Constructing Hard Functions from Learning Algorithms

Adam Klivans
UT Austin

Pravesh Kothari
UT Austin

Igor Oliveira
Columbia

Constructing Hard Functions from Learning Algorithms

Adam Klivans
UT Austin

Pravesh Kothari
UT Austin

Igor Oliveira
Columbia

Motivation

Circuit Lower Bounds are Hard to prove

What we'd like: $f \in$ NP such that $f \notin P/poly$.

Motivation

Circuit Lower Bounds are Hard to prove

What we'd like: $f \in$ NP such that $f \notin P/poly$.

Explicit circuit lower bounds known for very few circuit classes.

Motivation

Circuit Lower Bounds are Hard to prove

What we'd like: $f \in$ NP such that $f \notin P/poly$.

Explicit circuit lower bounds known for very few circuit classes.

Even for low depth circuits of AND, OR, NOT and mod-m gates, a lower bound eluded us till Williams-2011 breakthrough.

Motivation

Learning Algorithms and Circuit Lower Bounds

"Circuit Lower Bounds precede Learning Algorithms for specific circuit classes."

Furst-Saxe-Sipser-81, Ajtai-83, Yao-85, Hastad 1987,...

Parity $\notin$ AC0

Motivation

Learning Algorithms and Circuit Lower Bounds

"Circuit Lower Bounds precede Learning Algorithms for specific circuit classes."

Furst-Saxe-Sipser-81, Ajtai-83, Yao-85, Hastad 1987,...

Parity $\notin$ AC0

Linial-Mansour-Nisan 1989

Learning AC0 in Quasi-Poly time.

Is there a formal connection between learning algorithms and circuit lower bounds?

Learning Algorithms and Circuit Lower Bounds

Klivans-Fortnow 2005

"Efficient Learning Algorithms Yield Circuit Lower Bounds."

"learning expressive classes is as hard as proving strong circuit lower bounds"

Exact Learning Yields Lower Bounds

Klivans-Fortnow 2005

If C^* is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{EXP}^{\mathrm{NP}} \not\subseteq C$$

*C is any poly-size circuit class.

Exact Learning Yields Lower Bounds

Angluin's Exact Learning Model.

Klivans-Fortnow

If C^* is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{EXP}^{\mathrm{NP}} \not\subseteq C$$

*C is any poly-size circuit class.

Exact Learning Yields Lower Bounds

Hitchcock-Harkins 2011

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\text{EXP} \not\subseteq C$$

PAC Learning with MQs Yields Lower Bounds

Klivans-Fortnow 2005

If C is efficiently PAC learnable with membership queries on the uniform distribution, then:

$$\text{BPEXP} \not\subseteq \mathcal{C}$$

Our Result: I

If C is efficiently exactly learnable
with membership and equivalence queries, then:

$$\mathsf{DTIME}(n^{\omega(1)}) \not\subseteq C$$

Our Result: II

If C is efficiently PAC learnable with membership queries on the uniform distribution, then either:

1. PSPACE $\not\subseteq C$.
2. PSPACE $\subseteq$ BPP .

Our Result: III

If C is efficiently learnable via
statistical queries,
then there exists an explicit function f,
that is average case hard for C.

uses connection to discrepancy minimization.

Lower Bounds from Exact Learning

*C is a poly-size circuit class on the hypercube.

Exact Learning Algorithms

Target Concept $c \in \mathcal{C}$

Learner can ask two types of queries about c

Exact Learning Algorithms

Target Concept $c \in \mathcal{C}$

Learner can ask two types of queries about c

Membership Query

For any x, what is c(x)?

Exact Learning Algorithms

Target Concept $c \in \mathcal{C}$

Learner can ask two types of queries about c

Membership Query

For any x, what is c(x)?

Equivalence Query

For any circuit q, is c = q?

Exact Learning Algorithms

Target Concept $c \in \mathcal{C}$

Learner can ask two types of queries about c

Membership Query

For any x, what is c(x)?

