Strengths and weaknesses of quantum examples

Srinivasan Arunachalam (MIT)

joint with Ronald de Wolf (CWI, Amsterdam) and others

Machine learning

Classical machine learning

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go
- Why the recent interest? Flood of available data, increasing computational power, growing progress in algorithms

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go
- Why the recent interest? Flood of available data, increasing computational power, growing progress in algorithms

Quantum machine learning

- What can quantum computing do for machine learning?

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go
- Why the recent interest? Flood of available data, increasing computational power, growing progress in algorithms

Quantum machine learning

- What can quantum computing do for machine learning?
- The learner will be quantum, the data may be quantum

Machine learning

Classical machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go
- Why the recent interest? Flood of available data, increasing computational power, growing progress in algorithms

Quantum machine learning

- What can quantum computing do for machine learning?
- The learner will be quantum, the data may be quantum
- Some examples are known of reduction in time complexity:
  - clustering (Aïmeur et al. '13)
  - Principal component analysis (Lloyd et al. '13)
  - perceptron learning (Wiebe et al. '16)
  - recommendation systems (Kerenidis & Prakash '16)

Learning using classical examples

Learning using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)

Learning using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$ (Unknown)

Learning using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$ (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$

Learning using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$ (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$
- Labeled example for $c \in \mathcal{C}$: $(x, c(x))$ where $x \sim D$

Classical learner using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$. (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$
- Labeled example for $c \in \mathcal{C}$: $(x, c(x))$ where $x \sim D$

$\mathcal{C}$

$\downarrow$

c
target concept

Learner is trying to learn c

Classical learner using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$. (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$
- Labeled example for $c \in \mathcal{C}$: $(x, c(x))$ where $x \sim D$

$\mathcal{C}$

$\downarrow$

c
target concept

$$x_1 \sim D \quad \longrightarrow \quad (x_1, c(x_1))$$

Learner is trying to learn c

Classical learner using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$. (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$
- Labeled example for $c \in \mathcal{C}$: $(x, c(x))$ where $x \sim D$

$$\mathcal{C}$$

$$\downarrow$$

c
target concept

$$x_1 \sim D \longrightarrow (x_1, c(x_1))$$

$$x_2 \sim D \longrightarrow (x_2, c(x_2))$$

Learner is trying to learn c

Classical learner using classical examples

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target concept c: some function $c \in \mathcal{C}$. (Unknown)
- Distribution $D : \{0,1\}^n \to [0,1]$
- Labeled example for $c \in \mathcal{C}$: $(x, c(x))$ where $x \sim D$

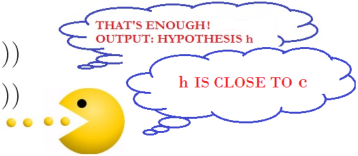
$\mathcal{C}$

$\downarrow$

c

target concept

$$
\begin{array}{ccc}
x_1 \sim D & \longrightarrow & (x_1, c(x_1)) \\
x_2 \sim D & \longrightarrow & (x_2, c(x_2)) \\
\vdots & & \vdots \\
x_T \sim D & \longrightarrow & (x_T, c(x_T))
\end{array}
$$

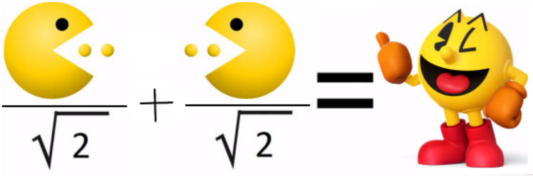


Learner is trying to learn c

Quantum learning using quantum examples

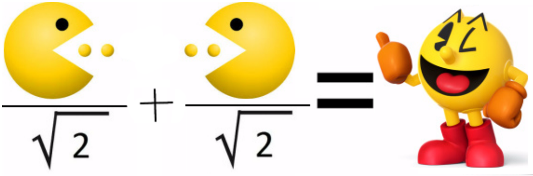
Quantum learning using quantum examples

- Learner is quantum:



