Predictive Quantum Learning

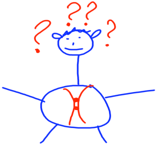
Dmitry Gavinsky

NEC Labs, Princeton

Learning under Standard Setting

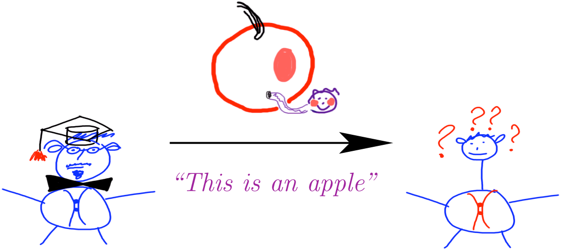
- The teacher knows what *apples* and *oranges* are.

Learning under Standard Setting



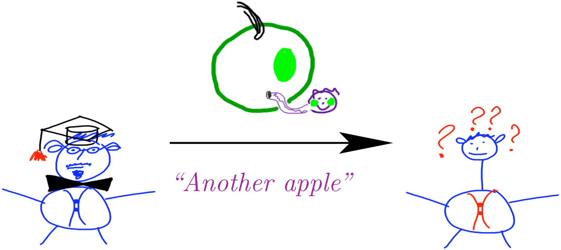
- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.

Learning under Standard Setting



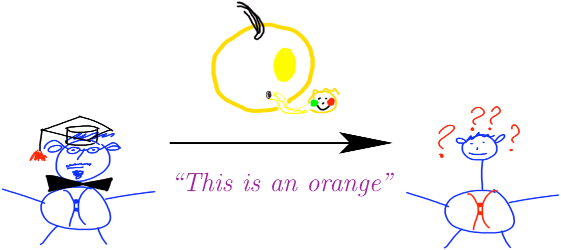
- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.
- ▶ The teacher shares his valuable knowledge with the student.

Learning under Standard Setting



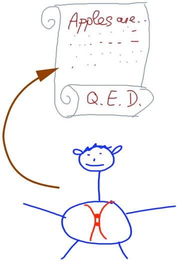
- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.
- The teacher shares his valuable knowledge with the student.

Learning under Standard Setting



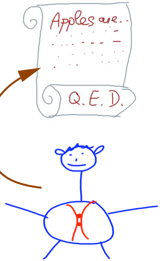
- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.
- The teacher shares his valuable knowledge with the student.

Learning under Standard Setting



- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.
- The teacher shares his valuable knowledge with the student.
- In the *testing phase* the the student writes an *essay (hypothesis), explaining how to distinguish apples from oranges*.

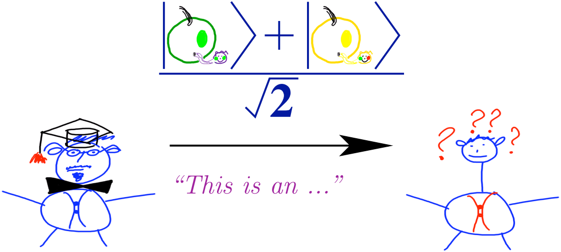
Learning under Standard Setting



- The teacher knows what *apples* and *oranges* are.
- But the student doesn't.
- The teacher shares his valuable knowledge with the student.
- In the *testing phase* the the student writes an *essay (hypothesis), explaining how to distinguish apples from oranges*.

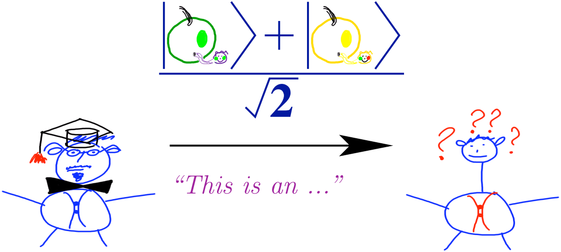
► This model is called *Probably Approximately Correct (PAC)*, it has been introduced by Valiant [V84].

