Title: The Learnability of Quantum States

Date: Sep 20, 2006 02:00 PM

URL: http://pirsa.org/06090011

Abstract: Traditional quantum state tomography requires a number of measurements that grows exponentially with the number of qubits n. But using ideas from computational learning theory, I'll show that "for most practical purposes" one can learn a quantum state using a number of measurements that grows only linearly with n. I'll discuss applications of this result in experimental physics and quantum computing theory, as well as possible implications for the foundations of quantum mechanics. quant-ph/0608142

The Learnability of Quantum States

Scott Aaronson
University of Waterloo

Quantum State Tomography

Suppose we have a physical process that produces as many copies as we like of a quantum state ρ

Quantum State Tomography

Suppose we have a physical process that produces as many copies as we like of a quantum state ρ

To each copy, we can apply a two-outcome measurement E, which yields '1' with probability $\mathrm{Tr}(E\rho)$ and '0' otherwise

Quantum State Tomography

Suppose we have a physical process that produces as many copies as we like of a quantum state ρ

To each copy, we can apply a two-outcome measurement E, which yields '1' with probability $\mathrm{Tr}(E\rho)$ and '0' otherwise

Our goal is to learn an approximate description of ρ, by combining the various measurement outcomes

Quantum State Tomography

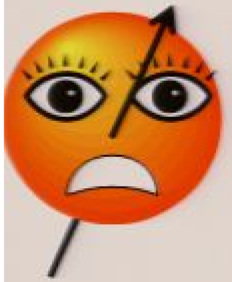
Suppose we have a physical process that produces as many copies as we like of a quantum state ρ

To each copy, we can apply a two-outcome measurement E, which yields '1' with probability $\mathrm{Tr}(E\rho)$ and '0' otherwise

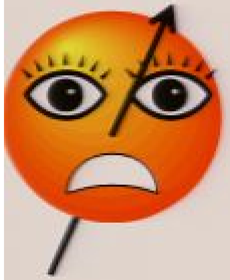
Our goal is to learn an approximate description of ρ, by combining the various measurement outcomes

EXPERIMENTALISTS ACTUALLY DO THIS

To learn about chemical reactions (Skovsen et al. 2003), test equipment (D'Ariano et al. 2002), study decoherence mechanisms (Resch et al. 2005), …

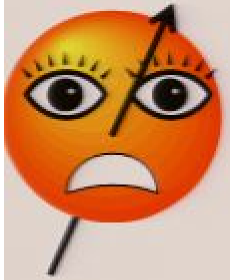


But there's a problem…



But there's a problem…

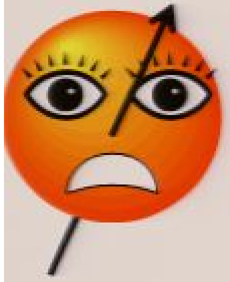
To do tomography on an entangled state of n qubits, we need $\Omega(4^n)$ measurements



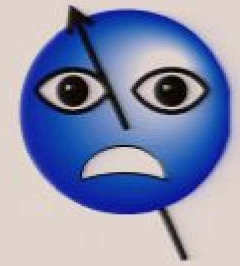
But there's a problem…

To do tomography on an entangled state of n qubits, we need $\Omega(4^n)$ measurements

The current record: 8 qubits (Häffner et al. 2005), requiring 656,100 experiments (!)



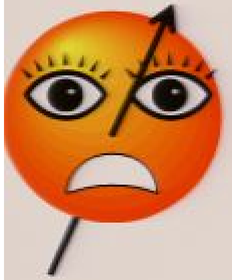
But there's a problem…



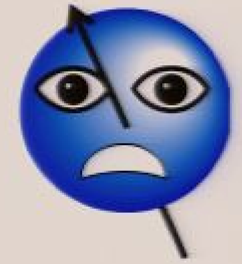
To do tomography on an entangled state of n qubits, we need $\Omega(4^n)$ measurements

The current record: 8 qubits (Häffner et al. 2005), requiring 656,100 experiments (!)

Does this mean that a generic state of (say) 10,000 particles can never be "learned" within the lifetime of the universe?



But there's a problem…



To do tomography on an entangled state of n qubits, we need $\Omega(4^n)$ measurements

The current record: 8 qubits (Häffner et al. 2005), requiring 656,100 experiments (!)

Does this mean that a generic state of (say) 10,000 particles can never be "learned" within the lifetime of the universe?

If so, this is certainly a practical problem—but to me, it's a **conceptual** problem as well

What **is** a quantum state?