Quantum learning using quantum examples

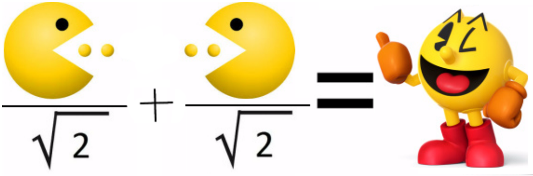
- Learner is quantum:



- Data is quantum: Bshouty-Jackson'95 introduced a quantum example as a superposition

$$\sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$$

- Learner is quantum:



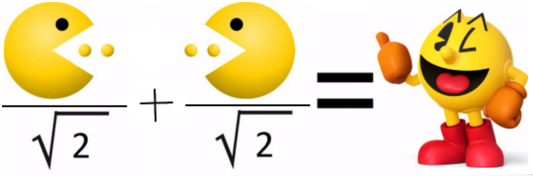
- Data is quantum: Bshouty-Jackson'95 introduced a quantum example as a superposition

$$\sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$$

Measuring this state gives a $(x, c(x))$ with probability $D(x)$,

Quantum learning using quantum examples

- Learner is quantum:



- Data is quantum: Bshouty-Jackson'95 introduced a quantum example as a superposition

$$\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$$

Measuring this state gives a $(x, c(x))$ with probability $D(x)$, so quantum examples are at least as powerful as classical

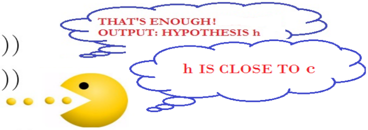
Motivating question for this talk

Fix a concept class $\mathcal{C}$, distribution $D : \{0,1\}^n \to [0,1]$

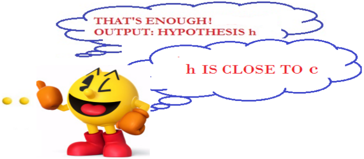
Motivating question for this talk

Fix a concept class $\mathcal{C}$, distribution $D : \{0,1\}^n \to [0,1]$

$$x_1 \sim D \longrightarrow (x_1, c(x_1))$$
$$x_2 \sim D \longrightarrow (x_2, c(x_2))$$
$$\vdots \qquad\qquad \vdots$$
$$x_T \sim D \longrightarrow (x_T, c(x_T))$$



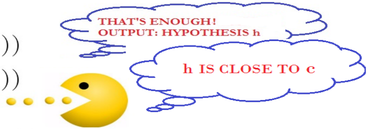
$$\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle \longrightarrow$$
$$\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle \longrightarrow$$
$$\vdots \qquad\qquad \vdots$$
$$\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle \longrightarrow$$



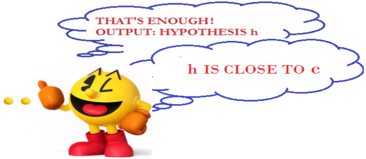
Motivating question for this talk

Fix a concept class $\mathcal{C}$, distribution $D : \{0,1\}^n \to [0,1]$

$$
\begin{aligned}
x_1 \sim D &\longrightarrow (x_1, c(x_1)) \\
x_2 \sim D &\longrightarrow (x_2, c(x_2)) \\
\vdots \quad & \qquad\quad \vdots \\
x_T \sim D &\longrightarrow (x_T, c(x_T))
\end{aligned}
$$



$$
\begin{aligned}
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle &\longrightarrow \\
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle &\longrightarrow \\
\vdots \qquad\quad & \quad\ \vdots \\
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle &\longrightarrow
\end{aligned}
$$



Question

Understanding the concept classes $\mathcal{C}$ and distributions D where fewer quantum examples suffice for a quantum learner

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$
- Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is close to c w.r.t. D,

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$
- Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is close to c w.r.t. D, i.e., $err_D(c,h) = \Pr_{x\sim D}[c(x) \neq h(x)]$ should be small

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$
- Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is close to c w.r.t. D, i.e., $err_D(c,h) = \Pr_{x\sim D}[c(x) \neq h(x)]$ should be small

Distribution-dependent learning (for a fixed D)

