Optimal Bounds for Quantum Learning via Information Theory

Pulkit Sinha
joint work with Ashwin Nayak, Shima Bab Hadiashar

July 2023

Contents

Introduction

Introduction: Quantum Learning Theory

Introduction: Quantum Learning Theory

- Classical learning problems:

  Given iid copies of an unknown random variable X, tell me something about X

Introduction: Quantum Learning Theory

- Classical learning problems:

  Given iid copies of an unknown random variable X, tell me something about X

- Examples:
  1. Finding mean, variance etc.

Introduction: Quantum Learning Theory

- Classical learning problems:

  Given iid copies of an unknown random variable X, tell me something about X

- Examples:
  1. Finding mean, variance etc.
  2. Finding any model for prediction (training data = samples)

Introduction: Quantum Learning Theory

- Classical learning problems:

  Given iid copies of an unknown random variable X, tell me something about X

- Examples:
  1. Finding mean, variance etc.
  2. Finding any model for prediction (training data = samples)

- *Sample Complexity*: number of copies/samples of X needed

Introduction: Quantum Learning Theory

- Classical learning problems:

  Given iid copies of an unknown random variable X, tell me something about X
- Examples:
  1. Finding mean, variance etc.
  2. Finding any model for prediction (training data = samples)
- *Sample Complexity*: number of copies/samples of X needed
  For instance: $\Theta(\frac{1}{\epsilon^2})$ samples for finding mean upto ϵ error

Introduction: Quantum Learning Theory

Introduction: Quantum Learning Theory

- Quantum Analogue:

  Given iid copies of a *Quantum State* ρ, tell me something about ρ.

Introduction: Quantum Learning Theory

- Quantum Analogue:

  Given iid copies of a *Quantum State* ρ, tell me something about ρ.

- Example: State tomography: Give a $\rho' \approx_\epsilon \rho$. Needs $\Theta(\frac{d^2}{\epsilon^2})$ samples[HHJ$^+$17].

Introduction: Quantum Learning Theory

- Quantum Analogue:

  Given iid copies of a *Quantum State* ρ, tell me something about ρ.

- Example: State tomography: Give a $\rho' \approx_\epsilon \rho$. Needs $\Theta(\frac{d^2}{\epsilon^2})$ samples[HHJ+17].

- Freedom to choose measurements, including entangled measurements.

An aside: Quantum Machine Learning

An aside: Quantum Machine Learning

This is NOT QML

An aside: Quantum Machine Learning

This is NOT QML:

- QML looks at computational advantage to solve classical problems

An aside: Quantum Machine Learning

This is NOT QML:

- QML looks at computational advantage to solve classical problems
- We look at the predictive power of the states themselves.

Our work

Our work

- We show lower bounds for the following problems:

Our work

- We show lower bounds for the following problems:
  1. Quantum PAC Learning
  2. Quantum Agnostic Learning
  3. Quantum Coupon Collector

Our work

- We show lower bounds for the following problems:
  1. Quantum PAC Learning
  2. Quantum Agnostic Learning
  3. Quantum Coupon Collector

- 1 and 2 shown by Arunachalem and de Wolfe[AdW18]

Our work

- We show lower bounds for the following problems:
  1. Quantum PAC Learning
  2. Quantum Agnostic Learning
  3. Quantum Coupon Collector
- 1 and 2 shown by Arunachalem and de Wolfe[AdW18]
- 3 shown by Arunachalam, Belovs, Childs, Kothari, Rosmanis, and de Wolf [ABC$^+$20]

Our work

- We show lower bounds for the following problems:
  1. Quantum PAC Learning
  2. Quantum Agnostic Learning
  3. Quantum Coupon Collector
- 1 and 2 shown by Arunachalem and de Wolfe[AdW18]
- 3 shown by Arunachalam, Belovs, Childs, Kothari, Rosmanis, and de Wolf [ABC$^+$20]
- Found lower bound with optimal leading term for QCC.

