

Title: The quantum Boltzmann machine

Date: Aug 24, 2018 11:00 AM

URL: <http://pirsa.org/18080083>

Abstract: <p>We propose to generalise classical maximum likelihood learning to density matrices. As the objective function, we propose a quantum likelihood that is related to the cross entropy between density matrices. We apply this learning criterion to the quantum Boltzmann machine (QBM), previously proposed by Amin et al. We demonstrate for the first time learning a quantum Hamiltonian from quantum statistics using this approach. For the anti-ferromagnetic Heisenberg and XYZ model we recover the true ground state wave function and Hamiltonian. The second contribution is to apply quantum learning to learn from classical data. Quantum learning uses in addition to the classical statistics also quantum statistics for learning. These statistics may violate the Bell inequality, as in the quantum case. Maximizing the quantum likelihood yields results that are significantly more accurate than the classical maximum likelihood approach in several cases. We give an example how the QBM can learn a strongly non-linear problem such as the parity problem. The solution shows entanglement, quantified by the entanglement entropy.</p>

Learning quantum models from quantum or classical data

H.J. (Bert) Kappen
Department of Biophysics, Radboud University,
the Netherlands

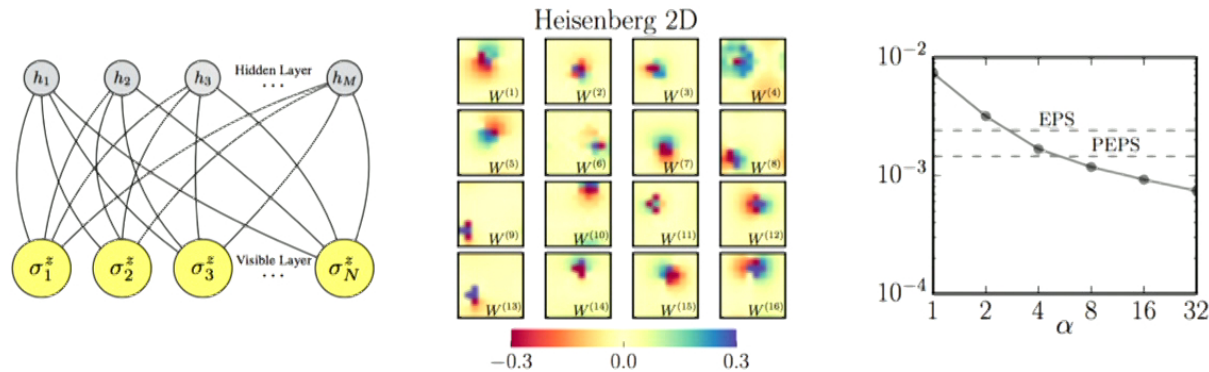
August 20, 2018



Bert Kappen

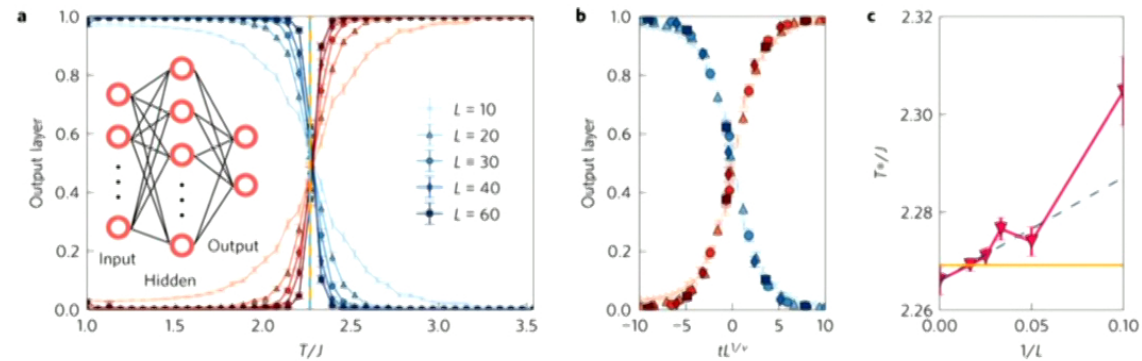
Machine learning → Quantum physics

Solving the Quantum Many-Body Problem with Artificial Neural Networks



Carleo, Troyer, Science 2017

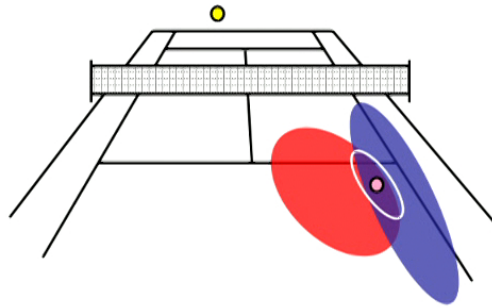