- An algorithm (ε, δ)-learns $\mathcal{C}$ under D if:

$$\forall c \in \mathcal{C}: \quad \Pr[\underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}}] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$
- Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is close to c w.r.t. D, i.e., $err_D(c,h) = \Pr_{x \sim D}[c(x) \neq h(x)]$ should be small

Distribution-dependent learning (for a fixed D)

- An algorithm (ε,δ)-learns $\mathcal{C}$ under D if:

$$\forall c \in \mathcal{C} : \quad \Pr[\ \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}}\ ] \geq \underbrace{1-\delta}_{\text{Probably}}$$

PAC learning (Distribution-independent learning for every D)

Distribution (in)dependent PAC learning

- Focus on Probably Approximately Correct (PAC) model of learning
- Fix $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and $D : \{0,1\}^n \to [0,1]$
- Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is close to c w.r.t. D, i.e., $err_D(c,h) = \Pr_{x\sim D}[c(x) \neq h(x)]$ should be small

Distribution-dependent learning (for a fixed D)

- An algorithm (ε, δ)-learns $\mathcal{C}$ under D if:

$$\forall c \in \mathcal{C} : \quad \Pr[\ \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}}\ ] \geq \underbrace{1-\delta}_{\text{Probably}}$$

PAC learning (Distribution-independent learning for every D)

- An algorithm (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \ \ \forall c \in \mathcal{C} : \quad \Pr[\ \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}}\ ] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

In this talk

- Strengths of quantum examples

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

In this talk

- Strengths of quantum examples

  ACLW'18: Sample complexity of learning Fourier-sparse Boolean functions under uniform D

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

In this talk

- Strengths of quantum examples

  ACLW'18: Sample complexity of learning Fourier-sparse Boolean functions under uniform D

  Bshouty-Jackson'95: Quantum polynomial time learnability of DNFs under uniform D

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

In this talk

- Strengths of quantum examples

  ACLW'18: Sample complexity of learning Fourier-sparse Boolean functions under uniform D

  Bshouty-Jackson'95: Quantum polynomial time learnability of DNFs under uniform D

  ACKW'18: Quantum examples can help the coupon collector

Complexity of learning

How to measure the efficiency of the classical or quantum learner?

- Sample complexity: number of labeled examples used by learner
- Time complexity: number of time-steps used by learner

In this talk

- Strengths of quantum examples

  ACLW'18: Sample complexity of learning Fourier-sparse Boolean functions under uniform D

  Bshouty-Jackson'95: Quantum polynomial time learnability of DNFs under uniform D

  ACKW'18: Quantum examples can help the coupon collector

- Weaknesses of quantum examples

  AW'17: Quantum examples are not more powerful than classical examples for PAC learning

Fourier sampling: a useful trick under uniform D

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$.

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 =$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2]$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\widehat{c}(S)^2\}_S$ forms a probability distribution

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\widehat{c}(S)^2\}_S$ forms a probability distribution
- Given quantum example under uniform D:

$$\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\widehat{c}(S)^2\}_S$ forms a probability distribution
- Given quantum example under uniform D:

$$\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle \xrightarrow{\text{Hadamard}} \sum_S \widehat{c}(S)\,|S\rangle$$

Fourier sampling: a useful trick under uniform D

- Let $c : \{0,1\}^n \to \{-1,1\}$. Then the Fourier coefficients are

$$\widehat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0,1\}^n$$

- Parseval's identity: $\sum_S \widehat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\widehat{c}(S)^2\}_S$ forms a probability distribution
- Given quantum example under uniform D:

$$\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle \xrightarrow{\text{Hadamard}} \sum_S \widehat{c}(S) \, |S\rangle$$

- Measuring allows to sample from the Fourier distribution $\{\widehat{c}(S)^2\}_S$

Applications of Fourier sampling

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c$ is a ℓ-junta$\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

Definition: We say c is k-Fourier sparse if $|\{S : \widehat{c}(S) \neq 0\}| \leq k$.