Quantum PAC Learning

PAC learning

PAC learning

- PAC learning[Val84] : (Informal)

Given an unknown binary function $f \in \mathcal{F}$ and random samples of f over any distribution of its domain, find an ϵ approximation of f over the same distribution.

PAC learning

- PAC learning[Val84] : (Informal)

Given an unknown binary function $f \in \mathcal{F}$ and random samples of f over any distribution of its domain, find an ϵ approximation of f over the same distribution.

Random sample: $(x, f(x))$ with probability p_x.

PAC learning

- PAC learning[Val84] : (Informal)

  Given an unknown binary function $f \in \mathcal{F}$ and random samples of f over any distribution of its domain, find an ϵ approximation of f over the same distribution.

  Random sample: $(x, f(x))$ with probability p_x.

- Well studied classically[Han16, BEHW89], sample complexity is $\Theta\left(\frac{d}{\epsilon}\right)$, d is the *VC-dimension* of $\mathcal{F}$.

PAC learning

- PAC learning[Val84] : (Informal)

  Given an unknown binary function $f \in \mathcal{F}$ and random samples of f over any distribution of its domain, find an ϵ approximation of f over the same distribution.

  Random sample: $(x, f(x))$ with probability p_x.

- Well studied classically[Han16, BEHW89], sample complexity is $\Theta\left(\frac{d}{\epsilon}\right)$, d is the *VC-dimension* of $\mathcal{F}$.

Quantum PAC learning

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x}\,|x\rangle\,|f(x)\rangle$$

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x}\,|x\rangle\,|f(x)\rangle$$

Atleast as powerful as random samples.

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x} \, |x\rangle \, |f(x)\rangle$$

Atleast as powerful as random samples.

Theorem (Quantum PAC learning lower bound(Simplified))

Sample complexity for Quantum PAC learning is $\Omega(\frac{d}{\epsilon})$

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x} \, |x\rangle \, |f(x)\rangle$$

Atleast as powerful as random samples.

Theorem (Quantum PAC learning lower bound(Simplified))

Sample complexity for Quantum PAC learning is $\Omega(\frac{d}{\epsilon})$

For simplicity:

$$\mathcal{F} = \{f : \mathbb{N}_{d+1} \rightarrow \{0, 1\}, f(0) = 0\}$$

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x}\,|x\rangle\,|f(x)\rangle$$

Atleast as powerful as random samples.

Theorem (Quantum PAC learning lower bound(Simplified))

Sample complexity for Quantum PAC learning is $\Omega(\frac{d}{\epsilon})$

For simplicity:

$$\mathcal{F} = \{f : \mathbb{N}_{d+1} \to \{0,1\}, f(0) = 0\}$$

indexed by $\mathbf{a} \in \{0,1\}^d$ with $f(i) = \mathbf{a}_i$.

Quantum PAC learning

Quantum PAC learning: use *Quantum Samples* instead [BJ98]

$$|\psi_f\rangle = \sum \sqrt{p_x} |x\rangle |f(x)\rangle$$

Atleast as powerful as random samples.

Theorem (Quantum PAC learning lower bound(Simplified))

Sample complexity for Quantum PAC learning is $\Omega(\frac{d}{\epsilon})$

For simplicity:

$$\mathcal{F} = \{f : \mathbb{N}_{d+1} \to \{0,1\}, f(0) = 0\}$$

indexed by $\mathbf{a} \in \{0,1\}^d$ with $f(i) = \mathbf{a}_i$.

$$p_0 = 1 - 4\epsilon, p_i = \frac{4\epsilon}{d} \forall i \in [d]$$

Data Processing inequality

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A : B) \geq I(A : \chi(B))$$

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A : B) \geq I(A : \chi(B))$$

In our setting,

1. A holds index **a** for Quantum Samples

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A : B) \geq I(A : \chi(B))$$

In our setting,

1. A holds index **a** for Quantum Samples
2. B holding t copies of corresponding quantum samples

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A : B) \geq I(A : \chi(B))$$

In our setting,

1. A holds index $\mathbf{a}$ for Quantum Samples
2. B holding t copies of corresponding quantum samples
3. χ is *Learner*, for learning A from B.

