Strengths and Weaknesses of Quantum Maching Learning

Srinivasan Arunachalam

Centrum Wiskunde & Informatica

Research Center for Quantum Software

Machine learning

- Grand goal: enable AI systems to improve themselves

Machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data

Machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at

Machine learning

- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing

Machine learning

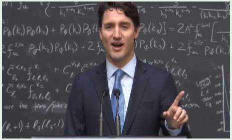
- Grand goal: enable AI systems to improve themselves
- Practical goal: learn "something" from given data
- Recent success: deep learning is extremely good at image recognition, natural language processing, even the game of Go

Quantum learning

Quantum learning

Quantum computing: 2 minute crash course

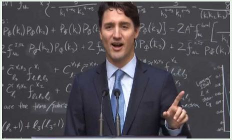
YouTube: “Canadian Prime Minister Justin Trudeau schools reporter”



Quantum learning

Quantum computing: 2 minute crash course

YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"

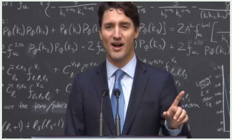


- Classically, a bit can be either 0 *or* 1

Quantum learning

Quantum computing: 2 minute crash course

YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"

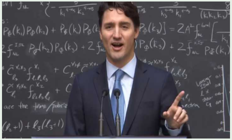


- Classically, a bit can be either 0 *or* 1
- Superposition: Qubit can be 0 *and* 1, each with an *amplitude*

Quantum learning

Quantum computing: 2 minute crash course

YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"

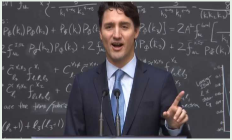


- Classically, a bit can be either 0 *or* 1
- Superposition: Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$

Quantum learning

Quantum computing: 2 minute crash course

YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"

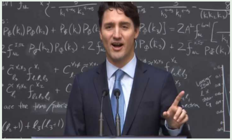


- Classically, a bit can be either 0 *or* 1
- Superposition: Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: "probability of outcome = amplitude^2"

Quantum learning

Quantum computing: 2 minute crash course

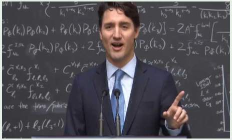
YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"



- Classically, a bit can be either 0 *or* 1
- Superposition: Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: "probability of outcome = amplitude^2"
  - Obtain $|0\rangle$ with probability $1/6$. $|\psi\rangle$ is lost

Quantum computing: 2 minute crash course

YouTube: "Canadian Prime Minister Justin Trudeau schools reporter"



- Classically, a bit can be either 0 *or* 1
- Superposition: Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: "probability of outcome = amplitude^2"
  - Obtain $|0\rangle$ with probability $1/6$. $|\psi\rangle$ is lost
  - Obtain $|1\rangle$ with probability $5/6$. $|\psi\rangle$ is lost

Quantum learning

Quantum computing: 2 minute crash course

- Classically, a bit can be either 0 *or* 1
- Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: Obtain $|0\rangle$ w.p. 1/6, $|1\rangle$ w.p. 5/6. $|\psi\rangle$ is lost

What can quantum computing do for machine learning?

Quantum learning

Quantum computing: 2 minute crash course

- Classically, a bit can be either 0 *or* 1
- Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: Obtain $|0\rangle$ w.p. 1/6, $|1\rangle$ w.p. 5/6. $|\psi\rangle$ is lost

What can quantum computing do for machine learning?

- The learner may be quantum

Quantum learning

Quantum computing: 2 minute crash course

- Classically, a bit can be either 0 *or* 1
- Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: Obtain $|0\rangle$ w.p. 1/6, $|1\rangle$ w.p. 5/6. $|\psi\rangle$ is lost

What can quantum computing do for machine learning?

- The learner may be quantum
- The data can also be quantum

Quantum learning

Quantum computing: 2 minute crash course

- Classically, a bit can be either 0 *or* 1
- Qubit can be 0 *and* 1, each with an *amplitude*
- Example: $|\psi\rangle = \frac{1}{\sqrt{6}}|0\rangle - \sqrt{\frac{5}{6}}|1\rangle$
- Measurement of $|\psi\rangle$: Obtain $|0\rangle$ w.p. 1/6, $|1\rangle$ w.p. 5/6. $|\psi\rangle$ is lost

What can quantum computing do for machine learning?