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

Definition: We say c is k-Fourier sparse if $|\{S : \widehat{c}(S) \neq 0\}| \leq k$.

Note that $\mathcal{C}_1$ is 1-Fourier sparse and $\mathcal{C}_2$ is 2^ℓ-Fourier sparse

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c$ is a ℓ-junta$\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

Definition: We say c is k-Fourier sparse if $|\{S : \widehat{c}(S) \neq 0\}| \leq k$.

Note that $\mathcal{C}_1$ is 1-Fourier sparse and $\mathcal{C}_2$ is 2^ℓ-Fourier sparse

Consider the concept class $\mathcal{C} = \{c : \{0,1\}^n \to \{-1,1\} : c$ is k-Fourier sparse$\}$

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

Definition: We say c is k-Fourier sparse if $|\{S : \widehat{c}(S) \neq 0\}| \leq k$.

Note that $\mathcal{C}_1$ is 1-Fourier sparse and $\mathcal{C}_2$ is 2^ℓ-Fourier sparse

Consider the concept class $\mathcal{C} = \{c : \{0,1\}^n \to \{-1,1\} : c \text{ is } k\text{-Fourier sparse}\}$

Observe that $\mathcal{C}_1 \subseteq \mathcal{C}$. $\mathcal{C}$ contains linear functions

Applications of Fourier sampling

- Consider the concept class of linear functions $\mathcal{C}_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

  Classical: $\Omega(n)$ classical examples needed

  Quantum: 1 quantum example suffices to learn $\mathcal{C}_1$ (Bernstein-Vazirani'93)

- Consider $\mathcal{C}_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on ℓ bits of x

  Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform D

  Quantum: $\mathcal{C}_2$ can be *exactly* learnt using $\widetilde{O}(2^\ell)$ quantum examples and in time $\widetilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

Definition: We say c is k-Fourier sparse if $|\{S : \widehat{c}(S) \neq 0\}| \leq k$.

Note that $\mathcal{C}_1$ is 1-Fourier sparse and $\mathcal{C}_2$ is 2^ℓ-Fourier sparse

Consider the concept class $\mathcal{C} = \{c : \{0,1\}^n \to \{-1,1\} : c \text{ is } k\text{-Fourier sparse}\}$

Observe that $\mathcal{C}_1 \subseteq \mathcal{C}$. $\mathcal{C}$ contains linear functions
Observe that $\mathcal{C}_2 \subseteq \mathcal{C}$. $\mathcal{C}$ contains $(\log k)$-juntas

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

- Use Fourier sampling to sample $S \sim \{\widehat{c}(S)^2\}_S$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

- Use Fourier sampling to sample $S \sim \{\widehat{c}(S)^2\}_S$
- Collect Ss until the learner learns the Fourier span of c, $\mathcal{V} = \text{span}\{S : \widehat{c}(S) \neq 0\}$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

- Use Fourier sampling to sample $S \sim \{\widehat{c}(S)^2\}_S$
- Collect Ss until the learner learns the Fourier span of c, $\mathcal{V} = \text{span}\{S : \widehat{c}(S) \neq 0\}$
- Suppose $\dim(\mathcal{V}) = r$, then $\widetilde{O}(rk)$ quantum examples suffice to find $\mathcal{V}$

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

- Use Fourier sampling to sample $S \sim \{\widehat{c}(S)^2\}_S$
- Collect Ss until the learner learns the Fourier span of c, $\mathcal{V} = \text{span}\{S : \widehat{c}(S) \neq 0\}$
- Suppose $\dim(\mathcal{V}) = r$, then $\widetilde{O}(rk)$ quantum examples suffice to find $\mathcal{V}$
- Use the result of [HR'15] to learn c' completely using $\widetilde{O}(rk)$ classical examples

Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

- Exact learning $\mathcal{C}$ under the uniform distribution D
- Classically (Haviv-Regev'15): $\widetilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (**A**CLW'18): $\widetilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size n)