Equivalence Query

For any circuit q, is c = q?

If $q \neq c$, the oracle sends x such that

q(x) = -c(x).

Learner must identify the correct hypothesis c.

Exact Learning Yields Lower Bounds

Klivans-Fortnow 2005

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{EXP}^{\mathrm{NP}} \not\subseteq C$$

Exact Learning Yields Lower Bounds

Klivans-Fortnow 2005

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{EXP}^{\mathrm{NP}} \not\subseteq C$$

- Outline similar to Impagliazzo-Kabanets-04's proof.
- Toda's theorem, collapse\hierarchy theorems, downward self-reducibility of Permanent...

Exact Learning Yields Lower Bounds

Hitchcock-Harkins 2011

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\text{EXP} \not\subseteq C$$

- Uses ideas from Betting Games and Resource Bounded Measure.

High Level Idea of Previous Results

1.Start from a purported hard function (Permanent) and assume it is computed by *C*.

2. Now show that efficient learnability of *C* violates some hierarchy theorem.

Our Result: I

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

Our Result: I

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

- Self-contained, simple and elementary proof.

Our Result: I

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

- Self-contained, simple and elementary proof
- Uses the learning algorithm in a non-standard way to diagonalize against the circuit class and compute a hard function.

Our Result: I

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

Corollary: Efficient exact learning algorithm for linear size circuits implies P = BPP. (Using Impagliazzo-Wigderson-98 generator.)

Our Result: I

If C is efficiently exactly learnable with membership and equivalen

$\mathrm{DTIME}(n^{\omega}$

Learning implies derandomization!

Corollary: Efficient exact learning algorithm for linear size circuits implies P = BPP.

Lower Bounds from PAC Learning with Membership Queries

PAC Learning with MQs

Target Concept $c \in \mathcal{C}$, Uniform Distribution

MQ: Learner can find c(x) for any x on the hypercube.

Learner w.h.p. must output hypothesis h such that:

$$\Pr[h(x) \neq c(x)] \leq \epsilon$$

PAC Learning with MQs Yields Lower Bounds

Klivans-Fortnow 2005

If C is efficiently PAC learnable with membership queries on the uniform distribution, then:

$$\text{BPEXP} \not\subseteq \mathcal{C}$$

Our Result: II

If C is efficiently PAC learnable with membership queries on the uniform distribution, then either:

1. PSPACE $\not\subseteq C$
2. PSPACE $\subseteq$ BPP

PAC Learning with MQs Yields Lower Bounds

Klivans-Fortnow 2005

If C is efficiently PAC learnable with membership queries on the uniform distribution, then:

$$\text{BPEXP} \not\subseteq C$$

IW01-If Permanent is computed by C and C is PAC learnable, then Permanent is computed by BPP.

Our Result: II

If C is efficiently PAC learnable with membership queries on the uniform distribution, then either:

1. PSPACE $\not\subseteq C$
2. PSPACE $\subseteq$ BPP

- Generalize Klivans-Fortnow argument.
- Use Trevisan-Vadhan07's downward self-reducible, self-correctible PSPACE complete function.

Our Result II

If C is efficiently PAC learnable with membership queries on the uniform distribution, then either:

1. PSPACE $\not\subseteq \mathcal{C}$
2. PSPACE $\subseteq$ BPP

As a corollary yields the Fortnow-Klivans result.

Proof

Lower Bounds from Mistake Bound* Learning

*more restrictive (special case of) Exact Learning

Mistake Bound Learning

Target Function: $c \in \mathcal{C}$

Online Learning Model.

1. In each round, learner receives a point x from the hypercube.

Mistake Bound Learning

Target Function: $c \in \mathcal{C}$

Online Learning Model.

1. In each round, learner receives a point x from the hypercube.

2. Learner predicts the label of x based on the current hypothesis.

Mistake Bound Learning

Target Function: $c \in \mathcal{C}$

Online Learning Model.