Quantum Learning



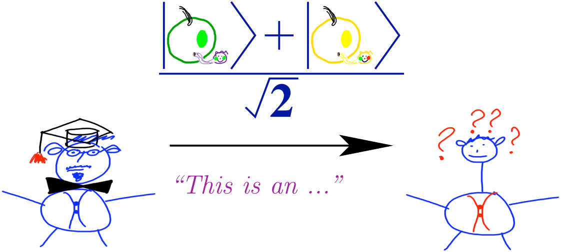
- We may allow *quantum messages* from the teacher.

Quantum Learning



- We may allow *quantum messages* from the teacher.
- The student may be a *quantum algorithm*.

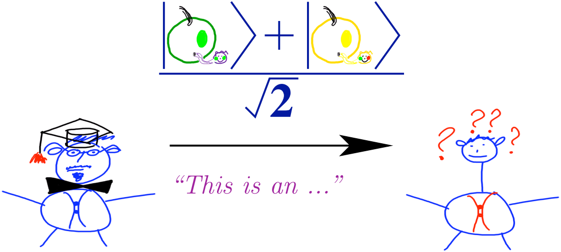
Quantum Learning



- We may allow *quantum messages* from the teacher.
- The student may be a *quantum algorithm*.

▶ Quantum models are at least as strong as their classical analogues.

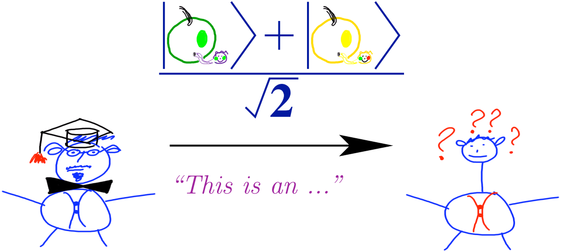
Quantum Learning



- We may allow *quantum messages* from the teacher.
- The student may be a *quantum algorithm*.
- Quantum models are at least as strong as their classical analogues.

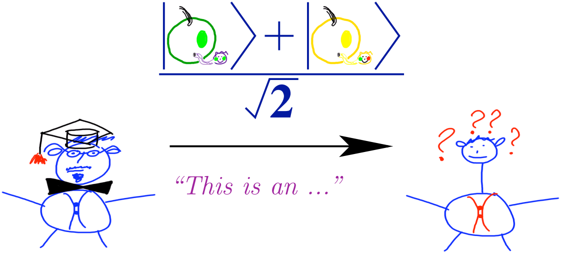
▶ A *concept class* is *learnable* in a given model if an *efficient algorithm* can play the role of the student.

Quantum Learning



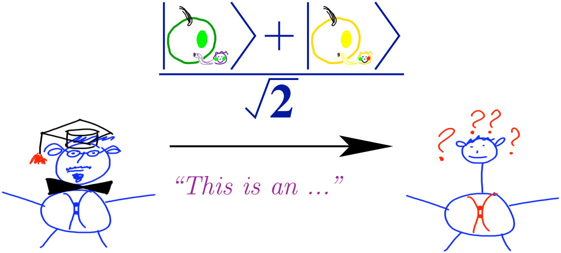
- We may allow *quantum messages* from the teacher.
- The student may be a *quantum algorithm*.
- Quantum models are at least as strong as their classical analogues.
- A *concept class* is *learnable* in a given model if an *efficient algorithm* can play the role of the student.
- Are quantum models stronger than classical?

Earlier Work



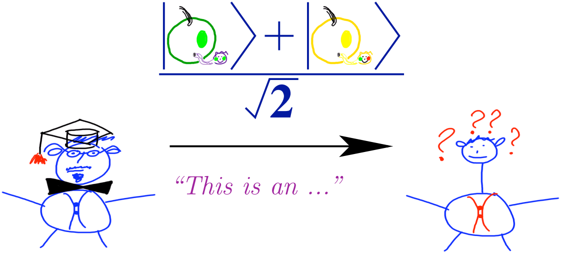
- In [BJ95] Bshouty and Jackson gave an *efficient algorithm for quantum PAC learning of DNF formulas* – this is currently not known to be possible from classical examples (even by a quantum algorithm).