  $\widetilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

Sketch of upper bound

- Use Fourier sampling to sample $S \sim \{\widehat{c}(S)^2\}_S$
- Collect Ss until the learner learns the Fourier span of c, $\mathcal{V} = \text{span}\{S : \widehat{c}(S) \neq 0\}$
- Suppose $\dim(\mathcal{V}) = r$, then $\widetilde{O}(rk)$ quantum examples suffice to find $\mathcal{V}$
- Use the result of [HR'15] to learn c' completely using $\widetilde{O}(rk)$ classical examples
- Since $r \leq \widetilde{O}(\sqrt{k})$ for every $c \in \mathcal{C}$ [Sanyal'15], we get $\widetilde{O}(k^{1.5})$ upper bound

Learning Disjunctive normal Forms (DNF)

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables.

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c$ is an s-term DNF in n variables$\}$ under uniform D

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c$ is an s-term DNF in n variables$\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$
We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

- Structural property: if c is an s-term DNF, then there exists U s.t. $|\widehat{c}(U)| \geq \frac{1}{s}$

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c$ is an s-term DNF in n variables$\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

- Structural property: if c is an s-term DNF, then there exists U s.t. $|\widehat{c}(U)| \geq \frac{1}{s}$
- Fourier sampling! Sample $T \sim \{\widehat{c}(T)^2\}_T$, poly$(s)$ many times to see such a U

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

- Structural property: if c is an s-term DNF, then there exists U s.t. $|\widehat{c}(U)| \geq \frac{1}{s}$
- Fourier sampling! Sample $T \sim \{\widehat{c}(T)^2\}_T$, poly$(s)$ many times to see such a U
- Construct a "weak learner" who outputs χ_U s.t. $\Pr[\chi_U(x) = c(x)] = \frac{1}{2} + \frac{1}{s}$

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$

We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

- Structural property: if c is an s-term DNF, then there exists U s.t. $|\widehat{c}(U)| \geq \frac{1}{s}$
- Fourier sampling! Sample $T \sim \{\widehat{c}(T)^2\}_T$, poly$(s)$ many times to see such a U
- Construct a "weak learner" who outputs χ_U s.t. $\Pr[\chi_U(x) = c(x)] = \frac{1}{2} + \frac{1}{s}$
- Not good enough! Want an hypothesis that agrees with c on most inputs x's

Learning Disjunctive normal Forms (DNF)

DNFs

Simply an OR of AND of variables. For example, $(x_1 \wedge x_4 \wedge \overline{x_3}) \vee (\overline{x_4} \wedge x_6 \wedge x_7 \wedge \overline{x_8})$
We say a DNF on n variables is an s-term DNF if number of clauses is $\leq s$

Learning $\mathcal{C} = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\}$ under uniform D

- Classically: Efficient learning using examples is a longstanding open question. Best known upper bound is $n^{O(\log n)}$ [Verbeurgt'90]
- Quantumly: Bshouty-Jackson'95 gave a polynomial-time quantum algorithm!

Proof sketch of quantum upper bound

- Structural property: if c is an s-term DNF, then there exists U s.t. $|\widehat{c}(U)| \geq \frac{1}{s}$
- Fourier sampling! Sample $T \sim \{\widehat{c}(T)^2\}_T$, poly$(s)$ many times to see such a U
- Construct a "weak learner" who outputs χ_U s.t. $\Pr[\chi_U(x) = c(x)] = \frac{1}{2} + \frac{1}{s}$
- Not good enough! Want an hypothesis that agrees with c on most inputs x's
- Boosting: Run weak learner many times in some manner to obtain a strong learner who outputs h satisfying $\Pr[h(x) = c(x)] \geq 2/3$

Pretty good measurement for state identification

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions.
  Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\, |x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\, |x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated,

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated,

  but we can always use the Pretty Good Measurement (PGM)

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated,
  but we can always use the Pretty Good Measurement (PGM)
- If P_{opt} is the success probability of the optimal measurement,

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions.
  Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated,
  but we can always use the Pretty Good Measurement (PGM)
- If P_{opt} is the success probability of the optimal measurement,
  P_{pgm} is the success probability of the PGM,