1. In each round, learner receives a point x from the hypercube.
2. Learner predicts the label of x based on the current hypothesis.
3. Learner is told the correct label of x, and can update the hypothesis.

Mistake Bound Learning

Target Function: $c \in \mathcal{C}$

Online Learning Model.

1. T

2. L

1. T(n): update time in each round
2. M(n): Maximum number of mistakes on any sequence of input points.

3. Learner is told the correct label of x, and can update the hypothesis.

Mistake Bound Learning Yields Lower Bounds

If C^* is learnable in time $T=T(n,s)$ and mistake bound $M=M(n,s)$, then there exists a function f such that

1. $f \in \mathsf{DTIME}(n + M.T)$
2. $f \notin \mathcal{C}$

*$\mathcal{C} = \mathcal{C}^s$: circuits in $\mathcal{C}$ of size at most s.

Mistake Bound Learning Yields Lower Bounds

1. Simulate the oracle for the learner, say A.

Mistake Bound Learning Yields Lower Bounds

1. Simulate the oracle for the learner, say L.
2. Present points in lexicographic order from the hypercube in each round.

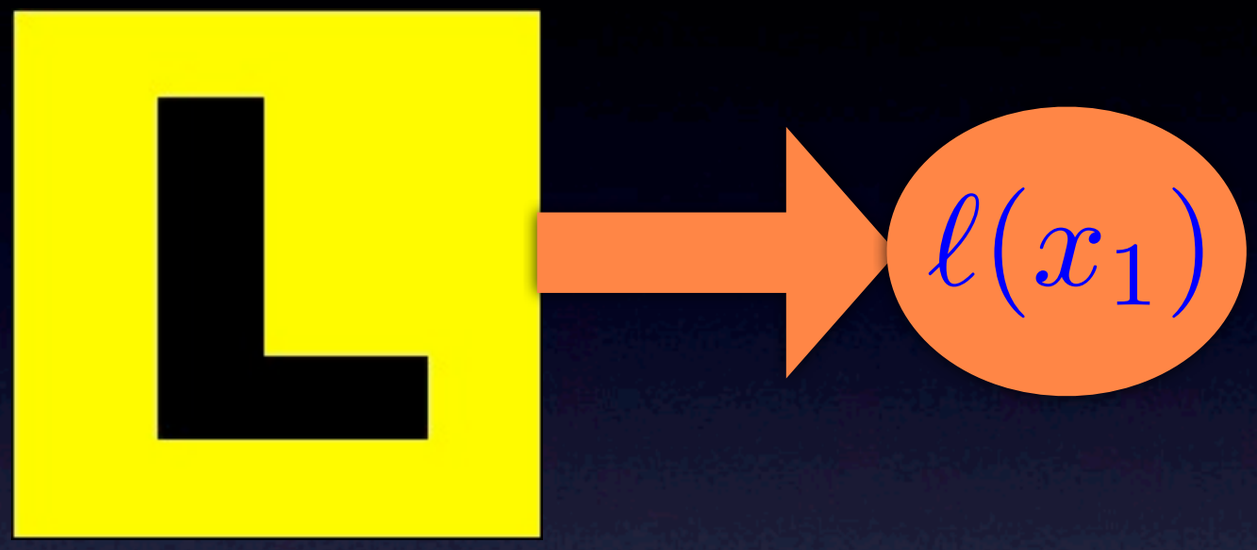
Mistake Bound Learning Yields Lower Bounds