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A:B) \geq I(A:\chi(B))$$

In our setting,

1. A holds index **a** for Quantum Samples
2. B holding t copies of corresponding quantum samples
3. χ is *Learner*, for learning A from B.

$I(A:\chi(B))$ is related to correctness, while $I(A:B)$ is increasing in the number of samples.

Data Processing inequality

Theorem (Data processing inequality)

Given a state ρ in Quantum Registers A, B, and any channel χ for B, we have:

$$I(A:B) \geq I(A:\chi(B))$$

In our setting,

1. A holds index **a** for Quantum Samples
2. B holding t copies of corresponding quantum samples
3. χ is *Learner*, for learning A from B.

$I(A:\chi(B))$ is related to correctness, while $I(A:B)$ is increasing in the number of samples.

$I(A:\chi(B))$ classical, need upper bound on $I(A:B)$.

Proof Outline

Proof Outline

- Fix a distribution of $\mathbf{a}$

Proof Outline

- Fix a distribution of **a**
- Find spectrum of ρ_B

Proof Outline

- Fix a distribution of **a**
- Find spectrum of ρ_B
- Find expression for $S(B)$

Proof Outline

- Fix a distribution of **a**
- Find spectrum of ρ_B
- Find expression for $S(B)$
- Upper bound $S(B)$ by treating $S(B)$ as expectation.

Proof Outline

- Fix a distribution of **a**
- Find spectrum of ρ_B
- Find expression for $S(B)$
- Upper bound $S(B)$ by treating $S(B)$ as expectation.

Spectrum of ρ_B

Spectrum of ρ_B

$$\rho_B = \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d} |\psi_{\mathbf{a}}\rangle\langle\psi_{\mathbf{a}}|^{\otimes t}$$

Spectrum of ρ_B

$$\rho_B = \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d} |\psi_{\mathbf{a}}\rangle\langle\psi_{\mathbf{a}}|^{\otimes t}$$

$$= \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d, \mathbf{x}, \mathbf{y} \in \mathbb{N}^t_{d+1}} \sqrt{p_{\mathbf{x}} p_{\mathbf{y}}}\, |\mathbf{x}\rangle\langle\mathbf{y}| \otimes |\mathbf{a}_{\mathbf{x}}\rangle\langle\mathbf{a}_{\mathbf{y}}|$$

Spectrum of ρ_B

$$\rho_B = \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d} |\psi_{\mathbf{a}}\rangle\langle\psi_{\mathbf{a}}|^{\otimes t}$$

$$= \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d, \mathbf{x}, \mathbf{y} \in \mathbb{N}^t_{d+1}} \sqrt{p_{\mathbf{x}} p_{\mathbf{y}}}\, |\mathbf{x}\rangle\langle\mathbf{y}| \otimes |\mathbf{a}_{\mathbf{x}}\rangle\langle\mathbf{a}_{\mathbf{y}}|$$

Block diagonalizes after applying $I \otimes H^{\otimes t}$

Spectrum of ρ_B

$$\rho_B = \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d} |\psi_{\mathbf{a}}\rangle\langle\psi_{\mathbf{a}}|^{\otimes t}$$

$$= \frac{1}{2^d} \sum_{\mathbf{a} \in \{0,1\}^d, \mathbf{x}, \mathbf{y} \in \mathbb{N}^t_{d+1}} \sqrt{p_{\mathbf{x}} p_{\mathbf{y}}}\, |\mathbf{x}\rangle\langle\mathbf{y}| \otimes |\mathbf{a}_{\mathbf{x}}\rangle\langle\mathbf{a}_{\mathbf{y}}|$$

