Strengths and weaknesses of quantum examples

Srinivasan Arunachalam (MIT)

joint with Ronald de Wolf (CWI, Amsterdam) and others
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 (Àmeur et al. '13)
- principal component analysis (Lloyd et al. '13)
- perceptron learning (Wiebe et al. '16)
- recommendation systems (Kerenidis & Prakash '16)
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

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 (A. ¨Ameur et al. '13)
- Principal component analysis (Lloyd et al. '13)
- perceptron learning (Wiebe et al. '16)
- recommendation systems (Kerenidis & Prakash '16)
## 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

- Quantum machine learning

  The learner will be quantum, the data may be quantum

  Some examples are known of reduction in time complexity:
  - clustering (A. }^{\ddagger}imeur et al. '13)
  - Principal component analysis (Lloyd et al. '13)
  - perceptron learning (Wiebe et al. '16)
  - recommendation systems (Kerenidis & Prakash '16)
**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?**
**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

**Basic definitions**

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

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

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 $x \sim D$.

\[
\begin{align*}
\mathcal{C} \\
\downarrow \\
\mathcal{C} \\
\text{target concept}
\end{align*}
\]

\[
x_1 \sim D \quad \longrightarrow \quad (x_1, c(x_1))
\]

*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 $x \sim D$

$Learner$ $is$ $trying$ $to$ $learn$ $c$
Classical learner using classical examples

Basic definitions

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

\[ C \]
\[ \downarrow \]
\[ \rightarrow \]
\[ C \]

```
x_1 \sim D
x_2 \sim D
\vdots
x_T \sim D
\rightarrow (x_1, c(x_1))
(x_2, c(x_2))
\vdots
(x_T, c(x_T))
```

Learner is trying to learn $c$
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.
Quantum learning using quantum examples

- Learner is quantum:

\[
\frac{\sqrt{2}}{\sqrt{2}} + \frac{\sqrt{2}}{\sqrt{2}} = \text{\figure}
\]
Quantum learning using quantum examples

- **Learner is quantum:**

  \[
  \frac{\sqrt{2}}{\sqrt{2}} + \frac{\sqrt{2}}{\sqrt{2}} = \text{Pacman}
  \]

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

  \[
  \sum_{x \in \{0,1\}^n} \sqrt{D(x)} |x, c(x)\rangle
  \]
Quantum learning using quantum examples

- **Learner is quantum:**

  $\frac{\sqrt{2}}{} + \frac{\sqrt{2}}{} = \equiv$

- **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

\[
\sum_{x \in \{0, 1\}^n} \sqrt{D(x)} \ket{x, c(x)}
\]

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 $C$, distribution $D : \{0, 1\}^n \rightarrow [0, 1]$
Motivating question for this talk

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

\[
\begin{align*}
    x_1 & \sim D \quad \rightarrow \quad (x_1, c(x_1)) \\
    x_2 & \sim D \quad \rightarrow \quad (x_2, c(x_2)) \\
    \vdots \\
    x_T & \sim D \quad \rightarrow \quad (x_T, c(x_T))
\end{align*}
\]
Motivating question for this talk

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

$x_1 \sim D \quad \rightarrow \quad (x_1, c(x_1))$
$x_2 \sim D \quad \rightarrow \quad (x_2, c(x_2))$
$\vdots$
$x_T \sim D \quad \rightarrow \quad (x_T, c(x_T))$

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

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

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

Question

Understanding the concept classes $C$ and distributions $D$ where fewer quantum examples suffice for a quantum learner
Focus on *Probably Approximately Correct* (PAC) model of learning.
Distribution (in)dependent PAC learning

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

Fix $C \subseteq \{c : \{0, 1\}^n \rightarrow \{0, 1\}\}$ and $D : \{0, 1\}^n \rightarrow [0, 1]$.