x_1

Mistake Bound Learning Yields Lower Bounds

3. Ask L to report its label: l(x)

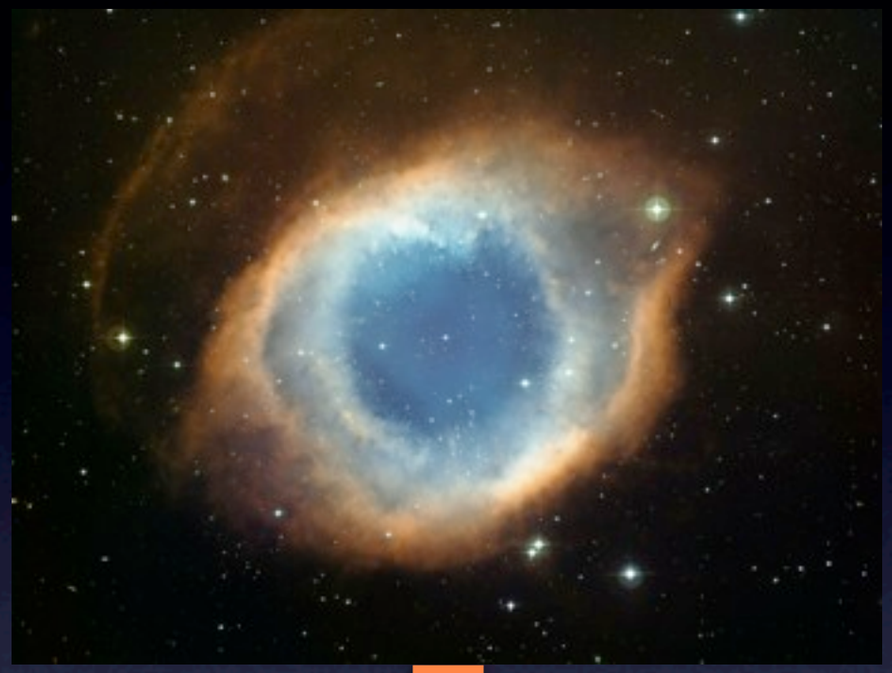
Mistake Bound Learning Yields Lower Bounds



Mistake Bound Learning Yields Lower Bounds

5. Set $f(x) = -l(x)$

Mistake Bound Learning Yields Lower Bounds



$$f(x_1) = -\ell(x_1)$$

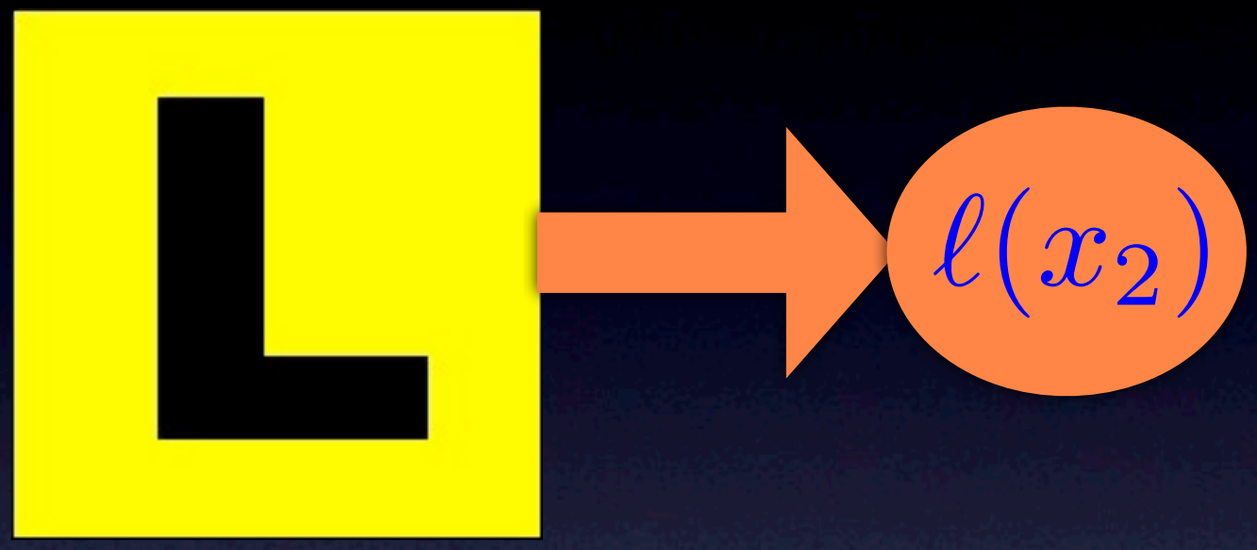
Mistake Bound Learning Yields Lower Bounds

Update Hypothesis.