Earlier Work



- In [BJ95] Bshouty and Jackson gave an *efficient algorithm for quantum PAC learning of DNF formulas* – this is currently not known to be possible from classical examples (even by a quantum algorithm).
- In [SG04] Servedio and Gortler showed that *from the information-theoretic point of view, classical PAC is not weaker than its quantum analogue*. On the other hand, they demonstrated a *separation modulo certain cryptographic assumptions*.

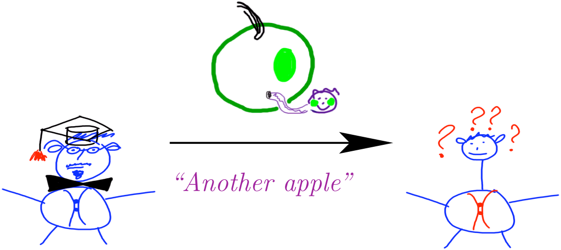
Earlier Work



- In [BJ95] Bshouty and Jackson gave an *efficient algorithm for quantum PAC learning of DNF formulas* – this is currently not known to be possible from classical examples (even by a quantum algorithm).
- In [SG04] Servedio and Gortler showed that *from the information-theoretic point of view, classical PAC is not weaker than its quantum analogue*. On the other hand, they demonstrated a *separation modulo certain cryptographic assumptions*.

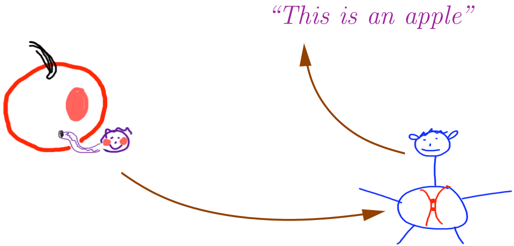
Learning under Predictive Setting

Learning under Predictive Setting



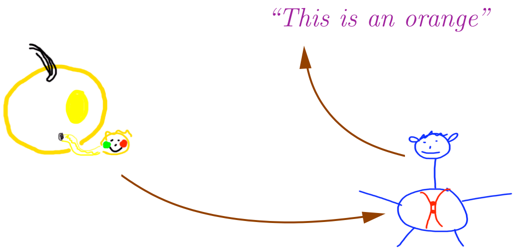
- The *learning phase* is the same.

Learning under Predictive Setting



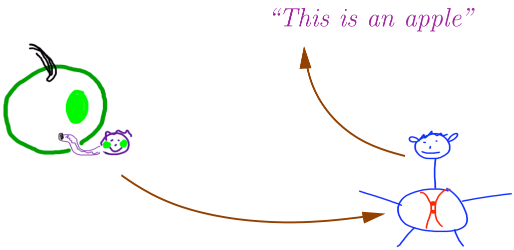
- The *learning phase* is the same.
- In the *testing phase* the student demonstrates *ability to distinguish apples from oranges*.

Learning under Predictive Setting



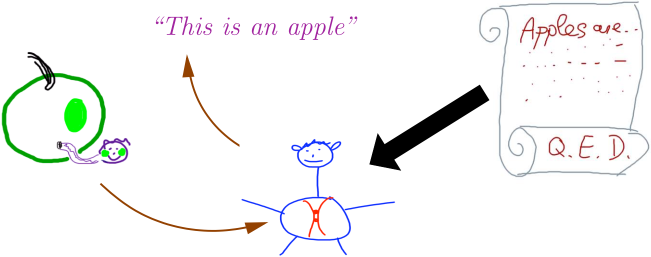
- The *learning phase* is the same.
- In the *testing phase* the student demonstrates *ability to distinguish apples from oranges*.

Learning under Predictive Setting



- The *learning phase* is the same.
- In the *testing phase* the student demonstrates *ability to distinguish apples from oranges*.