Block diagonalizes after applying $I \otimes H^{\otimes t}$

Block for $\mathbf{b} \in \{0,1\}^d$ has eigenvalue

$$\lambda_{\mathbf{b}} = \frac{1}{2^t} \sum_{\substack{\mathbf{c} \in \{0,1\}^t \\ \mathbf{x} \in \mathbb{N}^t_{d+1} \\ \mathrm{ps}(\mathbf{x}_{\mathbf{c}}) = \mathbf{b}}} p_{\mathbf{x}}$$

Spectrum of ρ_B

$$\rho_B = \frac{1}{2^d} \sum_{\mathbf{a}\in\{0,1\}^d} |\psi_{\mathbf{a}}\rangle\langle\psi_{\mathbf{a}}|^{\otimes t}$$
$$= \frac{1}{2^d} \sum_{\mathbf{a}\in\{0,1\}^d, \mathbf{x},\mathbf{y}\in\mathbb{N}^t_{d+1}} \sqrt{p_{\mathbf{x}} p_{\mathbf{y}}}\, |\mathbf{x}\rangle\langle\mathbf{y}| \otimes |\mathbf{a}_{\mathbf{x}}\rangle\langle\mathbf{a}_{\mathbf{y}}|$$

Block diagonalizes after applying $I \otimes H^{\otimes t}$

Block for $\mathbf{b} \in \{0,1\}^d$ has eigenvalue

$$\lambda_{\mathbf{b}} = \frac{1}{2^t} \sum_{\substack{\mathbf{c}\in\{0,1\}^t \\ \mathbf{x}\in\mathbb{N}^t_{d+1} \\ \mathrm{ps}(\mathbf{x}_{\mathbf{c}})=\mathbf{b}}} p_{\mathbf{x}}$$

$\lambda_{\mathbf{b}}$ is function of $|b|$, mutliplicity is $l_{|b|} = \binom{d}{|b|}$

Expression for $S(B)$

Expression for $S(B)$

With the above, $S(B)$ looks like

Expression for $S(B)$

With the above, $S(B)$ looks like

$$S(B) = \sum_{i=0}^{d} l_i \lambda_i \log \frac{1}{\lambda_i} = \sum_{i=0}^{d} (l_i \lambda_i) \cdot [\log l_i] + \quad O(\log(d+1))$$

Expression for $S(B)$

With the above, $S(B)$ looks like

$$S(B) = \sum_{i=0}^{d} l_i \lambda_i \log \frac{1}{\lambda_i} = \sum_{i=0}^{d} (l_i \lambda_i) \cdot [\log l_i] + \quad O(\log(d+1))$$

$$= \sum_{h=0}^{d} \left(\frac{1}{2^t} \binom{d}{h} \sum_{\substack{\mathbf{c} \in \{0,1\}^t \\ \mathbf{x} \in \mathbb{N}^t_{d+1} \\ |\mathrm{ps}(\mathbf{x_c})| = h}} p_{\mathbf{x}} \right) \log \binom{d}{h} + O(\log(d+1))$$

Expression for $S(B)$

With the above, $S(B)$ looks like

$$S(B) = \sum_{i=0}^{d} l_i \lambda_i \log \frac{1}{\lambda_i} = \sum_{i=0}^{d} (l_i \lambda_i) \cdot [\log l_i] + \quad O(\log(d+1))$$

$$= \sum_{h=0}^{d} \left(\frac{1}{2^t} \binom{d}{h} \sum_{\substack{\mathbf{c} \in \{0,1\}^t \\ \mathbf{x} \in \mathbb{N}_{d+1}^t \\ |\mathrm{ps}(\mathbf{x_c})| = h}} p_{\mathbf{x}} \right) \log \binom{d}{h} + O(\log(d+1))$$

Suffices to show concentration of h away from $\frac{d}{2}$.