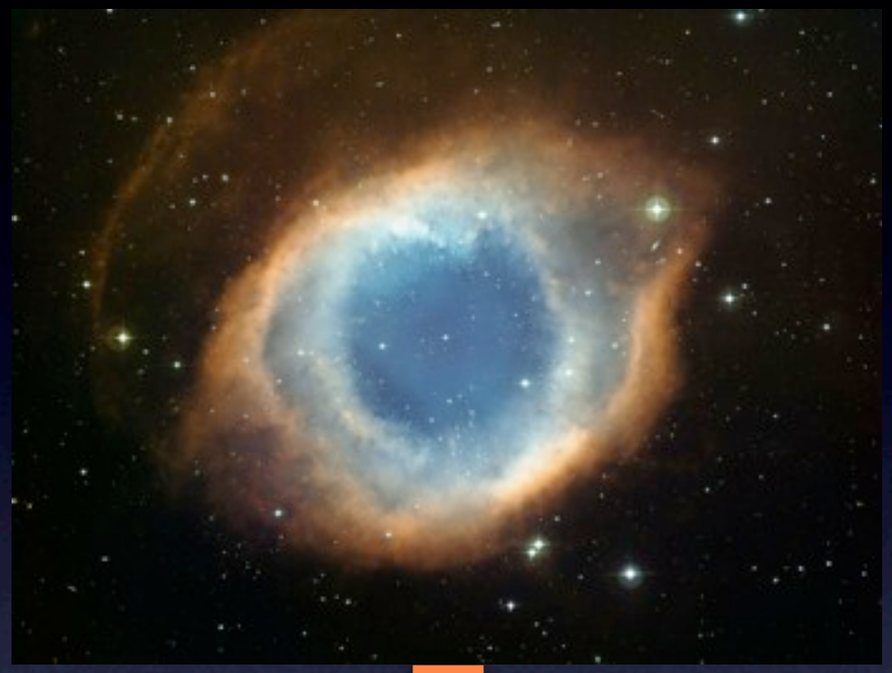
Mistake Bound Learning Yields Lower Bounds