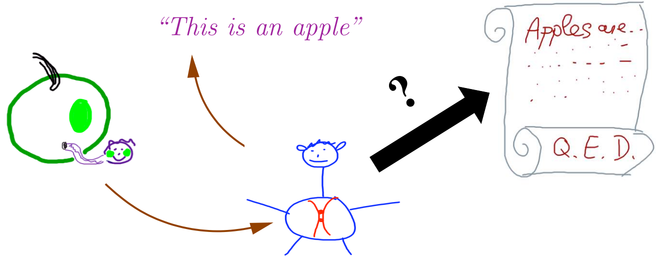
Learning under Predictive Setting



- The *learning phase* is the same.
- In the *testing phase* the student demonstrates *ability to distinguish apples from oranges*.

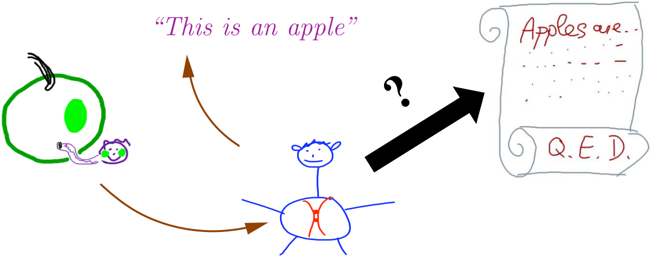
▶ Clearly, *standard learnability implies predictive learnability* (a hypothesis can be used as a distinguishing algorithm).

Does Predictive Learnability Equal Standard Learnability?



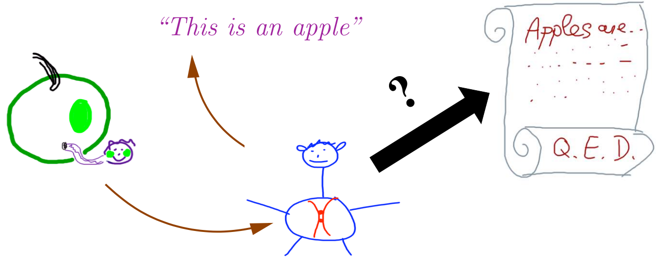
- It is well known that *in any classical learning model, a predictive learning algorithm can be turned into a standard one*.

Does Predictive Learnability Equal Standard Learnability?



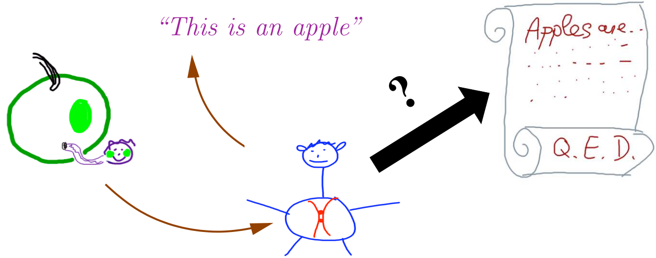
- It is well known that *in any classical learning model, a predictive learning algorithm can be turned into a standard one*.
  - The argument *does not translate to the case of quantum models*.

Does Predictive Learnability Equal Standard Learnability?



- It is well known that *in any classical learning model, a predictive learning algorithm can be turned into a standard one*.
- The argument *does not translate to the case of quantum models*.
- ▶ We will see that, indeed, *quantum predictive learning is qualitatively stronger than quantum standard learning*.

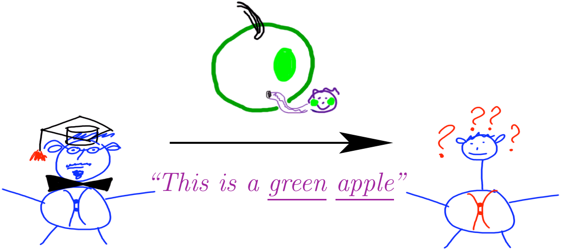
Does Predictive Learnability Equal Standard Learnability?



- It is well known that *in any classical learning model, a predictive learning algorithm can be turned into a standard one*.
- The argument *does not translate to the case of quantum models*.
- We will see that, indeed, *quantum predictive learning is qualitatively stronger than quantum standard learning*.
- ▶ Observe that *unconditional separation between quantum and classical learning* immediately follows (we will make a more formal statement later).