Quantum Coupon Collector

Coupon Collector Problem

Coupon Collector Problem

Simple definition:

Given random coupon out of k at each step, how many steps until each collected atleast once?

Coupon Collector Problem

Simple definition:

Given random coupon out of k at each step, how many steps until each collected atleast once?

Becomes harder to collect "new" coupons.

Coupon Collector Problem

Simple definition:

Given random coupon out of k at each step, how many steps until each collected atleast once?

Becomes harder to collect "new" coupons.

The answer is $k \ln k + \Theta(k)$, and has a very simple analysis.

Coupon Collector Problem

Simple definition:

Given random coupon out of k at each step, how many steps until each collected atleast once?

Becomes harder to collect “new” coupons.

The answer is $k \ln k + \Theta(k)$, and has a very simple analysis.

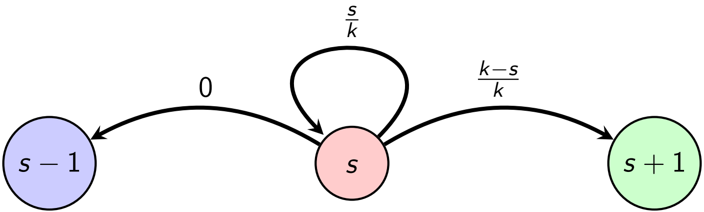
Can think of “learning” the set of coupons.

Coupon Collector Problem

Can define $X_t :=$ number of coupons collected as a random walk.

Coupon Collector Problem

Can define $X_t :=$ number of coupons collected as a random walk.



Quantum Coupon Collector

Quantum Coupon Collector

Let $S \subseteq \mathbb{N}_n$ be the set of coupons. $(|S| = k)$

QCC is learning S with quantum samples

$$|\psi_S\rangle = \frac{1}{\sqrt{k}} \sum_{i \in S} |i\rangle$$

Quantum Coupon Collector

Let $S \subseteq \mathbb{N}_n$ be the set of coupons. $(|S| = k)$

QCC is learning S with quantum samples

$$|\psi_S\rangle = \frac{1}{\sqrt{k}} \sum_{i \in S} |i\rangle$$

Complexity is $\Theta(k \log(\min(k, n-k+1)))$[ABC$^+$20]

Quantum Coupon Collector

Let $S \subseteq \mathbb{N}_n$ be the set of coupons. $(|S| = k)$

QCC is learning S with quantum samples

$$|\psi_S\rangle = \frac{1}{\sqrt{k}} \sum_{i \in S} |i\rangle$$

Complexity is $\Theta(k \log(\min(k, n-k+1)))$[ABC$^+$20]

We only consider $m := k - n << n$

Quantum Coupon Collector

Let $S \subseteq \mathbb{N}_n$ be the set of coupons. $(|S| = k)$

QCC is learning S with quantum samples

$$|\psi_S\rangle = \frac{1}{\sqrt{k}} \sum_{i \in S} |i\rangle$$

Complexity is $\Theta(k \log(\min(k, n - k + 1)))$[ABC$^+$20]

We only consider $m := k - n << n$

$$\rho_B = \frac{1}{\binom{n}{k}} \sum |\psi_S\rangle\langle\psi_S|^{\otimes t}$$

Spectrum of ρ_B

Spectrum of ρ_B

Spectrum matches with matrix M_t

$$M_t[S, S'] = \frac{1}{\binom{n}{k}} \langle \psi_S | \psi_{S'} \rangle^t$$

Spectrum of ρ_B

Spectrum matches with matrix M_t

$$M_t[S, S'] = \frac{1}{\binom{n}{k}} \langle \psi_S | \psi_{S'} \rangle^t$$

Using a matrix recurrence in [ABC$^+$20], we get a recurrence for the spectrum (related to a random walk W on $\mathbb{N}_{m+1}$)