What **is** a quantum state?

A “state of the world”? A “state of knowledge”?

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

How "useful" is a hypothesis that takes 10^{5000} bits even to write down? (E.g., "generic" many-particle entangled states)

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

How "useful" is a hypothesis that takes 10^{5000} bits even to write down? (E.g., "generic" many-particle entangled states)

Seems to bolster the arguments of quantum computing skeptics (Goldreich, Levin, 't Hooft, Wolfram, Penrose?), who think quantum mechanics will break down in the "large N limit"

Really we're talking about the Humean Problem of Induction…

You see 500 ravens. Every one is black. Why does that give you **any grounds whatsoever** for expecting the next raven to be black?

?

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

How "useful" is a hypothesis that takes 10^{5000} bits even to write down? (E.g., "generic" many-particle entangled states)

Seems to bolster the arguments of quantum computing skeptics (Goldreich, Levin, 't Hooft, Wolfram, Penrose?), who think quantum mechanics will break down in the "large N limit"

Really we're talking about the Humean Problem of Induction…

You see 500 ravens. Every one is black. Why does that give you **any grounds whatsoever** for expecting the next raven to be black?

?

Really we're talking about the Humean Problem of Induction…

You see 500 ravens. Every one is black. Why does that give you **any grounds whatsoever** for expecting the next raven to be black?

The answer, according to computational learning theory: In practice, we always restrict attention to some class of hypotheses vastly smaller than the class of all logically conceivable hypotheses

Probably Approximately Correct (PAC) Learning

Set S called the **sample space**

Probability distribution D over S

Class C of **hypotheses**: functions from S to {0,1}

Unknown function $f \in C$

Goal: Given $x_1,\ldots,x_m$ drawn independently from D, together with $f(x_1),\ldots,f(x_m)$, output a hypothesis $h \in C$ such that

$$\Pr_{x \in D}\left[h(x) = f(x)\right] \geq 1 - \varepsilon,$$

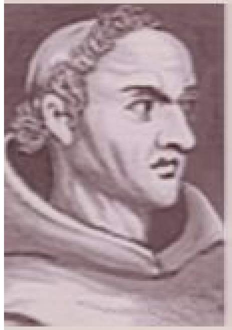
with probability at least $1-\delta$ over $x_1,\ldots,x_m$

Occam's Razor Theorem

Valiant 1984: If the hypothesis class C is finite, then any hypothesis consistent with

$$m = O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples



Occam's Razor Theorem

Valiant 1984: If the hypothesis class C is finite, then any hypothesis consistent with

$$m = O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

"Compression implies prediction"

Occam's Razor Theorem

Va… …thesis class C is finite, then any
b…

But the number of quantum states is infinite!

$$O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

"Compression implies prediction"

Occam's Razor Theorem

Va… …thesis class C is finite, then any h…

But the number of quantum states is infinite!

$O\left(\frac{1}{\varepsilon}\log \dots |C|\right)$

And even if we discretize, it's still **doubly** exponential in the number of qubits!

random samples will als…
fraction of future data, w…
the choice of samples

"Compression implies prediction"

My Result: A Quantum Occam's Razor Theorem

Let ρ be an n-qubit state. Let D be a distribution over two-outcome measurements. Suppose we draw m measurements $E_1,\ldots,E_m$ independently from D, and then output a "hypothesis state" σ such that $|\mathrm{Tr}(E_i\sigma)-\mathrm{Tr}(E_i\rho)|\leq\eta$ for all i. Then provided $\eta\leq\gamma\varepsilon/10$ and

$$m \geq \frac{K}{\gamma^2\varepsilon^2}\left(\frac{n}{\gamma^2\varepsilon^2}\log\frac{1}{\gamma\varepsilon}+\log\frac{1}{\delta}\right)$$

(for some constant K), we'll have

$$\Pr_{E\in D}\left[\left|\mathrm{Tr}(E\sigma)-\mathrm{Tr}(E\rho)\right|\leq\gamma\right]\geq 1-\varepsilon$$

with probability at least 1-δ over $E_1,\ldots,E_m$

Some Examples

If the distribution D over measurements is uniform (i.e., is the Haar measure), then the maximally mixed state works perfectly well as an “explanatory hypothesis”

Some Examples

If the distribution D over measurements is uniform (i.e., is the Haar measure), then the maximally mixed state works perfectly well as an "explanatory hypothesis"

If the distribution is concentrated on 1- and 2-qubit measurements, then we don't see much training data about many-particle entanglement, but we don't need it either

How do we actually find σ?