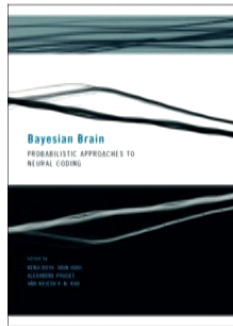
Machine learning phases of matter



Carrasquilla, Melko. Nature Physics 2017

Machine learning ← Quantum physics

The Bayesian brain



Bayesian models of inductive
learning and reasoning

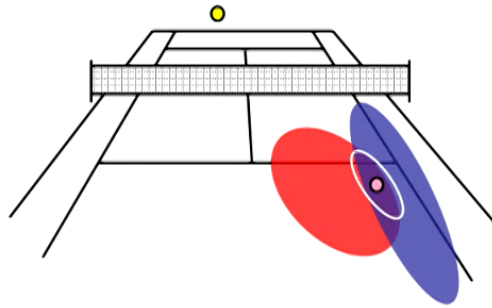
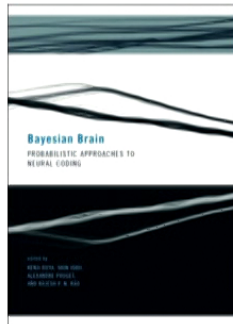
Josh Tenenbaum
MIT

Department of Brain and Cognitive Sciences
Computer Science and AI Lab (CSAIL)

Current dominant view in AI:

learn a probability distribution
MC sampling

The Bayesian brain



Bayesian models of inductive
learning and reasoning

Josh Tenenbaum
MIT

Department of Brain and Cognitive Sciences
Computer Science and AI Lab (CSAIL)

Current dominant view in AI:

learn a probability distribution
MC sampling

Aim:

learn quantum wave function
superposition through entanglement

Outline

Classical maximum likelihood framework

- learning probabilities
- Boltzmann Machine: learning a classical spin system

Quantum likelihood

- a likelihood for density matrices
- quantum spin 1/2 systems

Quantum tomography

- inferring wave function and Hamiltonian from quantum statistics

Generalize classical learning

- quantum statistics in classical data
- entanglement
- Bell inequality

Classical maximum likelihood learning

Data $s^\mu, \mu = 1, \dots, P$

Parametrized probability model $p(s|w)$

Maximize the probability that the data is generated by the model

$$e^{L_c(w)} = \prod_{\mu=1}^P p(s^\mu|w)$$

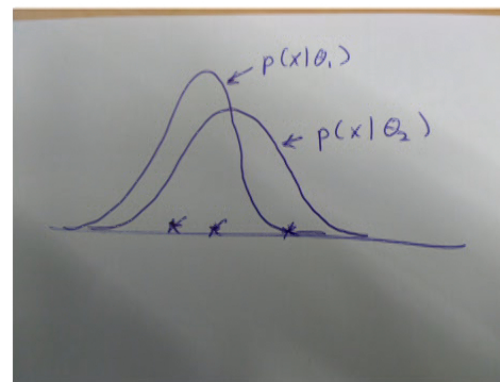
Maximizing L_c is equivalent to minimizing

$$KL(q|p) = \sum_x q(s) \log \frac{q(s)}{p(s)} \quad q(s) = \frac{1}{P} \sum_{\mu=1}^P \delta_{s, s^\mu}$$

q is the empirical (or data) distribution.