Spectrum of ρ_B

Spectrum matches with matrix M_t

$$M_t[S, S'] = \frac{1}{\binom{n}{k}} \langle \psi_S | \psi_{S'} \rangle^t$$

Using a matrix recurrence in [ABC+20], we get a recurrence for the spectrum (related to a random walk W on $\mathbb{N}_{m+1}$)

$$l_s \lambda_{s,t} = p_{s,0}\big(l_s \lambda_{s,t-1}\big) + p_{s-1,+1}\big(l_{s-1}\lambda_{s-1,t-1}\big) + p_{s+1,-1}\big(l_{s+1}\lambda_{s+1,t-1}\big)$$

Spectrum of ρ_B

Spectrum matches with matrix M_t

$$M_t[S, S'] = \frac{1}{\binom{n}{k}} \langle \psi_S | \psi_{S'} \rangle^t$$

Using a matrix recurrence in [ABC$^+$20], we get a recurrence for the spectrum (related to a random walk W on $\mathbb{N}_{m+1}$)

$$l_s \lambda_{s,t} = p_{s,0}\big(l_s \lambda_{s,t-1}\big) + p_{s-1,+1}\big(l_{s-1}\lambda_{s-1,t-1}\big) + p_{s+1,-1}\big(l_{s+1}\lambda_{s+1,t-1}\big)$$

where $s \in \{1 \ldots m\}$, $l_s = \binom{n}{s} - \binom{n}{s-1}$, $p_{i,j}$ refers to transition probability in W of moving from $i \to i+j$

Random Walk W

$$l_s \lambda_{s,t} = p_{s,0}\big(l_s \lambda_{s,t-1}\big) + p_{s-1,+1}\big(l_{s-1} \lambda_{s-1,t-1}\big) + p_{s+1,-1}\big(l_{s+1} \lambda_{s+1,t-1}\big)$$

Random Walk W

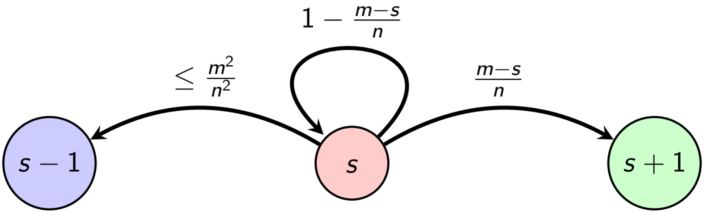
$$l_s\lambda_{s,t} = p_{s,0}\left(l_s\lambda_{s,t-1}\right) + p_{s-1,+1}\left(l_{s-1}\lambda_{s-1,t-1}\right) + p_{s+1,-1}\left(l_{s+1}\lambda_{s+1,t-1}\right)$$

$$\implies \Pr[W_t = s] = l_s\lambda_{s,t}$$

Random Walk W

$$l_s \lambda_{s,t} = p_{s,0}\left(l_s \lambda_{s,t-1}\right) + p_{s-1,+1}\left(l_{s-1} \lambda_{s-1,t-1}\right) + p_{s+1,-1}\left(l_{s+1} \lambda_{s+1,t-1}\right)$$

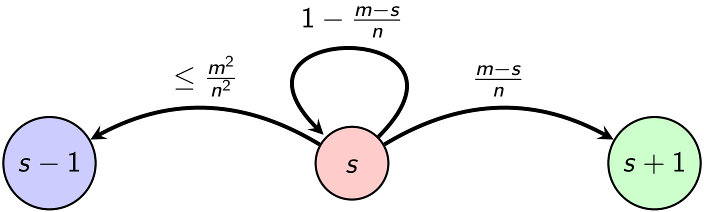
$$\implies \Pr[W_t = s] = l_s \lambda_{s,t}$$