How do we actually find σ?

Here's one way: let $b_1, \ldots, b_m$ be the binary outcomes of measurements $E_1, \ldots, E_m$

Then choose a hypothesis state σ to minimize

$$\sum_{i=1}^{m} \left(\mathrm{Tr}\left(E_i \sigma\right) - b_i \right)^2$$

How do we actually find σ?

Here's one way: let $b_1,\ldots,b_m$ be the binary outcomes of measurements $E_1,\ldots,E_m$

Then choose a hypothesis state σ to minimize

$$\sum_{i=1}^{m}\left(\mathrm{Tr}\left(E_i\sigma\right)-b_i\right)^2$$

This is a convex programming problem, which can be solved in time polynomial in $N=2^n$ (probably good enough in practice for $n\leq15$ or so)

Previous Approach to “Pretty Good” Quantum State Tomography
(Bužek et al.)

How do we actually find σ?

Here's one way: let $b_1, \ldots, b_m$ be the binary outcomes of measurements $E_1, \ldots, E_m$

Then choose a hypothesis state σ to minimize

$$\sum_{i=1}^{m} \left(\mathrm{Tr}(E_i \sigma) - b_i \right)^2$$

This is a convex programming problem, which can be solved in time polynomial in $N=2^n$ (probably good enough in practice for $n \leq 15$ or so)

Previous Approach to “Pretty Good” Quantum State Tomography

(Bužek et al.)

Previous Approach to "Pretty Good" Quantum State Tomography
(Bužek et al.)

(1) Assume a uniform prior over pure states

(2) Perform measurements

(3) Update the prior using Bayes' rule

Previous Approach to “Pretty Good” Quantum State Tomography
(Bužek et al.)

(1) Assume a uniform prior over pure states

(2) Perform measurements

(3) Update the prior using Bayes’ rule

Disadvantages:

- Staggering computational complexity
- Sensitive to choice of prior

Previous Approach to "Pretty Good" Quantum State Tomography

(Bužek et al.)

(1) Assume a uniform prior over pure states

(2) Perform measurements

(3) Update the prior using Bayes' rule

Disadvantages:

- Staggering computational complexity
- Sensitive to choice of prior

In the learning approach, we don't need a prior over states—just a prior over measurements

To prove the theorem, we need a notion introduced by Kearns and Schapire called

Fat-Shattering Dimension

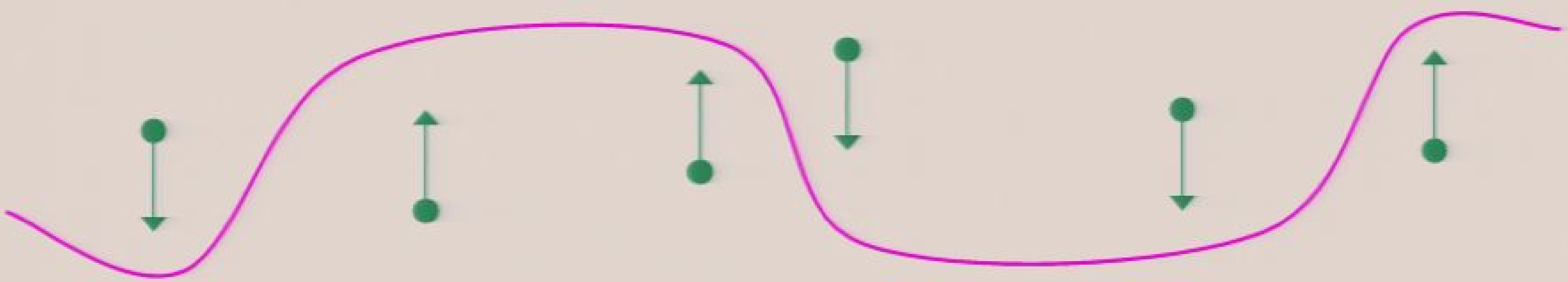
To prove the theorem, we need a notion introduced by Kearns and Schapire called

Fat-Shattering Dimension

Let C be a class of functions from S to [0,1]. We say a set $\{x_1,\ldots,x_k\} \subseteq S$ is **γ-shattered** by C if there exist reals $a_1,\ldots,a_k$ such that, for all 2^k possible statements of the form

$$f(x_1) \leq a_1 - \gamma \wedge f(x_2) \geq a_2 + \gamma \wedge \ldots \wedge f(x_k) \leq a_k - \gamma,$$

there's some $f \in C$ that satisfies the statement.