One can show:

$$KL(q|p) \geq 0 \quad KL(q|p) = 0 \quad \text{iff} \quad p = q$$



Boltzmann Machines

The state space consists of all vectors $s = (s_1, \dots, s_n)$ with $s_i = \pm 1$ binary spin variables.

$$p(s|w) = \frac{e^{H(s|w)}}{Z} \quad Z = \sum_s e^{H(s|w)} \quad H(s|w) = \sum_r w_r H_r(s)$$

The likelihood becomes

$$L_c(w) = \langle H \rangle_q - \log Z$$

Learning is gradient ascent on L_c , using $L_c(w + \Delta w) \geq L_c(w)$

$$\Delta w_r = \epsilon \frac{\partial L_c}{\partial w_r} \quad \frac{\partial L_c}{\partial w_r} = \langle H_r \rangle_q - \langle H_r \rangle_p$$

for ϵ small. Learning stops when the gradients are zero, ie. when $\langle H_r \rangle_q = \langle H_r \rangle_p$.

Boltzmann Machines

When $H(s|w) = \sum_i w_i s_i + \sum_{i>j} w_{ij} s_i s_j$, the learning rule becomes

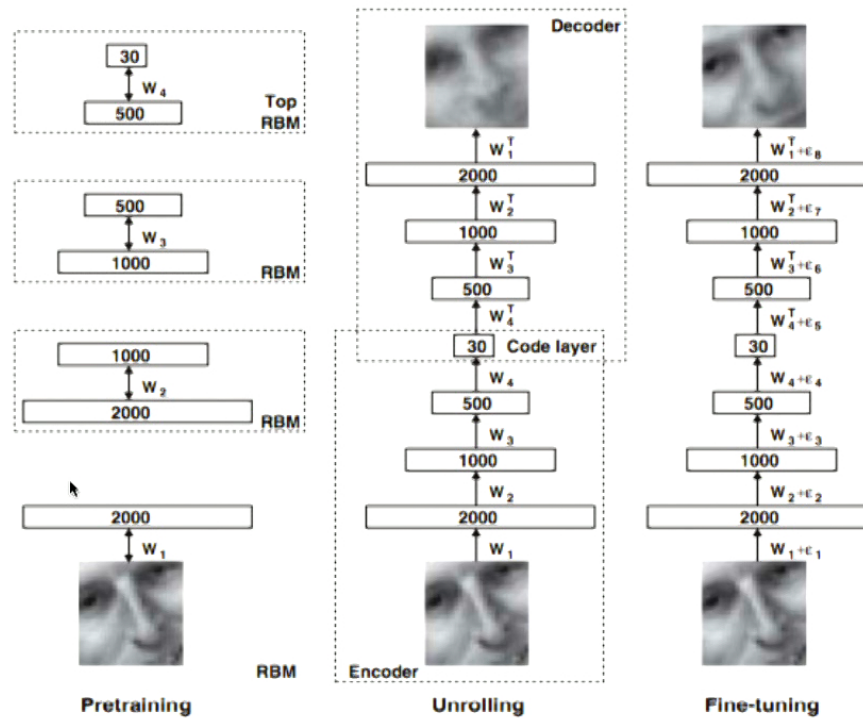
$$\Delta w_i \propto \langle s_i \rangle_q - \langle s_i \rangle_p \quad \Delta w_{ij} \propto \langle s_i s_j \rangle_q - \langle s_i s_j \rangle_p$$

This is the well-known Boltzmann Machine learning rule (Hinton 1984).

Computing $\langle \rangle_q$ is easy. Computing $\langle \rangle_p$ is hard (scales exponentially with n).