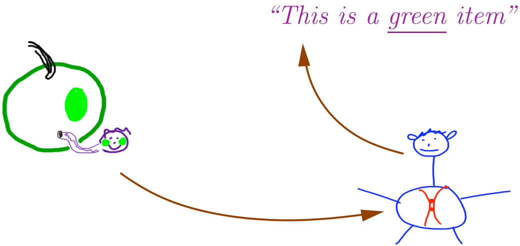
Relational Concept Classes

Relational Concept Classes



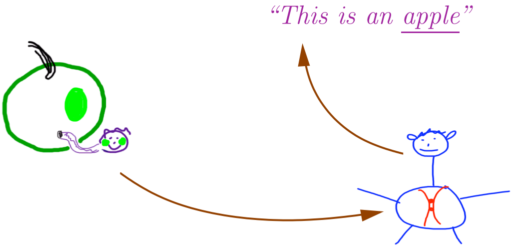
- Items may have *several* valid labels.

Relational Concept Classes



- Items may have *several* valid labels.
- In the testing phase, given an item the student may respond with any of its valid labels.

Relational Concept Classes



- Items may have *several* valid labels.
- In the testing phase, given an item the student may respond with any of its valid labels.

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class $\mathcal{H}$ that approximates $\mathcal{C}$ must be of double-exponential size*.

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class $\mathcal{H}$ that approximates $\mathcal{C}$ must be of double-exponential size*.

- Simple counting reveals that *almost all concept classes are unspeakable*.

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class* $\mathcal{H}$ *that approximates* $\mathcal{C}$ *must be of double-exponential size*.

- Simple counting reveals that *almost all concept classes are unspeakable*.

▶ It holds that

  - *no standard algorithm (either quantum or classical) can efficiently learn an unspeakable class* (trivially true);

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class* $\mathcal{H}$ *that approximates* $\mathcal{C}$ *must be of double-exponential size*.

- Simple counting reveals that *almost all concept classes are unspeakable*.

▶ It holds that

  - *no standard algorithm (either quantum or classical) can efficiently learn an unspeakable class* (trivially true);
  - *no classical algorithm can efficiently learn an unspeakable class*;

Unspeakable Concept Classes

- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class $\mathcal{H}$ that approximates $\mathcal{C}$ must be of double-exponential size*.

- Simple counting reveals that *almost all concept classes are unspeakable*.

▶ It holds that

  - *no <u>standard</u> algorithm (either quantum or classical) can efficiently learn an unspeakable class* (trivially true);
  - *no <u>classical</u> algorithm can efficiently learn an unspeakable class*;
  - *unspeakable classes of <u>functions</u> are not learnable efficiently (either quantumly or classically)*.

Unspeakable Concept Classes

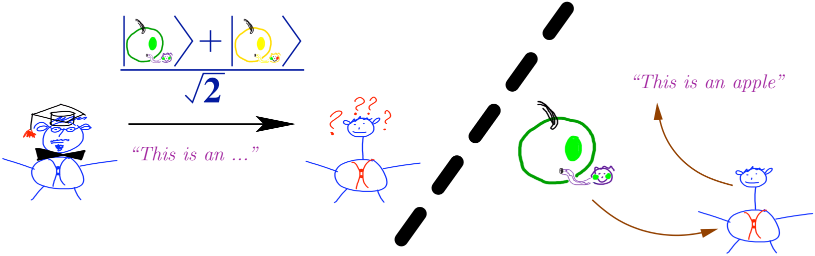
- A concept class $\mathcal{C}$ is *unspeakable* if it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis *approximates* the target.

  More precisely, *any concept class $\mathcal{H}$ that approximates $\mathcal{C}$ must be of double-exponential size*.

- Simple counting reveals that *almost all concept classes are unspeakable*.
- It holds that
  - *no standard algorithm (either quantum or classical) can efficiently learn an unspeakable class* (trivially true);
  - *no classical algorithm can efficiently learn an unspeakable class*;
  - *unspeakable classes of functions are not learnable efficiently (either quantumly or classically)*.