Using i.i.d. labeled examples, learner for $C$ should output hypothesis $h$ that is close to $c$ w.r.t. $D$. 
Focus on Probably Approximately Correct (PAC) model of learning

Fix $C \subseteq \{c : \{0, 1\}^n \rightarrow \{0, 1\}\}$ and $D : \{0, 1\}^n \rightarrow [0, 1]$

Using i.i.d. labeled examples, learner for $C$ 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 $C \subseteq \{c : \{0, 1\}^n \rightarrow \{0, 1\}\}$ and $D : \{0, 1\}^n \rightarrow [0, 1]$
- Using i.i.d. labeled examples, learner for $C$ 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 $(\varepsilon, \delta)$-learns $C$ under $D$ if:
  $$\forall c \in C : \Pr[err_D(c, h) \leq \varepsilon] \geq 1 - \delta$$

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

Fix \( C \subseteq \{ c : \{0, 1\}^n \rightarrow \{0, 1\} \} \) and \( D : \{0, 1\}^n \rightarrow [0, 1] \)

Using i.i.d. labeled examples, learner for \( C \) 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 \( (\varepsilon, \delta) \)-learns \( C \) under \( D \) if:

\[
\forall c \in C : \quad \Pr[err_D(c, h) \leq \varepsilon] \geq 1 - \delta
\]

PAC learning (Distribution-independent learning for every \( D \))
Focus on Probably Approximately Correct (PAC) model of learning

Fix $C \subseteq \{ c : \{0, 1\}^n \rightarrow \{0, 1\}\}$ and $D : \{0, 1\}^n \rightarrow [0, 1]$

Using i.i.d. labeled examples, learner for $C$ 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 $(\varepsilon, \delta)$-learns $C$ under $D$ if:

\[ \forall c \in C : \Pr[err_D(c, h) \leq \varepsilon] \geq 1 - \delta \]

### PAC learning (Distribution-independent learning for every $D$)

- An algorithm $(\varepsilon, \delta)$-PAC-learns $C$ if:

\[ \forall D \forall c \in C : \Pr[err_D(c, h) \leq \varepsilon] \geq 1 - \delta \]
How to measure the efficiency of the classical or quantum learner?
How to measure the efficiency of the classical or quantum learner?

- **Sample complexity**: number of labeled examples used by learner

Strengths of quantum examples:
- CLW’18: Sample complexity of learning Fourier-sparse Boolean functions under uniform
- DBshouty-Jackson’95: Quantum polynomial time learnability of DNFs under uniform
- CKW’18: Quantum examples can help the coupon collector

Weaknesses of quantum examples:
- W’17: Quantum examples are not more powerful than classical examples for PAC learning
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 $D$

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

$$\hat{c}(S) = \frac{1}{\sqrt{n}} \sum_{x \in \{0, 1\}} c(x) (-1)^{S \cdot x}$$

for all $S \in \{0, 1\}^n$.

Parseval's identity:

$$\sum_{S} \hat{c}(S)^2 = \mathbb{E}_{x}[c(x)^2] = 1$$

So $\{\hat{c}(S)^2\}_S$ forms a probability distribution.

Given quantum example under uniform $D$:

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

Measuring allows to sample from the Fourier distribution $\{\hat{c}(S)^2\}_S$. 
Fourier sampling: a useful trick under uniform $D$

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

- Parseval's identity: 
  
  $$\sum_S \hat{c}(S)^2 = E_x \left[c(x)^2\right] = 1$$

  So $\{\hat{c}(S)^2\}_S$ forms a probability distribution.

- Given quantum example under uniform $D$:
  
  $$\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle \xrightarrow{\text{Hadamard}} \sum_S \hat{c}(S) |S\rangle$$

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

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

$$\hat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0, 1\}^n$$
Fourier sampling: a useful trick under uniform $D$

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

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

Parseval's identity: $\sum_S \hat{c}(S)^2 =$
Let $c : \{0, 1\}^n \rightarrow \{-1, 1\}$. Then the Fourier coefficients are

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

Parseval's identity: $\sum_S \hat{c}(S)^2 = \mathbb{E}_x[c(x)^2]$
Fourier sampling: a useful trick under uniform $D$

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

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

- Parseval's identity: $\sum_S \hat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
Fourier sampling: a useful trick under uniform $D$

- Let $c : \{0, 1\}^n \to \{-1, 1\}$. Then the Fourier coefficients are
  \[
  \hat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0, 1\}^n
  \]

