sufficient for privately learning $\mathcal{C}$.

- Analogous to the VC dimension for non-private learners.

Prob. Reb vs. Det. Rep.

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Probabilistic Representation:

A **list** of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:

w.p. ¾, **a randomly chosen** $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Prob. Reb vs. Det. Rep.

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Probabilistic Representation:

A **list** of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, **a randomly chosen** $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Naïve construction: given **Rep=$\{\mathcal{H}_1,\mathcal{H}_2,\ldots,\mathcal{H}_r\}$** the union of $\mathcal{H}_1,\mathcal{H}_2,\ldots,\mathcal{H}_r$ is a deterministic representation.

- Size grows by ln r.

Prob. Reb vs. Det. Rep.

Deterministic Representation: A hypothesis class $\mathcal{H}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$, there exists a hypothesis $h_0 \in \mathcal{H}$ s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Probabilistic Representation:

A **list** of hypothesis classes $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_r\}$ s.t. for every $c \in \mathcal{C}$ and $\mathcal{D}$:
w.p. ¾, **a randomly chosen** $\mathcal{H}_i$ contains an h_0 s.t. $\mathbf{error}_{\mathcal{D}}(h_0) \leq ¼$.

Naïve construction: given **Rep=$\{\mathcal{H}_1,\mathcal{H}_2,\ldots,\mathcal{H}_r\}$** the union of $\mathcal{H}_1,\mathcal{H}_2,\ldots,\mathcal{H}_r$ is a deterministic representation.

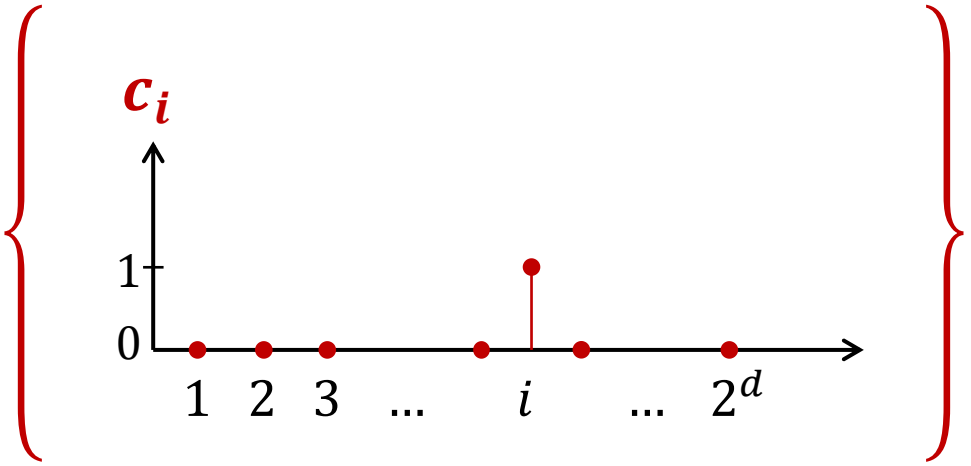
- Size grows by ln r.

Theorem: For every concept class C over the domain $\{0, 1\}^d$ there exists a deterministic representation of C of size O(RepDim(C) + ln d)

- Non-constructive proof via the probabilistic method.

Question

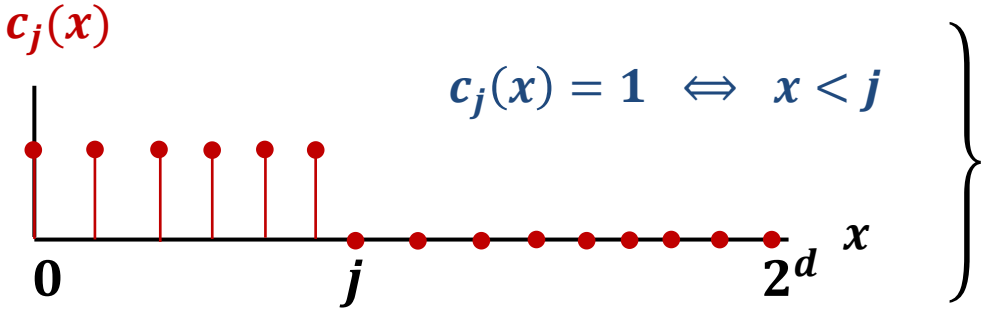
Clearly, RepDim(C) ≥ VC(C). Is there a concept class C for which RepDim(C) >> VC(C) ?

$\mathbf{POINT_d} =$ 

RepDim($\mathbf{POINT_d}$) = VC($\mathbf{POINT_d}$) = $O(1)$.

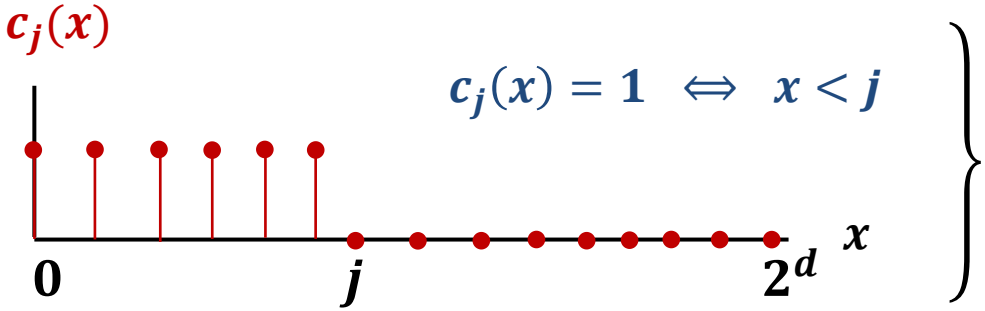
Intervals

$\mathbf{INTERVAL}_d =$



Intervals

$\mathbf{INTERVAL}_d =$



Facts about the sample complexity:

- Non-privately: $O(1)$
- [BKN 2010] ϵ-private proper-learner: $\Omega(d)$
- ϵ-private improper-learner: ???

Approx. D.P. to the Rescue

Dwork, McSherry, Nissim, Smith 2006

Dwork, Kenthapadi, McSherry, Mironov, Naor 2006

A (rand) algorithm $\mathcal{A}$ is (ϵ, δ) differentially private if for all neighboring databases S_1, S_2 and for all sets of outputs F:

$$\Pr[\mathcal{A}(S_1) \in F] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(S_2) \in F] + \delta$$

Approx. D.P. to the Rescue

Dwork, McSherry, Nissim, Smith 2006

Dwork, Kenthapadi, McSherry, Mironov, Naor 2006

A (rand) algorithm $\mathcal{A}$ is (ϵ, δ) differentially private if for all neighboring databases S_1, S_2 and for all sets of outputs F:

$$\Pr[\mathcal{A}(S_1) \in F] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(S_2) \in F] + \delta$$

We Show: (RANDOM'13)

We show (ϵ, δ)-private proper-learner with sample complexity $2^{O(\log^* d)}$.

Summary and Open Problems

What we saw:

- A rich picture for private learners
  - Even for simple classes such as Points and Intervals.
- **Sample complexity of private learners is characterized by the representation dimension.**
  - Similar characterization (not as clean) for other tasks.

Summary and Open Problems

What we saw:

- A rich picture for private learners
  - Even for simple classes such as Points and Intervals.
- **Sample complexity of private learners is characterized by the representation dimension.**
  - Similar characterization (not as clean) for other tasks.

Open problems:

- Tools for lower bounding the representation dimension
- Clearly, RepDim(C) ≥ VC(C).
  Is there a concept class C for which Repdim(C) >> VC(C) ?
- Characterization of the sample complexity for (ε,δ)-differential privacy.