Predictive Quantum Learning

Dmitry Gavinsky
NEC Laboratories America, Inc.
4 Independence Way, Suite 200
Princeton, NJ 08540, U.S.A.

Abstract

We demonstrate a relational concept class that is *efficiently learnable* in certain quantum analogue of the PAC model, while in *any* classical learning model exponential amount of training data would be required. We show that our separation is the best possible in several ways; in particular, there is no analogous result for a functional class, as well as for some weaker versions of quantum PAC.

This is the first (unconditional) separation of quantum and classical learning models.

1 Introduction

The main goal of quantum computing is to exhibit problems where quantum computers are much faster (or otherwise better) than classical ones. Preferably, exponentially better. To establish such separation through developing an efficient quantum algorithm for a computational task whose classical complexity is known to be high remains beyond the reach of human civilization. On the other hand, (unconditional) exponential complexity separations between quantum models and their classical analogues are known in some other fields, e.g., in communication complexity.

In this paper we look for unconditional exponential separations in *Machine Learning*. The field deals with the following kind of problems. In the *learning phase*, a *teacher* communicates with a *student*, in order to let the latter *learn* a concept ℓ, which is guaranteed to belong to certain *concept class* $\mathcal{C}$. Then in the *testing phase* it is checked how successful the student has been.

As an example, consider a teacher who trains a student to distinguish different types of fruits. A possible scenario for the learning phase would be that the teacher shows to the student different objects, and give the required explanations: "This is an apple", "This is an orange", and so on. Then in the testing phase different fruits are shown to the student, who is supposed to correctly identify them.

A *learning model* specifies the set of rules, governing the learning and the testing phases. One of the most natural models is that of *Probably Approximately Correct (PAC)*, defined by Valiant [V84]. In the learning phase of *PAC* a sequence of labeled examples $(x_1, \ell(x_1)), \ldots, (x_k, \ell(x_k))$ are sent by the teacher to the student – just like in the case considered above (where x_i-s were the items demonstrated and $\ell(\cdot)$ was the explanations given by the teacher). The examples are independently chosen by $x_i \sim D$, where D is some distribution over the domain. In the testing phase domain elements are offered to the student according to the same distribution D, and his goal is given x to respond with $\ell(x)$.

We say that certain concept class $\mathcal{C}$ that consists of possible candidates[1] for $\ell(\cdot)$ is *efficiently learnable* in a given learning model if there exists an efficient algorithm, successfully performing the student's task according to the model definition. Algorithm's efficiency usually (always in this paper) means that its running time is upper-bounded by a polynomial of the input length. More specifically, in the case of PAC we require that a learning algorithm runs in time polylogarithmic in the domain size (note that running time of a learning algorithm is, trivially, an upper bound on the number of training examples it uses).

1.1 Previous work

In [BJ95] Bshouty and Jackson introduced a natural quantum analogue of PAC, which we will denote by QAC.

It is natural to ask whether quantum learning models can offer any advantages over the classical ones, in terms of efficiency. The question has been considered by Servedio and Gortler [SG04], who showed that in the case of functional hypotheses the models PAC and QAC are equivalent from the information-theoretic point of view. On the other hand, they gave evidences that quantum models are more efficient than their classical analogues if certain cryptographic assumptions hold.

The main result of [BJ95] is an efficient algorithm that learns DNF formulas from independently chosen *quantum* examples – this is currently not known to be possible from classical examples (even with a quantum learning algorithm).

1.2 Our results

We consider several possible generalizations of PAC and QAC. First, we allow *relational* concept classes.[2] Second, we classify all learning models as follows:

- We call *standard* a learning model that requires that in the end of the learning phase the student produces a *final hypothesis*, which is an algorithmic procedure able to answer the questions in the testing phase "on student's behalf".
- We say that a model is *quasi-predictive* if the student learns in order to answer queries in the testing phase. The number of questions that will be asked is unknown during the learning phase.
- We call a model *predictive* if the student should answer queries in the testing phase, and an upper bound on the number of queries is an efficiency parameter, i.e., the learning complexity should be at most polynomial in the number of test question.

We allow standard, quasi-predictive, and predictive versions of both PAC and QAC.

We will call a concept class $\mathcal{C}$ *unspeakable* if, informally speaking, it *does not admit efficient hypotheses representation*, even if one only requires that a hypothesis approximates

[1] Formally speaking, in the example considered before there are several concepts taught by the teacher at the same time – one corresponding to each type of fruits. A more conventional PAC setting would be training in order to learn, say, *what is apple*: A teacher would show objects and tell which of them are apples and which are not. Similarly, in the testing phase the student would only have to distinguish apples from "anything else".