x_2

Mistake Bound Learning Yields Lower Bounds



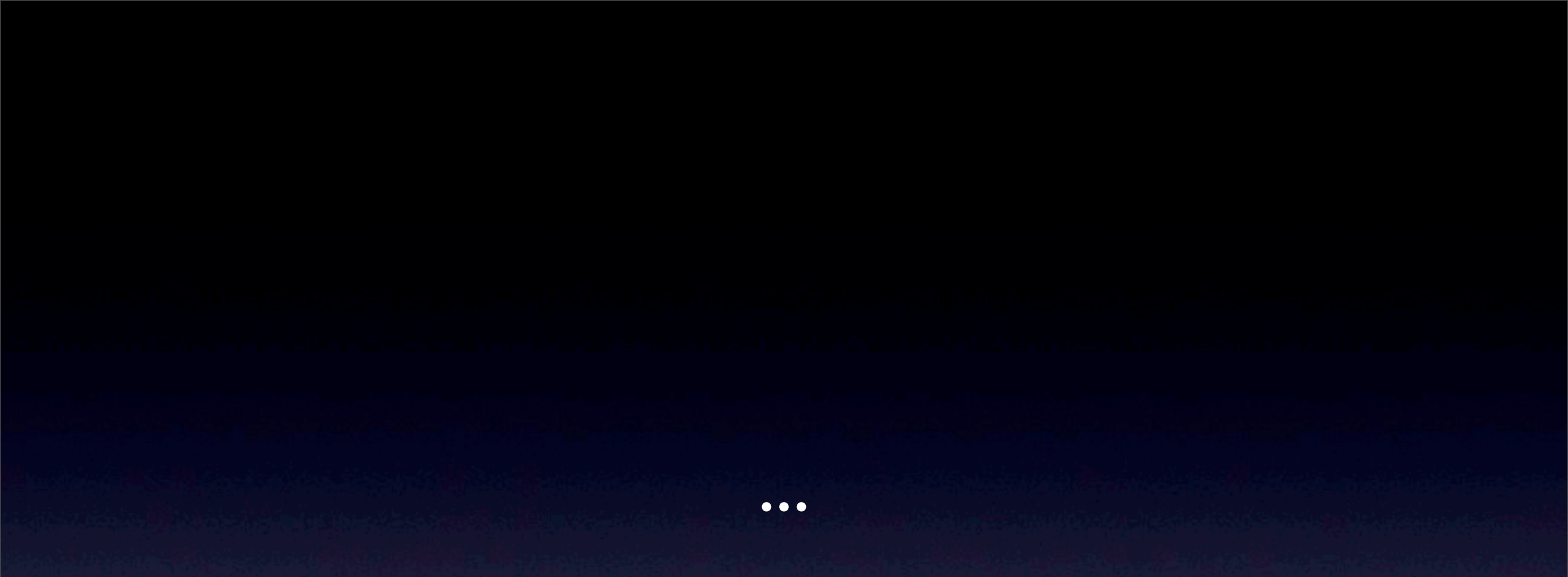
Mistake Bound Learning Yields Lower Bounds



$$f(x_2) = -\ell(x_2)$$

Mistake Bound Learning Yields Lower Bounds

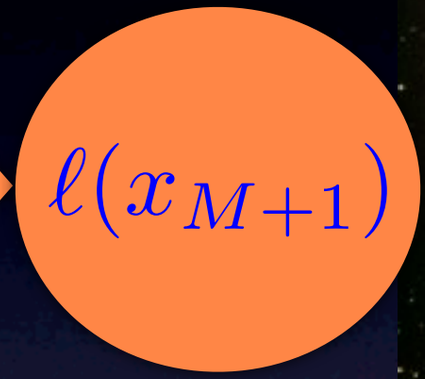
Update Hypothesis.



Mistake Bound Learning Yields Lower Bounds

x_{M+1}

Mistake Bound Learning Yields Lower Bounds



Mistake Bound Learning Yields Lower Bounds

$$f(x_{M+1}) = -\ell(x_{M+1})$$

Mistake Bound Learning Yields Lower Bounds

Update Hypothesis.

Mistake Bound Learning Yields Lower Bounds

For M+1 rounds:

1. Simulate the oracle for the learner, say A.
2. Present points in lexicographic order from the hypercube in each round.
3. Ask A to report its label: l(x)
4. Report that A made a mistake.
5. Set f(x) = -l(x)

Mistake Bound Learning Yields Lower Bounds

- To compute f at y, check if y is the lexicographically first M+1 points of the hypercube.

- If yes, then f(y) is already set in the procedure we ran and we return that value.

- Otherwise, we return -1.

Mistake Bound Learning Yields Lower Bounds

For M+1 rounds:

1. Sim…, say A.
2. Pre… fro… …er
3. Ask A to report its label: l(x)
4. Report that A made a mistake.
5. Set f(x) = -l(x)

Key Intuition

Learning Algorithm is used to diagonalize over the circuit class!

Mistake Bound Learning Yields Lower Bounds

For M+1 rounds:

1.
2.
3.
4.
5.

1. All labels reported are consistent with f.
2. A guaranteed to be correct if $f \in \mathcal{C}$.
3. In which case, A makes M+1 mistakes.

Contradiction!