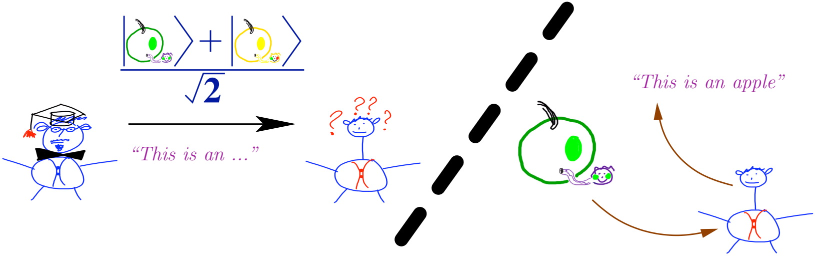
► Therefore, *willing to learn unspeakable concepts*, we can only hope to do so for a *relational* class, in a *quantum predictive* model.

Our Results



- We define a *quantum predictive version of PAC*, denoted by *QAC**.

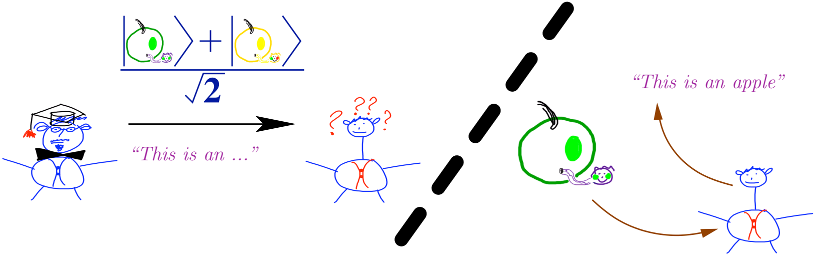
Our Results



- We define a *quantum predictive version of PAC*, denoted by QAC^*.

▶ There exists a relational concept class $\mathcal{C}$ that is *unspeakable* but can be *efficiently learnt in* QAC^*.

Our Results



- We define a *quantum predictive version of PAC*, denoted by *QAC**.
- There exists a relational concept class $\mathcal{C}$ that is *unspeakable* but can be *efficiently learnt in QAC**.

► In particular, *$\mathcal{C}$ is not learnable classically*.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.
- The set of queries is $[N-1]$, represented by binary strings of length $n = \lceil \log N \rceil$.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.
- The set of queries is $[N-1]$, represented by binary strings of length $n = \lceil \log N \rceil$.
- A pair $(x, c_x \oplus c_{x+j}) \in \mathbb{Z}_N \times \{0,1\}$ is a *valid answer* to query j w.r.t. $c \in \mathcal{C}$.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.
- The set of queries is $[N-1]$, represented by binary strings of length $n = \lceil \log N \rceil$.
- A pair $\underline{(x, c_x \oplus c_{x+j})} \in \mathbb{Z}_N \times \{0,1\}$ is a *valid answer* to query j w.r.t. $c \in \mathcal{C}$.

► We will demonstrate that *$\mathcal{C}$ is unspeakable – in particular, it is not learnable classically*.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.
- The set of queries is $[N-1]$, represented by binary strings of length $n = \lceil \log N \rceil$.
- A pair $\underline{(x, c_x \oplus c_{x+j})} \in \mathbb{Z}_N \times \{0,1\}$ is a *valid answer* to query j w.r.t. $c \in \mathcal{C}$.
- We will demonstrate that *$\mathcal{C}$ is unspeakable – in particular, it is not learnable classically*.
- On the other hand, *$\mathcal{C}$ is efficiently learnable in QAC^**.

The Concept Class

- Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $c \in \{0,1\}^N$.
- The set of queries is $[N-1]$, represented by binary strings of length $n = \lceil \log N \rceil$.
- A pair $\underline{(x, c_x \oplus c_{x+j})} \in \mathbb{Z}_N \times \{0,1\}$ is a *valid answer* to query j w.r.t. $c \in \mathcal{C}$.
- We will demonstrate that *$\mathcal{C}$ is unspeakable – in particular, it is not learnable classically*.
- On the other hand, *$\mathcal{C}$ is efficiently learnable in QAC**.
- This construction has been inspired by a communication problem defined in [BJK04].