[2] Consider the situation where during the learning stage items are marked by several labels, e.g., "green" and "apple". Then if a similar object appears in the testing phase, the student is allowed to answer either "green" or "apple" – the both choices are accepted as correct.

the target (that is, only hypotheses classes of double-exponential size can approximate all elements of $\mathcal{C}$). Simple counting reveals that almost all concept classes are unspeakable. Clearly, *no standard learning algorithm (either quantum or classical) can efficiently learn an unspeakable concept class*, because its output length would have to be exponential.

It is well known that in any "reasonable" classical learning model, a predictive learning algorithm can be turned into a standard one.[3] Therefore, *no classical algorithm can efficiently learn an unspeakable concept class.*

The argument does not translate to the case of quantum models, because of the uncertainty principle which says that quantum data is not completely accessible through measurements, as well as the efficiency limitations of the learning model under consideration. On the one hand, we will show that the argument generalizes to show that certain "quasi-hypothesis" of polynomial length can be extracted from a quantum quasi-predictive learning algorithm. Therefore, *no quasi-predictive learning algorithm (either quantum or classical) can efficiently learn an unspeakable concept class.*

On the other hand, we will demonstrate an *efficient quantum predictive learning algorithm that learns an unspeakable relational concept class.*

We also show that considering relational concept classes is essential in order to efficiently learn an unspeakable concept class. Therefore, the combination of a relational concept and quantum predictive mode of learning is crucial for learning an unspeakable class.

The following is our main result(cf. Theorem 3.1, Lemma 2.1, and Lemma 4.1).

Theorem 1.1. *There exists a relational concept class that is unspeakable but can be efficiently learnt in the model of predictive quantum PAC.*

Classical learning of an unspeakable concept class is not possible from less than exponential amount of information from the teacher, even by a computationally unlimited student.

Neither standard nor quasi-predictive learning of an unspeakable concept class is possible from less than exponential amount of quantum (w.l.g.) information from the teacher, even by a computationally unlimited student.

Predictive learning of an unspeakable functional concept class is not possible from less than exponential amount of quantum (w.l.g.) information from the teacher, even by a computationally unlimited student.

2 Definitions and more

A good survey of quantum vs. classical learning can be found in [SG04].

We will usually ignore normalization factors and global phases of quantum states.

For any $a \in \mathbb{N}$ we denote $[a] \stackrel{\text{def}}{=} \{1, \ldots, a\}$. We view the elements of $\mathbb{Z}_a$ as integers $\{0, 1, , \ldots, a-1\}$, and accordingly we define their ordering $0 < 1 < \cdots < a-1$. For any $i \in \mathbb{N}$ and $b \in \mathbb{Z}_a$, let $i \cdot b = ib$ be the i'th power of b w.r.t. the group operation $+$.

For two functions $f_1, f_2 : A \to B$ we say that one approximates the other if $\mathbf{Pr}_x\left[f_1(x) = f_2(x)\right] \geq 2/3$, where x is uniformly distributed over A.

[3] That can be achieved by producing a final hypothesis consisting of a description of the predicting subroutine together with all the data available after the learning stage.

We will consider binary relations of the form $r \subseteq A \times B$, and it will always be assumed that $\forall x \in A : \exists y \in B : (x, y) \in r$. We say that a hypothesis[4] $h : A \to B$ approximates a relation $r \subseteq A \times B$ if $\mathbf{Pr}_x [(x, h(x,)) \in r] \geq 2/3$, where the probability is taken over x uniformly distributed over A and possible randomized choices of h.

We allow relational concept classes containing subsets of $\{0,1\}^n \times B$, where n is the input length. Functional (Boolean) concept classes are a special case of relational ones, corresponding to $B = \{0,1\}$. We say that a concept class $\mathcal{C}$ is approximated by $\mathcal{C}'$ if for every $g \in \mathcal{C}$ there exists $g' \in \mathcal{C}'$, such that g' approximates g.