Mistake Bound Learning to Average Case Hardness

If C^* is learnable in time $T=T(n,s)$ and mistake bound $M=M(n,s)$, then there exists a function f such that

$$1. f \in \mathsf{DTIME}(n + M.T)$$

$$2. \forall c \in \mathcal{C},\ \Pr[f \neq c] \geq 1/2M$$

*$\mathcal{C} = \mathcal{C}^s$: circuits in $\mathcal{C}$ of size at most s.

Exact Learning to Lower Bounds

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

Extend the idea, handling some technical difficulties.

Exact Learning to Lower Bounds

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

Caveat: Learner needs to be deterministic. (All known algorithms happen to be deterministic.)

Summary

If C is efficiently exactly learnable with membership and equivalence queries, then:

$$\mathrm{DTIME}(n^{\omega(1)}) \not\subseteq C$$

If C is efficiently PAC learnable with membership queries on the uniform distribution, then either:

1. $\mathrm{PSPACE} \not\subseteq C$
2. $\mathrm{PSPACE} \subseteq \mathrm{BPP}$

Interpretation

Indicates hardness of designing learning algorithms for very expressive classes.

Interpretation

Indicates hardness of designing learning algorithms for very expressive classes.

Strategy to prove new lower bounds?

"Non-trivial algorithms yield lower bounds."

Karp-Lipton 1980

If SAT can be solved by polynomial size circuits, then PH collapses to Σ_2.

"Non-trivial algorithms yield lower bounds."

Karp-Lipton 1980

If SAT can be solved by polynomial size circuits, then PH collapses to Σ_2.

Impagliazzo-Kabanets-Wigderson 2004

If there is a deterministic algorithm for polynomial identity testing, then either:

1. NEXP $\not\subseteq$ P/poly or
2. Permanent does not have poly-size circuits.

"Non-trivial algorithms yield lower bounds."

Williams 2010,2011

If there exists an algorithm that solves satisfiability instances of a boolean circuit class C in "slightly-faster-than-trivial" time, then

$\mathrm{NEXP} \not\subseteq C$.

"Non-trivial algorithms yield lower bounds."

Williams 2010,2011

If there exists an algorithm that solves satisfiability instances of a boolean circuit class C in "slightly-faster-than-trivial" time, then

$NEXP \not\subseteq C$.

$NEXP \not\subseteq ACC$!


Questions?

Lower Bounds from Statistical Query Learning

Correlational* Statistical Query Learning

Target Concept $c \in \mathcal{C}$, Uniform Distribution*

1. Every query of the learner is a function q and expects E[q(x).c(x)].
2. Query must be answered within additive error of at most τ.

*more general statements possible.

Correlational Statistical Query Learning

Target Concept $c \in \mathcal{C}$, Uniform Distribution

1. Every query of the learner is a function q* and expects E[q(x).c(x)].
2. Query must be answered within additive error of at most τ.

*efficiently representable, computable

Correlational Statistical Query Learning

Target Concept $c \in \mathcal{C}$, Uniform Distribution

1. Every query of the learner is a function q and expects E[q(x)
2. Query must be answered within additive error of at most τ.

tolerance

*more general statements possible.

Correlational Statistical Query Learning

Target Concept $c \in \mathcal{C}$, Uniform Distribution

1. Every query of the learner is a function q and expects E[q(x).c(x)].
2. Query must be answered within additive error of at most τ.

Hypothesis h must satisfy:

$$\Pr[h(x) \neq c(x)] \leq \epsilon$$

Learning Algorithms and Circuit Lower Bounds

Learning Algorithms $\Rightarrow$ Circuit Lower Bounds?

Mistake Bound Learning Yields Lower Bounds

Arrange points on the hypercube in lexicographic order. Partition hypercube into consecutive elements of size M+1.

For M+1 rounds:

1. Simulate the oracle for the learner, say A.
2. Find the set that contains y. Present points x from the set of size M+1 containing y.
3. Ask A to report label: l(x)
4. Report that A made a mistake.
5. Set f(x) = -l(x)