- Parseval's identity: $\sum_S \hat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\hat{c}(S)^2\}_S$ 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
  $$\hat{c}(S) = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} c(x)(-1)^{S \cdot x} \quad \text{for all } S \in \{0, 1\}^n$$

- Parseval's identity: $\sum_S \hat{c}(S)^2 = \mathbb{E}_x[c(x)^2] = 1$
  So $\{\hat{c}(S)^2\}_S$ 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 $c : \{0, 1\}^n \rightarrow \{-1, 1\}$. Then the Fourier coefficients are

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

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

  So $\{\hat{c}(S)^2\}_S$ forms a probability distribution

- Given quantum example under uniform $D$:

$$\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle \xrightarrow{\text{Hadamard}} \sum_S \hat{c}(S) |S\rangle$$
Fourier sampling: a useful trick under uniform $D$

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

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

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

So $\{\hat{c}(S)^2\}_S$ forms a probability distribution

Given quantum example under uniform $D$:

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

Measuring allows to sample from the Fourier distribution $\{\hat{c}(S)^2\}_S$
Applications of Fourier sampling

Consider the concept class of linear functions $C_1 = \{ c \mathbb{S}(x) = \mathbb{S} \cdot x \mid \mathbb{S} \in \{0, 1\}^n \}$

Classical: $\Omega(n)$ classical examples needed
Quantum: 1 quantum example suffices to learn (Bernstein-Vazirani'93)

Consider $C_2 = \{ c \text{ is a } \ell\text{-junta} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform
Quantum: $C_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n^2\ell + 2^2\ell)$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

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

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse

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

Observe that $C_1 \subseteq C$.
Observe that $C_2 \subseteq C$.

$C$ contains linear functions

$C$ contains $(\log k)$-juntas
Applications of Fourier sampling

Consider the concept class of linear functions $C_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$.
Consider the concept class of linear functions $C_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

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

Consider $C_2 = \{c_{\text{is a } \ell \text{-junta}}\}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

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

**Quantum:** $C_2$ can be exactly learnt using $\tilde{O}(2^{\ell})$ quantum examples and in time $\tilde{O}(n^2 \ell + 2^{2\ell})$ (Atıcı-Servedio'09)

Generalizing both these concept classes?

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

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse

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

Observe that $C_1 \subseteq C$.

Observe that $C_2 \subseteq C$.

Observe that $C$ contains linear functions
Applications of Fourier sampling

Consider the concept class of linear functions $C_1 = \{ c_S(x) = S \cdot x \}_{S \in \{0,1\}^n}$

- **Classical:** $\Omega(n)$ classical examples needed
- **Quantum:** 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

Consider $C_2 = \{ c_{\text{is a } \ell\text{-junta}} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

- Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform
- Quantum: $C_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n^2\ell + 2^{2\ell})$ (Atıcı-Servedio’09)

Generalizing both these concept classes?

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

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse.

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

Observe that $C_1 \subseteq C$.

Observe that $C_2 \subseteq C$.

$C$ contains (log $k$)-juntas.
Applications of Fourier sampling

- Consider the concept class of linear functions $C_1 = \{ c_S(x) = S \cdot x \} \forall S \in \{0, 1\}^n$
  - Classical: $\Omega(n)$ classical examples needed
  - Quantum: 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)
- Consider $C_2 = \{ c \text{ is a } \ell\text{-junta} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$
Consider the concept class of linear functions $C_1 = \{c_S(x) = S \cdot x \mid S \in \{0,1\}^n\}$.

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

**Quantum:** 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

Consider $C_2 = \{c \text{ is a } \ell\text{-junta}, \text{i.e., } c(x) \text{ depends only on } \ell \text{ bits of } x\}$

**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 $C_1 = \{ c_S(x) = S \cdot x \}_{S \in \{0,1\}^n}$
  
  **Classical:** $\Omega(n)$ classical examples needed
  
  **Quantum:** 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

- Consider $C_2 = \{ c \text{ is a } \ell\text{-junta} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$
  
  **Classical:** Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform $D$
  
  **Quantum:** $C_2$ can be *exactly* learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio’09)
Consider the concept class of linear functions $C_1 = \{ c_S(x) = S \cdot x \}_{S \in \{0,1\}^n}$

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

Quantum: 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

Consider $C_2 = \{ c \text{ is a } \ell\text{-junta} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

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

Quantum: $C_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio’09)

Generalizing both these concept classes?