Random Walk W

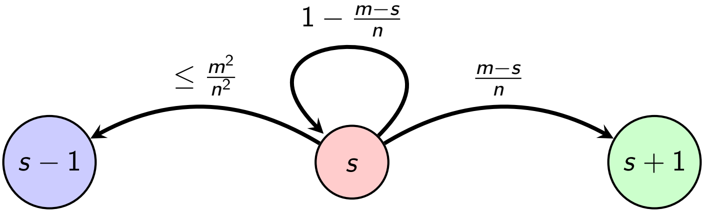
$$l_s \lambda_{s,t} = p_{s,0}\big(l_s \lambda_{s,t-1}\big) + p_{s-1,+1}\big(l_{s-1}\lambda_{s-1,t-1}\big) + p_{s+1,-1}\big(l_{s+1}\lambda_{s+1,t-1}\big)$$

$$\implies \Pr[W_t = s] = l_s \lambda_{s,t}$$

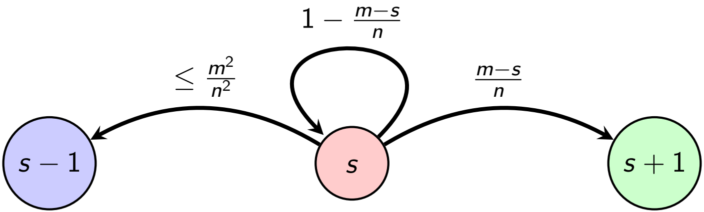


W very closely approximates a variant of coupon collector.

Random Walk W and $S(B)$

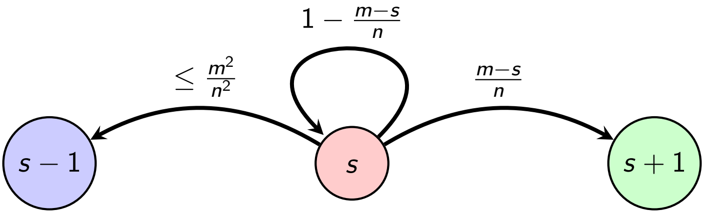


Random Walk W and $S(B)$



Can think of: out of n coupons, mark m, count number of marked coupons collected.

Random Walk W and $S(B)$



Can think of: out of n coupons, mark m, count number of marked coupons collected. Now,

$$S(B) = \sum l_s \lambda_{s,t} \log l_s + O(\log(m+1))$$

$$\approx \log(n) \cdot \sum (l_s \lambda_{s,t}) \cdot s = \log n \cdot \mathbb{E}[W_t]$$

Catch

Catch

Good hold on $I(A : B)$, but $I(A : \chi(B))$ can be too small.

Catch

Good hold on $I(A:B)$, but $I(A:\chi(B))$ can be too small.

Reason: working algorithms exist which "throw away" information

Catch

Good hold on $I(A : B)$, but $I(A : \chi(B))$ can be too small.

Reason: working algorithms exist which "throw away" information

Consider a working sample efficient algorithm:

Catch

Good hold on $I(A:B)$, but $I(A:\chi(B))$ can be too small.

Reason: working algorithms exist which "throw away" information

Consider a working sample efficient algorithm:

1. Use the algorithm to get a guess S'

Catch

Good hold on $I(A:B)$, but $I(A:\chi(B))$ can be too small.

Reason: working algorithms exist which "throw away" information

Consider a working sample efficient algorithm:

1. Use the algorithm to get a guess S'
2. Check correctness of S' (sample efficient)

Catch

Good hold on $I(A:B)$, but $I(A:\chi(B))$ can be too small.

Reason: working algorithms exist which “throw away” information

Consider a working sample efficient algorithm:

1. Use the algorithm to get a guess S'
2. Check correctness of S' (sample efficient)
3. If S' correct, output S', else output garbage.

Catch

Good hold on $I(A:B)$, but $I(A:\chi(B))$ can be too small.

Reason: working algorithms exist which "throw away" information

Consider a working sample efficient algorithm:

1. Use the algorithm to get a guess S'
2. Check correctness of S' (sample efficient)
3. If S' correct, output S', else output garbage.