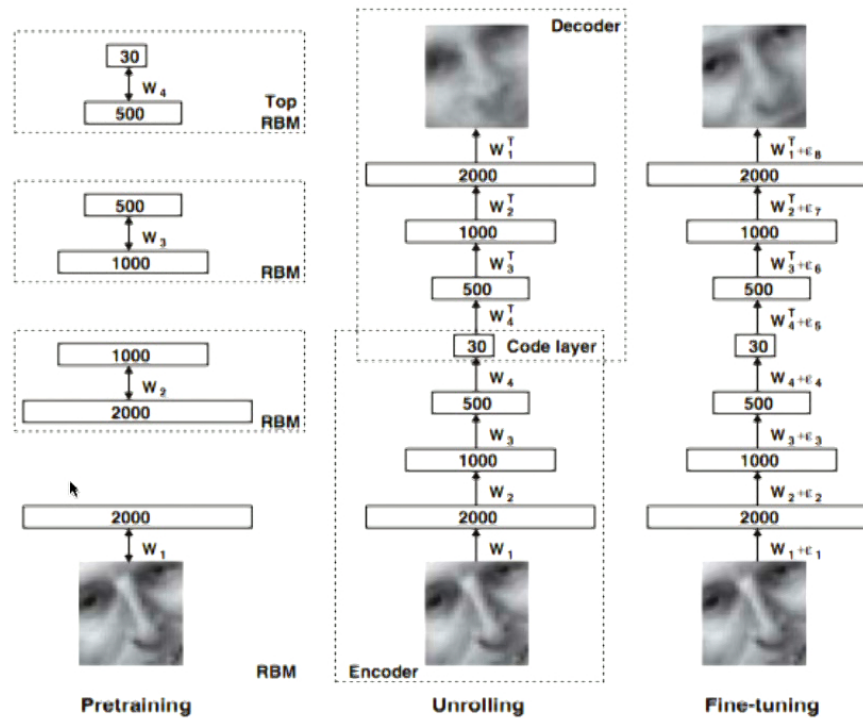


Deep learning



Hinton Salakhutdinov 2006

Deep learning



Hinton Salakhutdinov 2006

Density matrix formalism

Consider discrete state space labeled by s .

The density matrix $\rho(s, s')$ is a Hermitian positive semi-definite matrix with real eigenvalues $\lambda(s)$.

$$\lambda(s) \geq 0 \quad \text{Tr } \rho = \sum_s \lambda(s) = 1$$

When $\rho(s, s') = p(s)\delta_{s,s'}$ the density matrix reduces to a classical probability distribution.

The expectation value of a matrix A (quantum statistics) is defined as

$$\langle A \rangle_\rho = \text{Tr}(A\rho)$$

When A is diagonal,

$$\langle A \rangle_\rho = \langle A \rangle_p$$

with p the diagonal of ρ (classical statistics).

Density matrix formalism

The cross entropy between density matrices η, ρ is defined as

$$S(\eta, \rho) = \text{Tr}(\eta \log \eta) - \text{Tr}(\eta \log \rho)$$

with $\log \rho$ the matrix logarithm of ρ .

One can show (Klein's inequality):

- $S(\eta, \rho) \geq 0$
- $S(\eta, \rho) = 0$ iff $\eta = \rho$

S generalizes the concept of cross entropy between distributions to density matrices and reduces to the latter for diagonal density matrices.

Quantum learning

Denote η the data density matrix and ρ is the model density matrix.

We define quantum learning to find ρ that minimises S . This is equivalent to maximizing the quantum likelihood

$$L(\rho) = \text{Tr}(\eta \log \rho)$$

We consider density matrix models of the form (Quantum Boltzmann Machine)

$$\rho = \frac{1}{Z} e^H \quad Z = \text{Tr}(e^H) \quad H = \sum_r H_r w_r$$

H is the Hamiltonian of the quantum system

H_r are Hermitian matrices that describe the various quantum interactions.

$w = \{w_r, r = 1, \dots\}$ are real parameters that parametrize ρ .

e^H the matrix exponential of H .

Quantum learning

The quantum likelihood for the QBM is

$$L(w) = \langle H \rangle_\eta - \log Z$$

Learning is defined as gradient ascent on the quantum likelihood.