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated,
  but we can always use the Pretty Good Measurement (PGM)
- If P_{opt} is the success probability of the optimal measurement,
  P_{pgm} is the success probability of the PGM, then
  $P_{opt} \geq P_{pgm}$

Pretty good measurement for state identification

- Consider a concept class $\mathcal{C}$ consisting of n-bit Boolean functions. Let $D : \{0,1\}^n \to [0,1]$ be a distribution
- For $c \in \mathcal{C}$, a quantum example is $|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle$
- State identification: For uniform $c \in \mathcal{C}$ (unknown), given $|\psi_c\rangle^{\otimes T}$, identify c
- Optimal measurement could be quite complicated, but we can always use the Pretty Good Measurement (PGM)
- If P_{opt} is the success probability of the optimal measurement, P_{pgm} is the success probability of the PGM, then $P_{opt} \geq P_{pgm} \geq P_{opt}^2$ (Barnum-Knill'02)

Quantum examples help the coupon collector

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons.

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons.

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$.

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \setminus \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Suppose a quantum learner obtains quantum examples $\frac{1}{\sqrt{N-1}} \sum_{i \in (\{1,\ldots,N\} \backslash \{i^*\})} |i\rangle$.

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Suppose a quantum learner obtains quantum examples $\frac{1}{\sqrt{N-1}} \sum_{i \in (\{1,\ldots,N\} \backslash \{i^*\})} |i\rangle$.
How many quantum examples before learning i^*?

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Suppose a quantum learner obtains quantum examples $\frac{1}{\sqrt{N-1}} \sum_{i \in (\{1,\ldots,N\} \backslash \{i^*\})} |i\rangle$.
How many quantum examples before learning i^*?

Answer [**A**CKW'..]: Can learn i^* using $\Theta(N)$ quantum examples

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \setminus \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Suppose a quantum learner obtains quantum examples $\frac{1}{\sqrt{N-1}} \sum_{i \in (\{1,\ldots,N\} \setminus \{i^*\})} |i\rangle$. How many quantum examples before learning i^*?

Answer [**A**CKW'..]: Can learn i^* using $\Theta(N)$ quantum examples

Proof idea: Analyze the success probability using the pretty good measurement.

Quantum examples help the coupon collector

Standard coupon collector

Problem: Suppose there are N coupons. How many coupons to draw (with replacement) before having seen each coupon at least once?

Answer: Simple probability analysis shows $\Theta(N \log N)$

Variation to coupon collector

Problem: Suppose there are N coupons. Fix unknown $i^* \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \backslash \{i^*\}$ before learning i^*?

Answer: Same analysis as earlier shows $\Theta(N \log N)$

What if we are given "quantum examples"

Suppose a quantum learner obtains quantum examples $\frac{1}{\sqrt{N-1}} \sum_{i \in (\{1,\ldots,N\} \backslash \{i^*\})} |i\rangle$. How many quantum examples before learning i^*?

Answer [**A**CKW'..]: Can learn i^* using $\Theta(N)$ quantum examples

Proof idea: Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{opt} \geq P_{pgm} \geq 2/3$

Distribution-*in*dependent learning

Recall: PAC learning

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \quad \forall c \in \mathcal{C} \quad : \Pr[\quad \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}} \quad] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \ \ \forall c \in \mathcal{C} \ \ : \Pr[\ \underbrace{err_D(c,h) \le \varepsilon}_{\text{Approximately Correct}}\ ] \ge \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \ \ \forall c \in \mathcal{C} \ \ : \Pr[\ \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}}\ ] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \ \ \forall c \in \mathcal{C} \ \ : \Pr[\ \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}} \ ] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Classical PAC learning sample complexity is characterized by the VC-dimension of $\mathcal{C}$

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \;\; \forall c \in \mathcal{C} \;\; : \Pr[\; \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}} \;] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Classical PAC learning sample complexity is characterized by the VC-dimension of $\mathcal{C}$