Definition: We say $c$ is $k$-Fourier sparse if $|\{ S : \hat{c}(S) \neq 0 \}| \leq k$. 
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 $\mathcal{C}_2 = \{ c \text{ is a } \ell\text{-junta}, \text{ i.e., } c(x) \text{ depends only on } \ell \text{ bits of } x \}$
  - Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform $D$
  - Quantum: $\mathcal{C}_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n2^\ell + 2^{2^\ell})$ (Atıcı-Servedio’09)

**Generalizing both these concept classes?**

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

**Note that** $\mathcal{C}_1$ **is 1-Fourier sparse and** $\mathcal{C}_2$ **is** $2^\ell$-Fourier sparse.
Applications of Fourier sampling

- Consider the concept class of linear functions $C_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$
  - Classical: $\Omega(n)$ classical examples needed
  - Quantum: 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

- Consider $C_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$
  - Classical: Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform $D$
  - Quantum: $C_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio’09)

Generalizing both these concept classes?

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

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse

Consider the concept class $C = \{c : \{0,1\}^n \rightarrow \{-1,1\} : c \text{ is } k\text{-Fourier sparse}\}$
Applications of Fourier sampling

Consider the concept class of linear functions $C_1 = \{ c_S(x) = S \cdot x \}_{S \in \{0,1\}^n}$

- **Classical:** $\Omega(n)$ classical examples needed
- **Quantum:** 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

Consider $C_2 = \{ c \text{ is a } \ell\text{-junta} \}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

- **Classical:** Efficient learning is notoriously hard for $\ell = O(\log n)$ and uniform $D$
- **Quantum:** $C_2$ can be exactly learnt using $\tilde{O}(2^\ell)$ quantum examples and in time $\tilde{O}(n2^\ell + 2^{2\ell})$ (Atıcı-Servedio’09)

Generalizing both these concept classes?

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

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse

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

Observe that $C_1 \subseteq C$. $C$ contains linear functions
Applications of Fourier sampling

Consider the concept class of linear functions $C_1 = \{c_S(x) = S \cdot x\}_{S \in \{0,1\}^n}$

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

**Quantum:** 1 quantum example suffices to learn $C_1$ (Bernstein-Vazirani’93)

Consider $C_2 = \{c \text{ is a } \ell\text{-junta}\}$, i.e., $c(x)$ depends only on $\ell$ bits of $x$

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

**Quantum:** $C_2$ 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 $|\{S : \hat{c}(S) \neq 0\}| \leq k$.

Note that $C_1$ is 1-Fourier sparse and $C_2$ is $2^\ell$-Fourier sparse

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

Observe that $C_1 \subseteq C$. $C$ contains linear functions

Observe that $C_2 \subseteq C$. $C$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$

Sketch of upper bound

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

Collect $S$s until the learner learns the Fourier span of $c$, $V = \text{span}\{S: \hat{c}(S) \neq 0\}$

Suppose $\dim(V) = r$, then $\tilde{O}(rk)$ quantum examples suffice to find $V$

Use the result of [HR’15] to learn $c'$ completely using $\tilde{O}(rk)$ classical examples

Since $r \leq \tilde{O}(\sqrt{k})$ for every $c \in \mathcal{C}$ [Sanyal’15], we get $\tilde{O}(k^{1.5})$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2n}} \sum_x |x, c(x)\rangle$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$)
  $\tilde{\Omega}(k)$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW’18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$)

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

Sketch of upper bound

Use Fourier sampling to sample $S \sim \mathcal{E}c(S)^2$

Collect $S$s until the learner learns the Fourier span of $c$, $V = \text{span}\{S : \hat{c}(S) \neq 0\}$

Suppose $\text{dim}(V) = r$, then $\tilde{O}(rk)$ quantum examples suffice to find $V$

Use the result of [HR’15] to learn $c'$ completely using $\tilde{O}(rk)$ classical examples