Definition 1. *Let $\mathcal{C}$ be a concept class. We say that $\mathcal{C}$ is unspeakable if $|\mathcal{C}'| \in 2^{2^{\Omega(n)}}$ holds for any $\mathcal{C}'$ that approximates $\mathcal{C}$.*

2.1 Standard vs. quasi-predictive vs. predictive learning

To us, the setting of quasi-predictive learning is not very promising (even in the quantum case), in view of the following. Assume that we have a quasi-predictive learning algorithm L that receives at most k qubits from the teacher during the learning phase. If we ignore time complexity issues, and only require that all input and output be of length polynomial in k, then we can construct a standard learning algorithm L'. It is intuitively clear (and can be proved using standard methods, cf. [A04]) that it is possible to choose $O(k)$ queries, such that based on valid answers to them, a computationally unlimited student can, with high confidence, answer *all possible queries*, making no further use of the quantum data collected in the learning phase.[5] Consequently, a list of $O(k)$ queries together with their answers can represent a final hypothesis. Observe that such final hypothesis is short but may, in general, be hard to evaluate – as opposed to the case of standard models, where efficient hypothesis evaluation is required.

By the definition, unspeakable concepts do not admit concise hypotheses of any type, and therefore:

Lemma 2.1. *Both standard and quasi-predictive learning of an unspeakable (either functional or relational) concept class, even by a computationally unlimited student, requires exponential amount of quantum (w.l.g.) information from the teacher.*

On the other hand, our notion of predictive learning only requires that the student be efficient in terms of the number of question asked in the testing phase. We will show that in this case a predictive quantum student is able to learn certain unspeakable relational concept class.

[4] A hypothesis may be represented by any algorithm, or more generally, any rule for generating output, based on the input.

[5] As long as a question exists that cannot be addressed based on the previously given answers, but can be answered using a measurement of the quantum data – this measurement reveals $\Omega(1)$ amount of classical information contained in quantum data *before the testing phase had started*. And the total amount of such information is known to be $O(k)$.

2.2 On the number of test questions in predictive learning

Generally speaking, we want our predictive learning algorithms to be of complexity at most polynomial[6] in two parameters – the length of a question n and the number of testing queries ℓ. Alternatively, we can drop the parameter ℓ and require that our algorithm be efficient in terms of n only, and be able to answer a *single* testing query with sufficient accuracy. If we run ℓ independent copies of such single-query algorithm, then we obtain an efficient algorithm for answering ℓ queries. For notational convenience and w.l.g., in the rest of the paper we will only consider predictive learning algorithms aiming to answer a single testing query.

We define our predictive quantum version of PAC as follows.

Definition 2. *In QAC^* learning model a learning algorithm can ask for arbitrarily many copies of the state $\sum_{(x,y)\in C} |x, y\rangle$, where $C \in \mathcal{C}$ is a relational concept. In the end of the learning process the algorithm receives an element $x \in \{0,1\}^n$, and should output any y, such that $(x, y) \in C$.*

We say that an algorithm QAC^*-learns a concept class $\mathcal{C}$ if it succeeds with probability at least $3/4$, for any $x \in \{0,1\}^n$. A learning algorithm is efficient if its running time is at most polynomial in n. A concept class is efficiently learnable in QAC^* if there exists an efficient algorithm that QAC^*-learns it. [7]

3 Our concept class

Let us define a concept class $\mathcal{C}$, that will be shown to be both unspeakable and efficiently QAC^*-learnable, as follows (the definition has been inspired by a communication problem defined in [BJK04]).

Definition 3. *Let N be prime. Every concept in the class $\mathcal{C}$ is represented by $C \in \{0,1\}^N$. The set of queries is $[N-1]$, naturally represented by binary strings of length $n = \lceil \log N \rceil$. A pair $(x, i) \in \mathbb{Z}_N \times \{0,1\}$ is a valid answer to query j w.r.t. $C \in \mathcal{C}$ if $C_x \oplus C_{x+j} = i$.*

We will abuse the notation by viewing each $C \in \mathcal{C}$ either as a binary string of length N or as $\big\{(j, x, i) \big| (x, i) \text{ is a valid answer to } j \text{ w.r.t. } C\big\}$.

Theorem 3.1. *The concept class $\mathcal{C}$ is unspeakable (in particular, it is not classically efficiently learnable). On the other hand, $\mathcal{C}$ is efficiently learnable in QAC^*.*

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

A student will need k QAC-examples in order to answer to the testing query with probability $1 - 1/2^k$, and whenever an answer is produced it is correct.[8]

---

[6] Our separations remain valid if we allow any strictly subexponential complexity to be considered efficient, since our lower bounds are exponential.

[7] To keep the notation simple, we've taken the freedom to set the distribution of learning examples to uniform and to make the testing query "distribution-free".