Fundamental theorem of PAC learning

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \ \ \forall c \in \mathcal{C} \ \ : \Pr[\ \underbrace{err_D(c,h) \le \varepsilon}_{\text{Approximately Correct}}\ ] \ge \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Classical PAC learning sample complexity is characterized by the VC-dimension of $\mathcal{C}$

Fundamental theorem of PAC learning

Suppose VC-dim$(\mathcal{C}) = d$

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \;\; \forall c \in \mathcal{C} \;\; : \Pr[\; \underbrace{err_D(c,h) \le \varepsilon}_{\text{Approximately Correct}} \;] \ge \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Classical PAC learning sample complexity is characterized by the VC-dimension of $\mathcal{C}$

Fundamental theorem of PAC learning

Suppose VC-dim$(\mathcal{C}) = d$

- Blumer-Ehrenfeucht-Haussler-Warmuth'86:
  every (ε, δ)-PAC learner for $\mathcal{C}$ needs $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ examples

Distribution-*in*dependent learning

Recall: PAC learning

Given $(x, c(x))$ examples where $x \sim D$, a learner (ε, δ)-PAC-learns $\mathcal{C}$ if:

$$\forall D \;\; \forall c \in \mathcal{C} \;\; : \Pr[\; \underbrace{err_D(c,h) \leq \varepsilon}_{\text{Approximately Correct}} \;] \geq \underbrace{1-\delta}_{\text{Probably}}$$

Complexity measure: Number of labelled examples

For a concept class $\mathcal{C}$, associate a combinatorial parameter called VC-dimension of $\mathcal{C}$.

Classical PAC learning sample complexity is characterized by the VC-dimension of $\mathcal{C}$

Fundamental theorem of PAC learning

Suppose VC-dim$(\mathcal{C}) = d$

- Blumer-Ehrenfeucht-Haussler-Warmuth'86:
  every (ε, δ)-PAC learner for $\mathcal{C}$ needs $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ examples
- Hanneke'16: exists an (ε, δ)-PAC learner for $\mathcal{C}$ using $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ examples

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification.

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification. For a good learner $P_{opt} \geq 2/3$, so $P_{pgm} \geq P_{opt}^2 \geq 4/9$.

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification. For a good learner $P_{opt} \geq 2/3$, so $P_{pgm} \geq P_{opt}^2 \geq 4/9$. If $P_{pgm} \geq 4/9$, then $T = \Omega\left(\frac{d}{\varepsilon}\right)$

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification. For a good learner $P_{opt} \geq 2/3$, so $P_{pgm} \geq P_{opt}^2 \geq 4/9$. If $P_{pgm} \geq 4/9$, then $T = \Omega\left(\frac{d}{\varepsilon}\right)$

Quantum examples are no better than classical examples for PAC learning

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification. For a good learner $P_{opt} \geq 2/3$, so $P_{pgm} \geq P_{opt}^2 \geq 4/9$. If $P_{pgm} \geq 4/9$, then $T = \Omega\left(\frac{d}{\varepsilon}\right)$

Quantum examples are no better than classical examples for PAC learning

Let's get real!

- In computational learning theory, agnostic learning and learning under classification noise is a theoretical way to model noise in data

VC-dimension and quantum sample complexity

Quantum bounds

- Classical upper bound $O\left(\frac{d}{\varepsilon}+\frac{\log(1/\delta)}{\varepsilon}\right)$ carries over to quantum
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\varepsilon}+\frac{\log(1/\delta)}{\varepsilon}\right)$
- AW'17: Showed $\Omega\left(\frac{d}{\varepsilon}+\frac{\log(1/\delta)}{\varepsilon}\right)$ quantum examples are necessary

  Proof idea: Reduce to state identification. For a good learner $P_{opt} \geq 2/3$, so $P_{pgm} \geq P_{opt}^2 \geq 4/9$. If $P_{pgm} \geq 4/9$, then $T = \Omega\left(\frac{d}{\varepsilon}\right)$

Quantum examples are no better than classical examples for PAC learning

Let's get real!

- In computational learning theory, agnostic learning and learning under classification noise is a theoretical way to model noise in data
- Again, in these realistic models we show that quantum sample complexity equals classical sample complexity

Future directions

Future directions

More mileage out of Fourier sampling?