Since $r \leq \tilde{O}(\sqrt{k})$ for every $c \in \mathcal{C}$ [Sanyal’15], we get $\tilde{O}(k^{1.5})$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$)
  $\tilde{\Omega}(k)$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
  - Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$)
  - $\tilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

**Sketch of upper bound**

- Use Fourier sampling to sample $S \sim \{\hat{c}(S)^2\}_S$
- Collect $S$s until the learner learns the Fourier span of $c$, $\mathcal{V} = \text{span}\{S : \hat{c}(S) \neq 0\}$
Learning $\mathcal{C} = \{c \text{ is } k\text{-Fourier sparse}\}$

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

**Sketch of upper bound**

- Use Fourier sampling to sample $S \sim \{\hat{c}(S)^2\}_S$
- Collect $S$s until the learner learns the Fourier span of $c$, $\mathcal{V} = \text{span}\{S : \hat{c}(S) \neq 0\}$
- Suppose $\dim(\mathcal{V}) = r$, then $\tilde{O}(rk)$ 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 $\mathcal{D}$
- Classically (Haviv-Regev'15): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$).
  $\tilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

**Sketch of upper bound**

- Use Fourier sampling to sample $S \sim \{\hat{c}(S)^2\}_S$
- Collect $S$s until the learner learns the Fourier span of $c$, $\mathcal{V} = \text{span}\{S : \hat{c}(S) \neq 0\}$
- Suppose $\text{dim}(\mathcal{V}) = r$, then $\tilde{O}(rk)$ quantum examples suffice to find $\mathcal{V}$
- Use the result of [HR’15] to learn $c'$ completely using $\tilde{O}(rk)$ 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): $\tilde{\Theta}(nk)$ classical examples $(x, c(x))$ are necessary and sufficient to learn the concept class $\mathcal{C}$
- Quantumly (ACLW'18): $\tilde{O}(k^{1.5})$ quantum examples $\frac{1}{\sqrt{2^n}} \sum_x |x, c(x)\rangle$ are sufficient to learn $\mathcal{C}$ (independent of the universe size $n$)
  $\tilde{\Omega}(k)$ examples are necessary to learn $\mathcal{C}$

**Sketch of upper bound**

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

Simply an OR of AND of variables. For example, 

\[(x_1 \land x_4 \land x_3) \lor (x_4 \land x_6 \land x_7 \land x_8)\]

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

Learning \(C = \{c\text{ is an }s\text{-term DNF in }n\text{ variables}\}\) 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. 

\[|\hat{c}(U)| \geq 1\]

Fourier sampling! Sample \(T \sim \{\hat{c}(T)\}_{2}^{poly(s)}\) many times to see such a \(U\).

Construct a "weak learner" who outputs \(\chi_U\) s.t. 

\[\Pr[\chi_U(x) = c(x)] = \frac{1}{2^{s}} + \frac{1}{s}\]

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 

\[\Pr[h(x) = c(x)] \geq \frac{2}{3}\]
DNFs

Simply an OR of AND of variables.
Learning Disjunctive normal Forms (DNF)

<table>
<thead>
<tr>
<th>DNFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simply an <strong>OR of AND</strong> of variables. For example, ((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8}))</td>
</tr>
</tbody>
</table>
DNFs

Simply an **OR of AND** of variables. For example, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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

### Learning

**Learning \(C = \{c \text{ is an } s\text{-term DNF in } n \text{ variables}\} \text{ under uniform } D\)**
Learning Disjunctive normal Forms (DNF)

DNFs

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

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]
DNFs
Simply an OR of AND of variables. For example, \((x_1 \land x_4 \land \overline{x}_3) \lor (\overline{x}_4 \land x_6 \land x_7 \land \overline{x}_8)\)
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!
DNFs

Simply an OR of AND of variables. For example, \((x_1 \land x_4 \land \overline{x}_3) \lor (\overline{x}_4 \land x_6 \land x_7 \land \overline{x}_8)\)

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, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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. \(|\hat{c}(U)| \geq \frac{1}{s}\)
Learning Disjunctive normal Forms (DNF)