[8] If we allow a slightly modified version of examples, where i is represented through the amplitude as $\sum_{(j,x,i)\in C} (-1)^i |j, x\rangle$, then it is possible to QAC^*-learn $\mathcal{C}$ *exactly* from one such example.

By definition, a student receives from the oracle k copies of the state

$$\sum_{(j,x,i)\in C} |j,x,i\rangle\,,$$

where $C \in \mathcal{C}$ is the target concept to be learnt (recall that we ignore global phases and normalization factors of quantum states). First of all, the student measures the last register of each copy in the basis $\{|0\rangle + |1\rangle\,, |0\rangle - |1\rangle\}$. With probability $1 - 1/2^k$ at least one measurement results in $|0\rangle - |1\rangle$, then the student abandons all other copies (otherwise he halts and gives up).

Next, the student measures the second register in the computational basis, thus obtaining in the first two registers

$$\sum_{(j,x_0,i)\in C} (-1)^i |j,x_0\rangle = \sum_{j\in[N-1]} (-1)^{C_{x_0}\oplus C_{x_0+j}} |j,x_0\rangle = \sum_{j\in[N-1]} (-1)^{C_{x_0+j}} |j,x_0\rangle$$

for some $x_0 \in \mathbb{Z}_N$. Then he performs the transformation $|j,x_0\rangle \to |j+x_0,x_0\rangle$, and the leftmost register of the state becomes

$$|\alpha_{x_0}\rangle \stackrel{\text{def}}{=} \sum_{j\in[N-1]} (-1)^{C_{x_0+j}} |x_0+j\rangle = \sum_{k\in\mathbb{Z}_N\setminus\{x_0\}} (-1)^{C_k} |k\rangle\,.$$

At this point the student is ready to face the testing phase. Assume that a question $q \in [N-1]$ has been asked. Define the following perfect matching over $\mathbb{Z}_N \setminus \{x_0\}$:

$$m_q \stackrel{\text{def}}{=} \left\{ \left(x_0 + (2i+1)q, x_0 + (2i+2)q\right) \middle| 0 \le i \le \frac{N-3}{2} \right\}.$$

Observe that pairwise disjointness of the edges and the fact that x_0 is isolated follow from primality of N. Next, the student performs projective measurement of $|\alpha_{x_0}\rangle$ onto $(N-1)/2$ subspaces, each spanned by a pair of vectors $|a\rangle$ and $|b\rangle$ where a and b are connected in m_q (to make the measurement complete we add $|x_0\rangle\langle x_0|$ to it – this outcome never occurs, due to the structure of $|\alpha_{x_0}\rangle$).

Assume that the outcome of the last measurement corresponds to the edge $(a, a+q) \in m_q$. Then the state of the register that contained $|\alpha_{x_0}\rangle$ becomes either $|a\rangle + |a+q\rangle$ or $|a\rangle - |a+q\rangle$ – the former corresponding to the situation when $C_a \oplus C_{a+q} = 0$ and the latter to $C_a \oplus C_{a+q} = 1$. As two states are orthogonal, the student is able to answer $(a, 0)$ in the first case and $(a, 1)$ in the second case, which is a correct answer to the query q. Observe that all quantum operations involved in the algorithm can be performed efficiently.

3.2 $\mathcal{C}$ is unspeakable

Assume that $\mathcal{C}$ can be approximated by a class $\mathcal{D}$. Then there exists some $h_0 \in \mathcal{D}$ that approximates at least $2^N / |\mathcal{D}|$ elements of $\mathcal{C}$ – denote those elements by $\mathcal{C}_0$.

Consider the[9] answers that h_0 gives to all possible queries $q \in [N-1]$. Denote $(x_q, i_q) \stackrel{\text{def}}{=} h_0(q)$ and

$$Q_0 \stackrel{\text{def}}{=} \left\{ q \middle| (x_q, i_q) \text{ is a good answer to } q \text{ w.r.t. at least } 3/5\text{'th of } \mathcal{C}_0\text{'s elements} \right\}.$$

[9]For notational simplicity we assume that the hypotheses are deterministic. If we allow randomized hypothesis classes, our results remain valid due to the MinMax Theorem.

A simple counting argument implies that $|Q_0| \geq \frac{N-1}{6}$.