- The learner may be quantum
- The data can also be quantum
- Some examples are known of reduction in time complexity:
  - k-means clustering
  - principal component analysis
  - perceptron learning
  - recommendation systems

Probably Approximately Correct (PAC) learning

Probably Approximately Correct (PAC) learning

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)

Probably Approximately Correct (PAC) learning

Basic definitions

- Concept class $\mathcal{C}$: collection of Boolean functions on n bits (Known)
- Target Concept c: some function $c \in \mathcal{C}$ (Unknown)

Probably Approximately Correct (PAC) learning

Basic definitions

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

Probably Approximately Correct (PAC) learning

Basic definitions

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

Probably Approximately Correct (PAC) learning

Basic definitions

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

$\mathcal{C}$

$\downarrow$

c

target concept

Learner is trying to learn c

Probably Approximately Correct (PAC) learning

Basic definitions

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

$$\mathcal{C}$$

$$\downarrow$$

c
target concept

$x_1 \sim D \quad \longrightarrow \quad (x_1, c(x_1))$ 

Learner is trying to learn c

Probably Approximately Correct (PAC) learning

Basic definitions

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

$\mathcal{C}$

$\downarrow$

c
target concept

$$x_1 \sim D \quad \longrightarrow \quad (x_1, c(x_1))$$
$$x_2 \sim D \quad \longrightarrow \quad (x_2, c(x_2))$$

Learner is trying to learn c

Probably Approximately Correct (PAC) learning

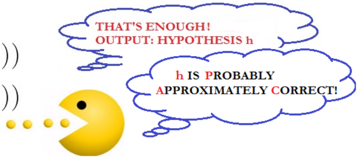
Basic definitions

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

$$\mathcal{C} \downarrow c$$

target concept

$$\begin{array}{ccc} x_1 \sim D & \longrightarrow & (x_1, c(x_1)) \\ x_2 \sim D & \longrightarrow & (x_2, c(x_2)) \\ \vdots & & \vdots \\ x_T \sim D & \longrightarrow & (x_T, c(x_T)) \end{array}$$



Learner is trying to learn c

Probably Approximately Correct (PAC) learning

Basic definitions

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

Goal of a PAC learner

Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is *Probably Approximately Correct*, i.e.,

Probably Approximately Correct (PAC) learning

Basic definitions

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

Goal of a PAC learner

Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is *Probably Approximately Correct*, i.e.,

> *for every D, for every $c \in \mathcal{C}$, with high probability, the hypothesis h should approximately look like c*

Probably Approximately Correct (PAC) learning

Basic definitions

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

Goal of a PAC learner

Using i.i.d. labeled examples, learner for $\mathcal{C}$ should output hypothesis h that is *Probably Approximately Correct*, i.e.,

> *for every D, for every $c \in \mathcal{C}$, with high probability, the hypothesis h should approximately look like c*

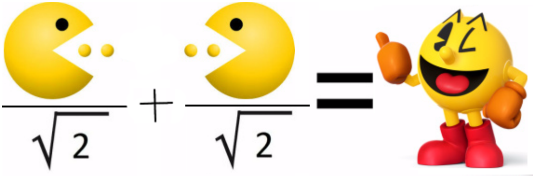
Complexity measure: *Sample complexity*

Minimum number of examples seen by the optimal PAC learner for $\mathcal{C}$

Quantum PAC learning

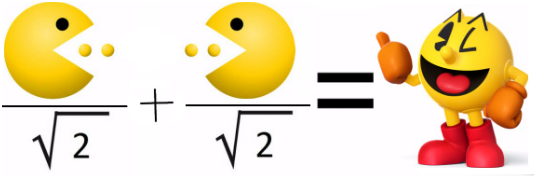
Quantum PAC learning

- Learner is quantum:



Quantum PAC learning

- Learner is quantum:

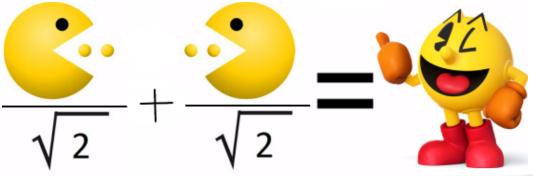


- Data is quantum: Quantum example is a superposition

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

Quantum PAC learning

- Learner is quantum:



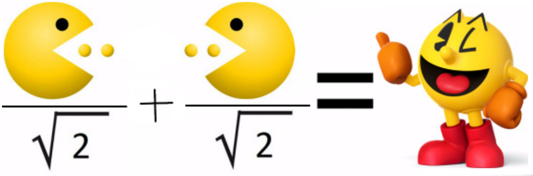
- Data is quantum: Quantum example is a superposition

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

Measuring this state gives a $(x, c(x))$ with probability $D(x)$,

Quantum PAC learning

- Learner is quantum:



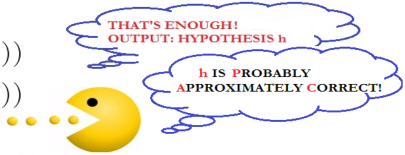
- Data is quantum: Quantum example is a superposition

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

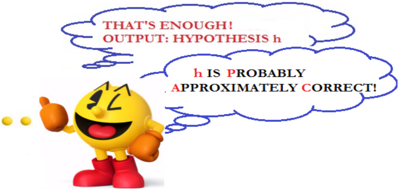
Measuring this state gives a $(x, c(x))$ with probability $D(x)$, so quantum examples are at least as powerful as classical

Quantum PAC learning

$$
\begin{array}{lcl}
x_1 \sim D & \longrightarrow & (x_1, c(x_1)) \\
x_2 \sim D & \longrightarrow & (x_2, c(x_2)) \\
\vdots & & \vdots \\
x_T \sim D & \longrightarrow & (x_T, c(x_T))
\end{array}
$$



$$
\begin{array}{lc}
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle & \longrightarrow \\
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle & \longrightarrow \\
\vdots & \vdots \\
\sum_{x\in\{0,1\}^n} \sqrt{D(x)}\,|x, c(x)\rangle & \longrightarrow
\end{array}
$$



Question

Fewer quantum examples suffice for a quantum learner in the PAC model?

Does quantum provide an advantage for PAC learning?

Vapnik and Chervonenkis (VC) dimension

Combinatorial parameter that can be defined for every concept class $\mathcal{C}$

Does quantum provide an advantage for PAC learning?

Vapnik and Chervonenkis (VC) dimension

Combinatorial parameter that can be defined for every concept class $\mathcal{C}$

Classical bounds

VC dimension characterizes classical sample complexity

Does quantum provide an advantage for PAC learning?

Vapnik and Chervonenkis (VC) dimension

Combinatorial parameter that can be defined for every concept class $\mathcal{C}$

Classical bounds

VC dimension characterizes classical sample complexity

Quantum bounds [Arunachalam, de Wolf'16]

VC dimension characterizes *quantum* sample complexity

Does quantum provide an advantage for PAC learning?

Vapnik and Chervonenkis (VC) dimension

Combinatorial parameter that can be defined for every concept class $\mathcal{C}$

Classical bounds

VC dimension characterizes classical sample complexity

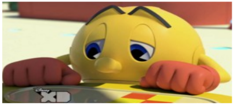
Quantum bounds [Arunachalam, de Wolf'16]

VC dimension characterizes *quantum* sample complexity

Classical sample complexity = Quantum sample complexity

Conclusion and future work

Classical PAC
Sample complexity
=
Quantum PAC
Sample complexity



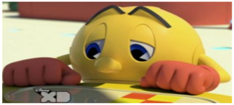
Take home

Quantum examples do not provide an advantage for PAC learning

Classical PAC Sample complexity

=

Quantum PAC Sample complexity



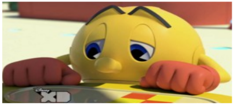
Take home

Quantum examples do not provide an advantage for PAC learning

Future work

Quantum machine learning is still young! Don't have convincing examples where quantum significantly improve machine learning.

Classical PAC
Sample complexity
=
Quantum PAC
Sample complexity



Take home

Quantum examples do not provide an advantage for PAC learning

Future work

Quantum machine learning is still young! Don't have convincing examples where quantum significantly improve machine learning.

Our goal is to find examples where quantum speeds up classical learning.