This throws away correlation between S and S' when $S' \neq S$.

Proof Using Holevo-Curlander bounds

Proof Using Holevo-Curlander bounds

Theorem (HC-bound [Tys10](simplified))

For N equiprobable pure states, with density matrix ρ, then the simultaneous distinguishablity is upper bounded by

$$\frac{1}{\sqrt{N}}\mathrm{Tr}(\sqrt{\rho})$$

Proof Using Holevo-Curlander bounds

Theorem (HC-bound [Tys10](simplified))

For N equiprobable pure states, with density matrix ρ, then the simultaneous distinguishablity is upper bounded by

$$\frac{1}{\sqrt{N}}\mathrm{Tr}(\sqrt{\rho})$$

In our case, the upper bound is

$$\frac{1}{\sqrt{\binom{n}{m}}}\sum l_s\sqrt{\lambda_{s,t}} \approx \sqrt{\frac{l_m}{\binom{n}{m}}}\cdot\sqrt{l_m\lambda_{m,t}}$$

Proof Using Holevo-Curlander bounds

Theorem (HC-bound [Tys10](simplified))

For N equiprobable pure states, with density matrix ρ, then the simultaneous distinguishablity is upper bounded by

$$\frac{1}{\sqrt{N}}\mathrm{Tr}(\sqrt{\rho})$$

In our case, the upper bound is

$$\frac{1}{\sqrt{\binom{n}{m}}}\sum l_s\sqrt{\lambda_{s,t}} \approx \sqrt{\frac{l_m}{\binom{n}{m}}}\cdot\sqrt{l_m\lambda_{m,t}}$$

Theorem (Lower Bound for QCC)

Sample complexity of QCC is $(1-o(1))k\ln(\min k, n-k+1)$.

Conclusion and Additional Points

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning
- Can try to use this method on other problems considered in [AdW18]

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning
- Can try to use this method on other problems considered in [AdW18]
- Doesnt work directly on Quantum Coupon Collector

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning
- Can try to use this method on other problems considered in [AdW18]
- Doesnt work directly on Quantum Coupon Collector
- Works if working with approximation variant of coupon collector

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning
- Can try to use this method on other problems considered in [AdW18]
- Doesnt work directly on Quantum Coupon Collector
- Works if working with approximation variant of coupon collector
- Obtained spectrum gives optimal lower bound with HC Bound

Conclusion and Additional Points

- Quantum PAC learning lower bounds using information theory
- Quantum Agnostic learning very similar to PAC learning
- Can try to use this method on other problems considered in [AdW18]
- Doesnt work directly on Quantum Coupon Collector
- Works if working with approximation variant of coupon collector
- Obtained spectrum gives optimal lower bound with HC Bound

End

Thank You!

References

References I

Srinivasan Arunachalam, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf.
Quantum Coupon Collector.
In Steven T. Flammia, editor, *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, volume 158 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:17, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Srinivasan Arunachalam and Ronald de Wolf.
Optimal quantum sample complexity of learning algorithms.
Journal of Machine Learning Research, 19(1):2879–2878, January 2018.

References II

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth.
Learnability and the Vapnik-Chervonenkis dimension.
Journal of the ACM, 36(4):929–965, October 1989.

Nader H. Bshouty and Jeffrey C. Jackson.
Learning DNF over the uniform distribution using a quantum example oracle.
SIAM Journal on Computing, 28(3):1136–1153, 1998.

Steve Hanneke.
The optimal sample complexity of PAC learning.
Journal of Machine Learning Research, 17(38):1–15, January 2016.

Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu.
Sample-optimal tomography of quantum states.
IEEE Transactions on Information Theory, 63(9):5628–5641, 2017.

References III

Jon Tyson.
Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates.
Journal of Mathematical Physics, 51(9):092204, 2010.

Leslie G. Valiant.
A theory of the learnable.
Communications of the ACM, 27(11):1134–1142, 1984.

End