Let $e_q \stackrel{\text{def}}{=} (x_q, x_q + q)$ and $E_0 \stackrel{\text{def}}{=} \left\{e_q \middle| q \in Q_0\right\}$. Every edge e_q corresponds to at most 2 different values of $q \in [N-1]$, therefore $|E_0| \geq \frac{N-1}{12}$. Consider a graph G_0 over N nodes, whose edges are the elements of E_0. Observe that G_0 contains at least $\sqrt{2|E_0|} \geq \sqrt{\frac{N-1}{6}}$ non-isolated vertices.

Let $F_0 \subseteq G_0$ be a forest consisting of a spanning tree for each connected component of G_0. Then F_0 contains at least $\sqrt{\frac{N-1}{24}}$ edges, denote them by E_0'. Let $Q_0' \subseteq Q_0$ be a subset of size exactly $|E_0'|$, such that

$$E_0' = \left\{e_q \middle| q \in Q_0'\right\}.$$

View the elements of $\mathcal{C}$ as binary strings of length N. Let us consider two probability distributions, one corresponding to uniformly choosing $C \in \mathcal{C}$ and the other corresponding to uniformly choosing $C \in \mathcal{C}_0$ – denote them by D^C and D_0^C, respectively. Then

$$\log\left(\frac{|\mathcal{C}|}{|\mathcal{C}_0|}\right) = \mathbf{H}\left[D^C\right] - \mathbf{H}\left[D_0^C\right],$$

where $\mathbf{H}[\cdot]$ denotes the binary entropy.

Let us view C as a random variable, distributed either by D^C or by D_0^C. For every $e_q = (a,b)$ let $I_q \stackrel{\text{def}}{=} C_a \oplus C_b$. It is straightforward from the construction of Q_0' that if $C \sim D^C$ then the collection $\left\{I_q \middle| q \in Q_0'\right\}$ consists of mutually independent unbiased Boolean random variables.

Let $J \stackrel{\text{def}}{=} (I_q)_{q \in Q_0'}$ be a random string of length $|Q_0'|$. Then

$$\mathbf{H}[C] = \mathbf{H}[J] + \mathbf{H}\left[C \middle| J\right]$$

is true for any distribution of C, and

$$\begin{aligned}
\log\left(\frac{|\mathcal{C}|}{|\mathcal{C}_0|}\right) &= \mathbf{H}\left[D^C\right] - \mathbf{H}\left[D_0^C\right] = \mathop{\mathbf{H}}_{D^C}[J] - \mathop{\mathbf{H}}_{D_0^C}[J] + \mathop{\mathbf{H}}_{D^C}\left[C \middle| J\right] - \mathop{\mathbf{H}}_{D_0^C}\left[C \middle| J\right] \\
&\geq \mathop{\mathbf{H}}_{D^C}[J] - \mathop{\mathbf{H}}_{D_0^C}[J] = \left|Q_0'\right| - \mathop{\mathbf{H}}_{D_0^C}[J] \\
&\geq \left|Q_0'\right| - \sum_{q \in Q_0'} \mathop{\mathbf{H}}_{D_0^C}[I_q] = \sum_{q \in Q_0'}\left(1 - \mathop{\mathbf{H}}_{D_0^C}[I_q]\right),
\end{aligned} \tag{1}$$

where the first inequality follows from the fact that $\mathbf{H}_{D^C}\left[C \middle| J\right] = N - |Q_0'|$, which is the maximum of $\mathbf{H}\left[C \middle| J\right]$ under any distribution of C.

From the definition of Q_0 (and the fact that $Q_0' \subseteq Q_0$), we know that each of $\left\{I_q \middle| q \in Q_0'\right\}$ is at least 3/5-biased, therefore $\mathbf{H}_{D_0^C}[I_q] \leq \frac{49}{50}$, and (1) leads to

$$\log\left(\frac{|\mathcal{C}|}{|\mathcal{C}_0|}\right) \geq \frac{|Q_0'|}{50} = \frac{|E_0'|}{50} > \frac{\sqrt{N}}{250}$$

for sufficiently large N. According to our choice of h_0,

$$|\mathcal{D}| \geq \frac{|\mathcal{C}|}{|\mathcal{C}_0|} \in 2^{N^{\Omega(1)}} \in 2^{2^{\Omega(n)}},$$

which means that the concept class $\mathcal{C}$ is unspeakable.

4 Unspeakable functions cannot be learnt efficiently

We claim that no unspeakable *functional* concept class can be efficiently learnt even in a predictive learning model, either quantum or classical. More formally:

Lemma 4.1. *Predictive learning of an unspeakable functional concept class, even by a computationally unlimited student, requires exponential amount of quantum (w.l.g.) information from the teacher.*

Proof. Assume that for some functional concept class $\mathcal{F}$ that is unspeakable, the following holds. A teacher T knows some $f_0 \in \mathcal{F}$, hidden from a student S. Then T exchange at most k_q qubits with S. Finally, S is given some x_0 from the common domain D of the functions in $\mathcal{F}$, and is able to compute $f_0(x_0)$ with high confidence.