Then $fat_C(\gamma)$, the **γ-fat-shattering dimension** of C, is the size of the largest set γ-shattered by C.

Small Fat-Shattering Dimension Implies Small Sample Complexity

Let C be a class of functions from S to [0,1], and let $f \in C$. Suppose we draw m elements $x_1, \ldots, x_m$ independently from some distribution D, and then output a hypothesis $h \in C$ such that $|h(x_i)-f(x_i)| \leq \eta$ for all i. Then provided $\eta \leq \gamma\varepsilon/7$ and

$$m = \Omega\left(\frac{1}{\gamma^2\varepsilon^2}\left(\mathrm{fat}_C\left(\frac{\gamma\varepsilon}{35}\right)\log^2\frac{1}{\gamma\varepsilon} + \log\frac{1}{\delta}\right)\right),$$

we'll have

$$\Pr_{x \in D}\left[\left|h(x) - f(x)\right| \leq \gamma\right] \geq 1 - \varepsilon$$

with probability at least $1-\delta$ over $x_1, \ldots, x_m$.

Small Fat-Shattering Dimension Implies Small Sample Complexity

Let C be a class of functions from S to [0,1], and let $f \in C$. Suppose we draw m elements $x_1, \ldots, x_m$ independently from some distribution D, and then output a hypothesis $h \in C$ such that $|h(x_i)-f(x_i)| \leq \eta$ for all i. Then provided $\eta \leq \gamma\varepsilon/7$ and

$$m = \Omega\left(\frac{1}{\gamma^2\varepsilon^2}\left(\mathrm{fat}_C\left(\frac{\gamma\varepsilon}{35}\right)\log^2\frac{1}{\gamma\varepsilon} + \log\frac{1}{\delta}\right)\right),$$

we'll have

$$\Pr_{x \in D}\left[\left|h(x) - f(x)\right| \leq \gamma\right] \geq 1 - \varepsilon$$

with probability at least $1-\delta$ over $x_1, \ldots, x_m$.

Proof uses a 1996 result of Bartlett and Long, building on Alon et al. 1993, building on Blumer et al. 1989

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Small Fat-Shattering Dimension Implies Small Sample Complexity

Let C be a class of functions from S to [0,1], and let $f \in C$. Suppose we draw m elements $x_1,\ldots,x_m$ independently from some distribution D, and then output a hypothesis $h \in C$ such that $|h(x_i)-f(x_i)| \leq \eta$ for all i. Then provided $\eta \leq \gamma\varepsilon/7$ and

$$m = \Omega\left(\frac{1}{\gamma^2\varepsilon^2}\left(\mathrm{fat}_C\left(\frac{\gamma\varepsilon}{35}\right)\log^2\frac{1}{\gamma\varepsilon} + \log\frac{1}{\delta}\right)\right),$$

we'll have

$$\Pr_{x \in D}\left[\left|h(x) - f(x)\right| \leq \gamma\right] \geq 1 - \varepsilon$$

with probability at least $1-\delta$ over $x_1,\ldots,x_m$.

Proof uses a 1996 result of Bartlett and Long, building on Alon et al. 1993, building on Blumer et al. 1989

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Nayak 1999: If we want to "encode" k classical bits into n qubits, in such a way that any bit can be recovered with probability 1-p, then we need $n \geq (1-H(p))k$

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Nayak 1999: If we want to "encode" k classical bits into n qubits, in such a way that any bit can be recovered with probability 1-p, then we need $n \geq (1-H(p))k$

Corollary ("turning Nayak's result on its head"):
Let C_n be the set of functions that map an n-qubit measurement E to $\mathrm{Tr}(E\rho)$, for some ρ. Then

$$\mathrm{fat}_{C_n}(\gamma) = O\left(\frac{n}{\gamma^2}\right).$$

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Nayak 1999: If we want to "encode" k classical bits into n qubits, in such a way that any bit can be recovered with probability 1-p, then we need n≥(1-H(p))k

Corollary ("turning Nayak's result on its head"):
Let C_n be the set of functions that map an n-qubit measurement E to Tr(Eρ), for some ρ. Then

$$\mathrm{fat}_{C_n}(\gamma) = O\left(\frac{n}{\gamma^2}\right).$$

Quantum Occam's Razor Theorem follows easily…

Upper-Bounding the Fat-Shattering Dimension of Quantum States

Nayak 1999: If we want to "encode" k classical bits into n qubits, in such a way that any bit can be recovered with probability 1-p, then we need n≥(1-H(p))k

