Welcome 2017 Faculty Summit Attendees

Faculty Summit 2017
microsoftfacultysummit.com

Microsoft Research
Microsoft.com/research

Facebook
@microsoftresearch

Twitter
@MSFTResearch
#FacSumm
#EdgeofAI
Microsoft Research
Faculty Summit 2017

Quantum Neural Networks

Nathan Wiebe
Work with Maria Kieferova
Machine learning

The Ultimate Go Challenge
Game 1 of 5
9 March 2016

AlphaGo vs Lee Sedol

Result: W Res
Number of Moves: 186
Time White: 1h 55m
Time Black: 1h 32m
Quantum Machine Learning

• Fast scalable quantum computers will fundamentally change machine learning.

• Quadratic (or greater) speedups.

• Better models for data.

• Improved privacy and security.

Fantastic speedups and where to find them...

- Support vector machines.
- PCA.
- Approximate Bayesian Inference.
- Recommender Systems.
- Topological Analysis.
- Perceptron training.
- Nearest Neighbor Classification.
- Gradient Descent.
- Quantum Boltzmann Machine Training.
Quantum Machine Learning

• Fast scalable quantum computers will fundamentally change machine learning.

• Quadratic (or greater) speedups.

• Better models for data.

• Improved privacy and security.

Fantastic speedups and where to find them...
Support vector machines.
PCA.
Approximate Bayesian Inference.
Recommender Systems.
Topological Analysis.
Perceptron training.
Nearest Neighbor Classification.
Gradient Descent.

Quantum Boltzmann Machine Training.
Boltzmann Machines

Class of recurrent neural network that is strongly related to physics.

Energy

\[ E(x) = \sum_{<i,j>} w_{i,j} n_i n_j \]

\[ w = \text{argmin}_w (D_{KL}(P(x) \| P_{\text{train}}(x))) \]

\[ E(x) \rightarrow H = \sum_{<i,j>} \frac{w_{i,j}(1 - Z_i)(1 - Z_j)}{4} \]
Quantizing the Boltzmann Machine

Quantum states that mimic training data

Energy function $\sum_j w_j H_j$

Classically we have only energy penalties.

Quantumly, we control both the energies and the eigenbasis the energies are applied in.

Quantum energy operator (Hamiltonian)

KL-divergence between output dist. and training dist.

Quantum relative entropy between distributions.

Training data set

Quantum distribution over training data

Quantum Boltzmann machines learn directly from quantum states. They also generate quantum states that closely resemble the training data.
Training quantum Boltzmann Machines

We train using gradient descent on the quantum relative entropy (quantum KL divergence).

\[
S(\rho||\sigma) = \text{Tr}(\rho \log \rho - \rho \log \sigma) \nonumber \\
\sigma = \frac{e^{-\sum_j w_j H_j}}{\text{Tr}(e^{-\sum_j w_j H_j})} \\
\log(\sigma) = -\sum_j w_j H_j - \log(\text{Tr}(e^{-\sum_j w_j H_j})) \\
\partial_{w_j} \log(\sigma) = -H_j + \text{Tr}(\sigma H_j) \\
\partial_{w_j} S(\rho||\sigma) = -\text{Tr}(\rho H_j) + \text{Tr}(\sigma H_j) = \langle H_j \rangle_{\text{model}} - \langle H_j \rangle_{\text{data}}
\]

The number of times the training data is accessed and Gibbs states are prepared scales as \(O\left(\frac{M^2}{\epsilon^2}\right)\).

\(M = \) number of weights.
Relative entropy training

Training objective function
Learning a model for a quantum system

Assume a quantum Hamiltonian that is a transverse Ising model:

\[ H = \sum_j \alpha_j Z^j + \sum_j \beta_j X^j + \sum_{\langle i,j \rangle} \gamma_{i,j} Z^i Z^j. \]

Find, \( \alpha, \beta \) and \( \gamma \) that best represent the Hamiltonian given training data

\[ \rho = e^{-\tilde{H}} / Z \]

Want to minimize \( \Delta H = || H - \tilde{H} || \)
Building a quantum associative memory

Quantum Training vectors: $v_j = \psi_j \otimes e_j$

Pick $H_j$ from a universal set of Hamiltonians on 6 qubits. 1000 epochs, learning rate=0.2

Relative entropy $\approx 10^{-3}$. Address qubits

Data qubits
Conclusions

Quantum computers enable new forms of machine learning.

Quantum Neural networks can learn directly from quantum data.

This opens up much richer models of data that cannot be efficiently trained using classical computers.
What is machine learning?

Machine learning uses computers to find patterns in data, classify data, or perform tasks.

Computers are not usually told explicitly how to do this, or what “features” to use.
Thank you