**DNFs**

Simply an **OR of AND of variables**. For example, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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. \(|\hat{c}(U)| \geq \frac{1}{s}\)
- **Fourier sampling**: Sample \(T \sim \{\hat{c}(T)^2\}_T\), poly\((s)\) many times to see such a \(U\)
Learning Disjunctive normal Forms (DNF)

**DNFs**

Simply an **OR of AND** of variables. For example, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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}\} \text{ 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. \(|\hat{c}(U)| \geq \frac{1}{s}\)
- **Fourier sampling**! Sample \(T \sim \{\hat{c}(T)^2\}_T\), \(\text{poly}(s)\) many times to see such a \(U\)
- Construct a “**weak learner**” who outputs \(\chi_U\) 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, \((x_1 \wedge x_4 \wedge \overline{x}_3) \lor (\overline{x}_4 \wedge x_6 \wedge x_7 \wedge \overline{x}_8)\)
We say a DNF on \(n\) variables is an \(s\)-term DNF if number of clauses is \(\leq s\)

Learning \(C = \{c\text{ is an }s\text{-term DNF in }n\text{ variables}\}\) 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. \(|\hat{c}(U)| \geq \frac{1}{s}\)
- Fourier sampling! Sample \(T \sim \{\hat{c}(T)^2\}_T\), \(\text{poly}(s)\) many times to see such a \(U\)
- Construct a “weak learner” who outputs \(\chi_U\) 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, \((x_1 \land x_4 \land \overline{x_3}) \lor (\overline{x_4} \land x_6 \land x_7 \land \overline{x_8})\)

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}\}_{\text{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. \(|\hat{c}(U)| \geq \frac{1}{s}\)
- Fourier sampling! Sample \(T \sim \{\hat{c}(T)^2\}_T\), \(\text{poly}(s)\) many times to see such a \(U\)
- Construct a “weak learner” who outputs \(\chi_U\) 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

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

State identification: For uniform $c \in C$ (unknown), given $|\psi_c⟩ \otimes T$, 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 $P_{opt} \geq P_{pgm} \geq P_{2opt}$ (Barnum-Knill'02).
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ be a distribution.
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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 $P_{\text{opt}}$ is the success probability of the optimal measurement, $P_{\text{pgm}}$ is the success probability of the PGM, then $P_{\text{opt}} \geq P_{\text{pgm}} \geq P_{\text{2opt}}$ (Barnum-Knill'02).
Consider a concept class $C$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ be a distribution.

For $c \in C$, a quantum example is
$$|\psi_c\rangle = \sum_{x \in \{0,1\}^n} \sqrt{D(x)} |x, c(x)\rangle$$

State identification: For uniform $c \in C$ (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 $P_{\text{opt}}$ is the success probability of the optimal measurement, $P_{\text{pgm}}$ is the success probability of the PGM, then
$$P_{\text{opt}} \geq P_{\text{pgm}} \geq P_{\text{opt}}^2 \quad \text{(Barnum-Knill'02)}$$
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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,
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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).
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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,
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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,
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
\[
P_{opt} \geq P_{pgm}
\]
Consider a concept class $\mathcal{C}$ consisting of $n$-bit Boolean functions. Let $D : \{0, 1\}^n \rightarrow [0, 1]$ 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

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 $i^\ast \in \{1, \ldots, N\}$. How many coupons to draw (with replacement) from $\{1, \ldots, N\} \setminus \{i^\ast\}$ before learning $i^\ast$?

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

What if we are given "quantum examples"?

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

How many quantum examples before learning $i^\ast$?

Answer $[\text{ACKW'..}]:$ Can learn $i^\ast$ using $\Theta(N)$ quantum examples.

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

**Standard coupon collector**

**Problem:** Suppose there are $N$ coupons.

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

**Variation to coupon collector**

Problem: Suppose there are $N$ coupons. Fix unknown $i^* \in \{1,...,N\}$. How many coupons to draw (with replacement) from $\{1,...,N\}\{-i^*\}$ 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 $\frac{1}{\sqrt{N}} - \sum_{i \in (\{1,...,N\}\{-i^*\})} |i\rangle$.

How many quantum examples before learning $i^*$?