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i \left|j,x\right\rangle$

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i |j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i |j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}} |j,x_0\rangle$.

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i |j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i |j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}} |j,x_0\rangle$.
  - Then he performs $|j,x_0\rangle \to |j+x_0,x_0\rangle$, which results in

$$|\alpha_{x_0}\rangle \stackrel{def}{=} \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}}(-1)^{c_k} |k\rangle.$$

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i |j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i |j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}} |j,x_0\rangle$.
- Then he performs $|j,x_0\rangle \to |j+x_0,x_0\rangle$, which results in

$$|\alpha_{x_0}\rangle \stackrel{def}{=} \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}}(-1)^{c_k} |k\rangle.$$

The student is ready for the testing phase.

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i |j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i |j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}} |j,x_0\rangle$.
- Then he performs $|j,x_0\rangle \to |j+x_0,x_0\rangle$, which results in

$$|\alpha_{x_0}\rangle \stackrel{def}{=} \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}}(-1)^{c_k} |k\rangle.$$

The student is ready for the testing phase.

- Given a test question $q\in[N-1]$, the student performs projective measurement of $|\alpha_{x_0}\rangle$ onto $(N-1)/2$ subspaces, each spanned by a pair of vectors from

$$\left\{\left(|x_0+(2i+1)q\rangle, |x_0+(2i+2)q\rangle\right) \;\middle|\; 0\le i\le \tfrac{N-3}{2}\right\}.$$

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i |j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i |j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}} |j,x_0\rangle$.
- Then he performs $|j,x_0\rangle \to |j+x_0,x_0\rangle$, which results in

$$|\alpha_{x_0}\rangle \stackrel{def}{=} \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}}(-1)^{c_k} |k\rangle.$$

The student is ready for the testing phase.

- Given a test question $q \in [N-1]$, the student performs projective measurement of $|\alpha_{x_0}\rangle$ onto $(N-1)/2$ subspaces, each spanned by a pair of vectors from

$$\left\{ \left(|x_0+(2i+1)q\rangle,|x_0+(2i+2)q\rangle\right) \;\middle|\; 0 \le i \le \frac{N-3}{2} \right\}.$$

- ▶ The state becomes either $|a\rangle + |a+q\rangle$ or $|a\rangle - |a+q\rangle$, corresponding to $c_a \oplus c_{a+q} = 0$ or $c_a \oplus c_{a+q} = 1$, respectively.

Efficient QAC^*-Learning of $\mathcal{C}$

- The student receives from the teacher the state $\sum_{(j,x,i)\in c}(-1)^i\,|j,x\rangle$ and *measures the second register in the computational basis*, obtaining $\sum_{(j,x_0,i)\in c}(-1)^i\,|j,x_0\rangle = \sum_{j\in[N-1]}(-1)^{c_{x_0+j}}\,|j,x_0\rangle$.
- Then he performs $|j,x_0\rangle \to |j+x_0,x_0\rangle$, which results in

$$|\alpha_{x_0}\rangle \stackrel{def}{=} \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}}(-1)^{c_k}\,|k\rangle.$$

The student is ready for the testing phase.

- Given a test question $q\in[N-1]$, the student performs projective measurement of $|\alpha_{x_0}\rangle$ onto $(N-1)/2$ subspaces, each spanned by a pair of vectors from

$$\left\{\left(|x_0+(2i+1)q\rangle,|x_0+(2i+2)q\rangle\right)\ \middle|\ 0\le i\le \tfrac{N-3}{2}\right\}.$$

- The state becomes either $|a\rangle+|a+q\rangle$ or $|a\rangle-|a+q\rangle$, corresponding to $c_a\oplus c_{a+q}=0$ or $c_a\oplus c_{a+q}=1$, respectively.

▶ The student responds with $(a, c_a\oplus c_{a+q})$, as required.

$\mathcal{C}$ Is Unspeakable

- Assume that some concept class $\mathcal{H}$ approximates $\mathcal{C}$.