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time?
(One of his ten semi-grand challenges for quantum computing!)

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time? A theoretical way to understand neural networks

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time? A theoretical way to understand neural networks
- Can we learn constant-depth quantum circuits?

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time?
  (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time?
  A theoretical way to understand neural networks
- Can we learn constant-depth quantum circuits?

More open questions!

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time? A theoretical way to understand neural networks
- Can we learn constant-depth quantum circuits?

More open questions!

- Can we learn the concept class of k-Fourier sparse Boolean functions using $O(k \log k)$ samples matching our lower bound?

Future directions

More mileage out of Fourier sampling?

- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time? A theoretical way to understand neural networks
- Can we learn constant-depth quantum circuits?

More open questions!

- Can we learn the concept class of k-Fourier sparse Boolean functions using $O(k \log k)$ samples matching our lower bound?
- Theoretically, one could consider more optimistic PAC-like models where learner need *not* succeed $\forall c \in \mathcal{C}$ and $\forall D$

Future directions

More mileage out of Fourier sampling?

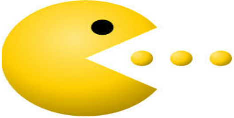
- Extend result of Bshouty-Jackson from depth-2 circuits (i.e., DNFs) to depth-3?
- Can we PAC-learn DNFs? If so, then we could possibly learn depth-3 circuits under the uniform distribution
- Scott Aaronson: Can AC^0 be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can TC^0 be learnt in quantum polynomial time? A theoretical way to understand neural networks
- Can we learn constant-depth quantum circuits?

More open questions!

- Can we learn the concept class of k-Fourier sparse Boolean functions using $O(k \log k)$ samples matching our lower bound?
- Theoretically, one could consider more optimistic PAC-like models where learner need *not* succeed $\forall c \in \mathcal{C}$ and $\forall D$
- Find more distributions (other than uniform) where quantum provides a speedup

Conclusion

For PAC learning, quantum examples are no better than classical examples

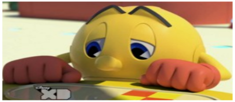


Classical PAC

=

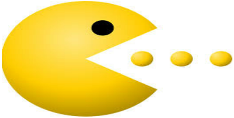
Quantum PAC

Sample complexity



Conclusion

For PAC learning, quantum examples are no better than classical examples

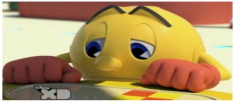


Classical PAC

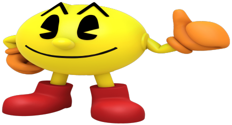
=

Quantum PAC

Sample complexity

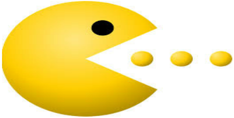


Under uniform D, quantum examples seem to help tremendously in some cases

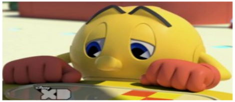


Conclusion

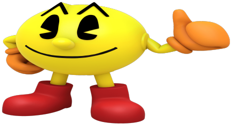
For PAC learning, quantum examples are no better than classical examples



Classical PAC = Quantum PAC Sample complexity



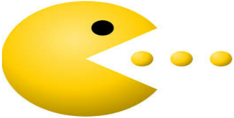
Under uniform D, quantum examples seem to help tremendously in some cases



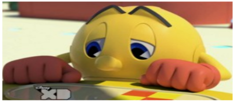
Quantum machine learning is still in its infancy! Not many strong examples where quantum significantly improves ML

Conclusion

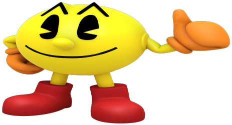
For PAC learning, quantum examples are no better than classical examples



Classical PAC
=
Quantum PAC
Sample complexity



Under uniform D, quantum examples seem to help tremendously in some cases



Quantum machine learning is still in its infancy! Not many strong examples where quantum significantly improves ML

Many recent surveys on quantum machine learning.