Let us consider the following communication task $\mathcal{G}$ in the model of *1-way communication.* Alice receives $f_0 \in \mathcal{F}$ and Bob receives $x_0 \in D$. Alice is allowed to send a single message to Bob, in order to let him compute $f_0(x_0)$ with high confidence. By our assumption regarding learnability of $\mathcal{F}$, it is enough for Alice to send k_q qubits to Bob. Indeed, even though in the learning protocol interaction between S and T is allowed, S possesses no information that would be hidden from T, and so if T is willing to cooperate then a single message from T to S is sufficient. To conclude, the quantum 1-way communication cost of the problem $\mathcal{G}$ is at most k_q.

Let k_c denote the cost of $\mathcal{G}$ in the model of 1-way classical communication. As $\mathcal{F}$ is unspeakable, $k_c \in 2^{\Omega(n)}$. It has been shown by Aaronson [A04] that classical 1-way communication cost of a *functional*[10] communication problem is at most the length of Bob's input times $k_q \log(k_q)$. The bit-length of Bob's input x_0 is n, and

$$k_q \geq \frac{k_c}{n \log(k_q)} \in 2^{\Omega(n)},$$

as required. ∎

5 Open problems

We have demonstrated that efficient quantum predictive learning of an unspeakable relational concept class is possible. The following questions seem interesting.

When we demonstrated the limitations of quantum quasi-predictive learning (in the proof of Lemma 2.1), we argued that certain "quasi-hypothesis" of polynomial length can be extracted from an efficient quantum quasi-predictive learning algorithm. But in fact, our construction does not rely upon the efficiency of the learning algorithm, and on the other hand, the quasi-hypothesis we construct cannot, in general, be efficiently evaluated. It would be interesting to come up with a stronger argument that would "preserve efficiency"; or otherwise, to get an example of a non-trivial quantum quasi-predictive learning algorithm. Similar observations can be made w.r.t. our proof of Lemma 4.1. The transformation in [A04] is, in general, not efficient. Are there non-trivial quantum predictive (or even quasi-predictive) learning algorithm for functional concepts?

[10] The same is not true for relational problems – which makes our main result possible – cf. [GRW08].

In the above questions by "non-trivial" we mean a quantum algorithm for learning a concept class that admits concise hypotheses, but only those that cannot be efficiently evaluated. Observe that such "quasi-unspeakable" concept classes cannot be learnt efficiently in any reasonable classical model (due to the fact that in the classical setting the equivalence between standard and predictive learning is efficiency-preserving). In fact, it might be the case that reasonable modifications of the known *conditional* separations between quantum and classical learning can lead to efficient quantum learnability of a concept class that is quasi-unspeakable under (non-uniform version of) the original hardness assumptions.

More generally, give new examples of efficient quantum (quasi-)predictive learning of concept classes which are not efficiently learnable classically. Such examples might be interesting even for models stronger than QAC^* (e.g., one may allow the student to make *membership queries*).

References

[A04] S. Aaronson. Limitations of Quantum Advice and One-Way Communication. *Proceedings of the 19th IEEE Conference on Computational Complexity, pages 320-332*, 2004.

[BJ95] N. Bshouty and J. Jackson. Learning DNF over the Uniform Distribution using a Quantum Example Oracle. *Proceedings of the 8th Annual Conference on Computational Learning Theory, pages 118-127*, 1995.

[BJK04] Z. Bar-Yossef, T. S. Jayram and I. Kerenidis. Exponential Separation of Quantum and Classical One-Way Communication Complexity. *Proceedings of 36th Symposium on Theory of Computing, pages 128-137*, 2004.

[GRW08] D. Gavinsky, O. Regev and R. de Wolf. Simultaneous Communication Protocols with Quantum and Classical Messages. *http://arxiv.org/abs/quant-ph/0807.2758*, 2008.

[SG04] R. Servedio and S. Gortler. Equivalences and Separations Between Quantum and Classical Learnability. *SIAM Journal on Computing 33(5), pages 1067-1092*, 2004.

[V84] L. Valiant. A Theory of Learnable. *Communications of the ACM 27(11), pages 1134-1142*, 1984.