$\mathcal{C}$ Is Unspeakable

- Assume that some concept class $\mathcal{H}$ approximates $\mathcal{C}$.

► We show that *each $h \in \mathcal{H}$ approximates at most $2^{O(\sqrt{N})} = 2^{O(2^{n/2})}$ concepts from $\mathcal{C}$* – this follows from the properties of $\mathcal{C}$, first observed in [BJK04].

$\mathcal{C}$ Is Unspeakable

- Assume that some concept class $\mathcal{H}$ approximates $\mathcal{C}$.
- We show that *each $h \in \mathcal{H}$ approximates at most* $2^{O(\sqrt{N})} = 2^{O(2^{n/2})}$ *concepts from $\mathcal{C}$* – this follows from the properties of $\mathcal{C}$, first observed in [BJK04].

► As $|\mathcal{C}| = 2^{2^n}$, it must hold that

$$|\mathcal{H}| \geq \frac{|\mathcal{C}|}{2^{O(2^{n/2})}} = 2^{\Omega(2^{n-n/2})} = 2^{2^{\Omega(n)}}.$$

$\mathcal{C}$ Is Unspeakable

- Assume that some concept class $\mathcal{H}$ approximates $\mathcal{C}$.
- We show that *each $h \in \mathcal{H}$ approximates at most* $2^{O\left(\sqrt{N}\right)} = 2^{O\left(2^{n/2}\right)}$ *concepts from $\mathcal{C}$* – this follows from the properties of $\mathcal{C}$, first observed in [BJK04].
- As $|\mathcal{C}| = 2^{2^n}$, it must hold that

$$|\mathcal{H}| \geq \frac{|\mathcal{C}|}{2^{O\left(2^{n/2}\right)}} = 2^{\Omega\left(2^{n-n/2}\right)} = 2^{2^{\Omega(n)}}.$$

► Therefore, *$\mathcal{C}$ is unspeakable*.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.

- One may consider *quasi-unspeakable* concept classes that *admit concise hypotheses, but only those that cannot be efficiently evaluated*.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.
- One may consider *quasi-unspeakable* concept classes that *admit concise hypotheses, but only those that cannot be efficiently evaluated*.
- Our impossibility of separation arguments generalize, implying that *quasi-unspeakable classes are not learnable classically*.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.
- One may consider *quasi-unspeakable* concept classes that *admit concise hypotheses, but only those that cannot be efficiently evaluated*.
- Our impossibility of separation arguments generalize, implying that *quasi-unspeakable classes are not learnable classically*.

► We do not know whether efficient quantum predictive learning of quasi-unspeakable <u>functional</u> concept classes is possible.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.
- One may consider *quasi-unspeakable* concept classes that *admit concise hypotheses, but only those that cannot be efficiently evaluated*.
- Our impossibility of separation arguments generalize, implying that *quasi-unspeakable classes are not learnable classically*.
- We do not know whether efficient quantum predictive learning of quasi-unspeakable functional concept classes is possible.

► Such an example would probably require (non-uniform) hardness assumptions, but on the other hand, might be viewed as a stronger separation between quantum and classical learning.

Further Work

- We have demonstrated that *efficient quantum predictive learning of an unspeakable relational concept class is possible*.

  The same would be *impossible for a functional concept class*.
- One may consider *quasi-unspeakable* concept classes that *admit concise hypotheses, but only those that cannot be efficiently evaluated*.
- Our impossibility of separation arguments generalize, implying that *quasi-unspeakable classes are not learnable classically*.
- We do not know whether efficient quantum predictive learning of quasi-unspeakable functional concept classes is possible.
- Such an example would probably require (non-uniform) hardness assumptions, but on the other hand, might be viewed as a stronger separation between quantum and classical learning.

▶ *Assuming $BQP \not\subseteq P/poly$, the answer is trivial* (let $\mathcal{C} \stackrel{\text{def}}{=} \{f_L\}$, for any $L \in BQP \setminus P/poly$) – the goal is to *weaken the assumptions*.


Thank you!