Corollary ("turning Nayak's result on its head"):
Let C_n be the set of functions that map an n-qubit meas... ρ), for some ρ. Then

$$\ldots\mathrm{at}_{C_n}(\gamma)=O\left(\frac{n}{\gamma^2}\right).$$

Quantum Occam's Razor Theorem follows easily…

Application to Quantum Computing:
Simulating Quantum One-Way Protocols

Application to Quantum Computing:
Simulating Quantum One-Way Protocols

Alice has an N-bit string x. Bob has a M-bit string y.
Together they want to evaluate a Boolean function f(x,y).
Only **one-way communication** from Alice to Bob is allowed.

Application to Quantum Computing:
Simulating Quantum One-Way Protocols

Alice has an N-bit string x. Bob has a M-bit string y.
Together they want to evaluate a Boolean function f(x,y).
Only **one-way communication** from Alice to Bob is allowed.

Theorem: The number of bits Alice needs to send Bob in a classical probabilistic protocol, is (up to a constant) at most M times the number of qubits she needs to send quantumly

Application to Quantum Computing: Simulating Quantum One-Way Protocols

Alice has an N-bit string x. Bob has a M-bit string y. Together they want to evaluate a Boolean function f(x,y). Only **one-way communication** from Alice to Bob is allowed.

Theorem: The number of bits Alice needs to send Bob in a classical probabilistic protocol, is (up to a constant) at most M times the number of qubits she needs to send quantumly

Intuition: In the classical protocol, first Alice sends random inputs $y_1,\ldots,y_T$, together with $f(x,y_1),\ldots,f(x,y_T)$. Then Bob searches for a quantum message ρ from Alice consistent with those $f(x,y_i)$ values. By the Quantum Occam's Razor Theorem, such a ρ (once he finds it) will probably yield the right outputs for most other y's as well

Application to Quantum Computing:
Using Trusted Classical Data to Verify an Untrusted Quantum State

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Application to Quantum Computing:
Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Theorem: There exists a set of "benchmark inputs" $x_1,\ldots,x_T$, where T=poly(n), such that if $|\psi\rangle$ works on the benchmark inputs, it will work on most other inputs as well

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Theorem: There exists a set of "benchmark inputs" $x_1,\ldots,x_T$, where T=poly(n), such that if $|\psi\rangle$ works on the benchmark inputs, it will work on most other inputs as well

Intuition: Again the Quantum Occam's Razor Theorem

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Theorem: There exists a set of "benchmark inputs" $x_1,\ldots,x_T$, where T=poly(n), such that if $|\psi\rangle$ works on the benchmark inputs, it will work on most other inputs as well

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Theorem: There exists a set of "benchmark inputs" $x_1,\ldots,x_T$, where T=poly(n), such that if $|\psi\rangle$ works on the benchmark inputs, it will work on most other inputs as well

Intuition: Again the Quantum Occam's Razor Theorem

Application to Quantum Computing: Using Trusted Classical Data to Verify an Untrusted Quantum State

At the quantum software store, you buy an n-qubit **quantum program** $|\psi\rangle$ to give your quantum computer new functionality

But you don't trust the software to work as expected

Theorem: There exists a set of "benchmark inputs" $x_1,\ldots,x_T$, where T=poly(n), such that if $|\psi\rangle$ works on the benchmark inputs, it will work on most other inputs as well

Intuition: Again the Quantum Occam's Razor Theorem

Technical part: How to test $|\psi\rangle$ on the benchmark inputs without destroying it?

Open Problems

Open Problems

Computationally-efficient learning algorithms

Open Problems

Computationally-efficient learning algorithms

Experimental implementation!

Open Problems

Computationally-efficient learning algorithms

Experimental implementation!

Tighter bounds on measurement complexity

Open Problems

Computationally-efficient learning algorithms

Experimental implementation!

Tighter bounds on measurement complexity

Further applications to quantum computing

Open Problems

Computationally-efficient learning algorithms

Experimental implementation!

Tighter bounds on measurement complexity

Further applications to quantum computing

Derive quantum theory from learnability?

Occam's Razor Theorem

Valiant 1984: If the hypothesis class C is finite, then any hypothesis consistent with

$$m = O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

"Compression implies prediction"

Open Problems

Computationally-efficient learning algorithms

Experimental implementation!

Tighter bounds on measurement complexity

Further applications to quantum computing

Derive quantum theory from learnability?