Answer: $\Theta(N)$ quantum examples

Proof idea: Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{opt} \geq P_{pgm} \geq \frac{2}{3}$.
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 $\frac{1}{\sqrt{N}} - \sum_{i \in (\{1, \ldots, N\} \setminus \{i^*\})} |i\rangle$.

How many quantum examples before learning $i^*$?

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

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

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 [ACKW’..]:** Can learn $i^*$ using $\Theta(N)$ quantum examples

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

If $T = O(N)$, then $P_{opt} \geq P_{pgm} \geq \frac{2}{3}$
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.

What if we are given "quantum examples"? Suppose a quantum learner obtains quantum examples $\frac{\sqrt{N} - 1}{\sum_{i \in \{1, \ldots, N\} \setminus \{i^*\}} |i\rangle}$. How many quantum examples before learning $i^*$?

Answer: Can learn $i^*$ using $\Theta(N)$ quantum examples

Proof idea: Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{opt} \geq P_{pgm} \geq \frac{2}{3}$
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 $i^* \in \{1, \ldots, N\}$.

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

What if we are given "quantum examples"? Suppose a quantum learner obtains quantum examples $\sqrt{N}^{-1} \sum_{i \in (\{1, \ldots, N\} \setminus \{i^*\})} |i\rangle$.

How many quantum examples before learning $i^*$?

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

**Proof idea:** Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{opt} \geq P_{pgm} \geq \frac{2}{3}$
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 $\{1, \ldots, N\}\setminus\{i^*\}$

Answer [ACKW’..]: Can learn $i^*$ using $\Theta(N)$ quantum examples

Proof idea: Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{\text{opt}} \geq P_{\text{pgm}} \geq \frac{2}{3}$
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 \( i^* \)?
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 $\Theta(N \log N)$
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 $\{1, \ldots, N\}\setminus\{i^*\}$ 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 [ACKW’..]:** Can learn $i^*$ using $\Theta(N)$ quantum examples

**Proof idea:** Analyze the success probability using the pretty good measurement. If $T = O(N)$, then $P_{\text{opt}} \geq P_{\text{pgm}} \geq \frac{2}{3}$
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 \( i^* \)?
### 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 [ACKW'..]:** 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 $\{1, \ldots, N\}\backslash\{i^*\}$ 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 [ACKW’..]: 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 \( \{1, \ldots, N\} \setminus \{i^*\} \) 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** [ACKW’..]: Can learn \( i^* \) using \( \Theta(N) \) quantum examples

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

If \( T = O(N) \), then \( P_{opt} \geq P_{pgm} \geq 2/3 \)
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\epsilon, \delta)-\text{PAC-learns} C\) if:

\[
\forall D \forall c \in C: \Pr[\text{err}_D(c, h) \leq \epsilon] \geq 1 - \delta.
\]

Approximately Correct

Probably

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called VC-dimension of \(C\).

Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\).

Fundamental theorem of PAC learning

Suppose VC-dim\((C) = d\): Blumer-Ehrenfeucht-Haussler-Warmuth'86:

\[
\text{every } (\epsilon, \delta)-\text{PAC learner for } C \text{ needs } \Omega\left(d \epsilon + \log\left(\frac{1}{\delta} \epsilon\right)\right) \text{ examples}
\]

Hanneke'16: exists an \((\epsilon, \delta)-\text{PAC learner for } C\) using \(O\left(d \epsilon + \log\left(\frac{1}{\delta} \epsilon\right)\right) \text{ examples}
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \quad \forall c \in C \quad : \quad \Pr[\text{err}_D(c, h) \leq \varepsilon] \geq 1 - \delta
\]

Approximately Correct \hspace{1cm} Probably
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \forall c \in C : \Pr[ err_D(c, h) \leq \varepsilon ] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \forall c \in C : \Pr[\text{err}_D(c, h) \leq \varepsilon] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called \(VC\)-dimension of \(C\).
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \forall c \in C : \Pr\left[ \frac{\text{err}_D(c, h)}{\varepsilon} \right] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called \textit{VC-dimension} of \(C\).

Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\)
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \forall c \in C : \Pr[ err_D(c, h) \leq \varepsilon ] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called VC-dimension of \(C\).

Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\)

Fundamental theorem of PAC learning
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \ \forall c \in C \ : \ \Pr[\text{err}_D(c, h) \leq \varepsilon] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called VC-dimension of \(C\). Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\).

Fundamental theorem of PAC learning

Suppose \(VC\text{-dim}(C) = d\)
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\epsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \quad \forall c \in C \quad \Pr[\err_D(c, h) \leq \epsilon] \geq 1 - \delta
\]

Approximately Correct Probably

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called VC-dimension of \(C\). Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\).

Fundamental theorem of PAC learning

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

- Blumer-Ehrenfeucht-Haussler-Warmuth'86: every \((\epsilon, \delta)\)-PAC learner for \(C\) needs \(\Omega\left(\frac{d}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right)\) examples
Recall: PAC learning

Given \((x, c(x))\) examples where \(x \sim D\), a learner \((\varepsilon, \delta)\)-PAC-learns \(C\) if:

\[
\forall D \forall c \in C : \Pr[err_D(c, h) \leq \varepsilon \geq 1 - \delta] \geq 1 - \delta
\]

Complexity measure: Number of labelled examples

For a concept class \(C\), associate a combinatorial parameter called VC-dimension of \(C\). Classical PAC learning sample complexity is characterized by the VC-dimension of \(C\)

Fundamental theorem of PAC learning

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

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

- Hanneke'16: exists an \((\varepsilon, \delta)\)-PAC learner for \(C\) using \(O \left( \frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon} \right)\) examples
VC-dimension and quantum sample complexity

Quantum bounds

Classical upper bound

\[ O(d^2 \varepsilon + \log(1/\delta)) \]

carries over to quantum

Atıcı-Servedio'04: lower bound \( \Omega\left(\sqrt{d^2 \varepsilon + \log(1/\delta)}\right) \)

AW'17: Showed \( \Omega\left(d^2 \varepsilon + \log(1/\delta)\right) \) quantum examples are necessary

Proof idea: Reduce to state identification.

For a good learner

\[ P_{opt} \geq 2/3, \text{ so } P_{gpm} \geq P_{opt} \geq 4/9. \]

If \( P_{gpm} \geq 4/9 \), then \( T = \Omega(d^2 \varepsilon) \)

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.
Quantum bounds

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

---

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.
Quantum bounds

- Classical upper bound $O\left(\frac{d}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right)$ carries over to quantum.
- Atıcı-Servedio'04: lower bound $\Omega\left(\frac{\sqrt{d}}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right)$
- AW’17: Showed $\Omega\left(\frac{d}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right)$ quantum examples are necessary.
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 $P_{\text{opt}} \geq 2/3$, so $P_{\text{pgm}} \geq P_{\text{opt}}^2 \geq 4/9$. 

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.
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

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 $\text{AC}^0$ be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
Can $\text{TC}^0$ 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 $\forall c \in C$ and $\forall D$.
Find more distributions (other than uniform) where quantum provides a speedup.
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 $\forall c \in C$ and $\forall D$.

Find more distributions (other than uniform) where quantum provides a speedup.
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 \( \text{AC}^0 \) be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)

Can \( \text{TC}^0 \) 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 \( \forall c \in C \) and \( \forall D \)

Find more distributions (other than uniform) where quantum provides a speedup.
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 \( \text{AC}^0 \) be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)

Can \( \text{TC}^0 \) 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 \( \forall c \in C \) and \( \forall D \)

Find more distributions (other than uniform) where quantum provides a speedup
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 $\text{AC}^0$ 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?
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 $\text{AC}^0$ be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can $\text{TC}^0$ 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 $\text{AC}^0$ be learnt in quantum polynomial time? (One of his ten semi-grand challenges for quantum computing!)
- Can $\text{TC}^0$ 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 $\forall D$. 

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 $\forall c \in C$ and $\forall D$.
- Find more distributions (other than uniform) where quantum provides a speedup.
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.

\[ \text{Classical PAC} = \text{Quantum PAC} \]

Sample complexity

Under uniform $D$, quantum examples seem to help tremendously in some cases.
For PAC learning, quantum examples are no better than classical examples

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