Since $H = \sum_r w_r H_r$, $\frac{\partial}{\partial w_r} \langle H \rangle_\eta = \langle H_r \rangle_\eta$.

The derivative $\frac{\partial}{\partial w_r} \log Z = \langle H_r \rangle_\rho$.¹ This gives the learning rule for the QBM:

$$\Delta w_r \propto \frac{\partial L}{\partial w_r} = \langle H_r \rangle_\eta - \langle H_r \rangle_\rho$$

¹ This is computed through the Trotter expansion $e^H = \lim_{m \rightarrow \infty} (e^{H/m})^m$. Then

$$\frac{\partial}{\partial w_r} e^H = \int_0^1 dt e^{Ht} H_r e^{H(1-t)} \quad \frac{\partial}{\partial w_r} \text{Tr}(e^H) = \text{Tr}(H_r e^H) \quad \frac{\partial}{\partial w_r} \log Z = \frac{1}{Z} \frac{\partial}{\partial w_r} \text{Tr}(e^H) = \langle H_r \rangle_\rho$$

Quantum Boltzmann Machine

For n quantum spin variables σ_i^k we consider the Hamiltonian

$$H = \sum_{i=1}^n \sum_{k=x,y,z} w_i^k \sigma_i^k + \sum_{i=1}^n \sum_{j>i} w_{ij}^k \sigma_i^k \sigma_j^k$$

$\sigma_i^{x,y,z}$ are Pauli spin 1/2 operators. For this Hamiltonian, the learning rule becomes

$$\Delta w_i^k = \langle \sigma_i^k \rangle_{\eta} - \langle \sigma_i^k \rangle_{\rho} \quad \Delta w_{ij}^k = \langle \sigma_i^k \sigma_j^k \rangle_{\eta} - \langle \sigma_i^k \sigma_j^k \rangle_{\rho} \quad k = x, y, z$$

The QBM reduces to classical BM when $w_i^{x,y} = w_{ij}^{x,y} = 0$.

Hamiltonian matrix representation

The state of an n spins system is the tensor product of the states $|s_i = \pm 1\rangle$ for each spin:

$$|s\rangle = |s_1, \dots, s_n\rangle$$

Thus a total of 2^n states.

On this basis, H has matrix elements ²

$$\begin{aligned} \langle s'|H|s\rangle &= \sum_{i=1}^n (w_i^x + i w_i^y s_i) \delta_{s', F_i s} + \sum_{i < j}^n (w_{ij}^x - w_{ij}^y s_i^k s_j^k) \delta_{s', F_i F_j s} \\ &\quad + \left(\sum_{i=1}^n w_i^z s_i + \sum_{i < j}^n w_{ij}^z s_i s_j \right) \delta_{s', s} \end{aligned}$$

²When $s = (s_1, \dots, s_i, \dots, s_n)$, $F_i s = (s_1, \dots, -s_i, \dots, s_n)$



Hamiltonian matrix representation

The state of an n spins system is the tensor product of the states $|s_i = \pm 1\rangle$ for each spin:

$$|s\rangle = |s_1, \dots, s_n\rangle$$

Thus a total of 2^n states.

On this basis, H has matrix elements ²

$$\begin{aligned} \langle s'|H|s\rangle &= \sum_{i=1}^n (w_i^x + i w_i^y s_i) \delta_{s', s_i} + \sum_{i < j}^n (w_{ij}^x - w_{ij}^y s_i^x s_j^x) \delta_{s', s_i} \delta_{s_j} \\ &\quad + \left(\sum_{i=1}^n w_i^z s_i + \sum_{i < j}^n w_{ij}^z s_i s_j \right) \delta_{s', s} \end{aligned}$$

²When $s = (s_1, \dots, s_n)$, $s' = (s'_1, \dots, s'_n)$

Bert Kappen

© 2010

Examples of quantum tomography

Given quantum statistics of an unknown quantum system

$$\langle s_i^k \rangle_\eta, \langle s_i^k s_j^k \rangle_\eta$$

Estimate the quantum Hamiltonian that is the source of these quantum statistics.

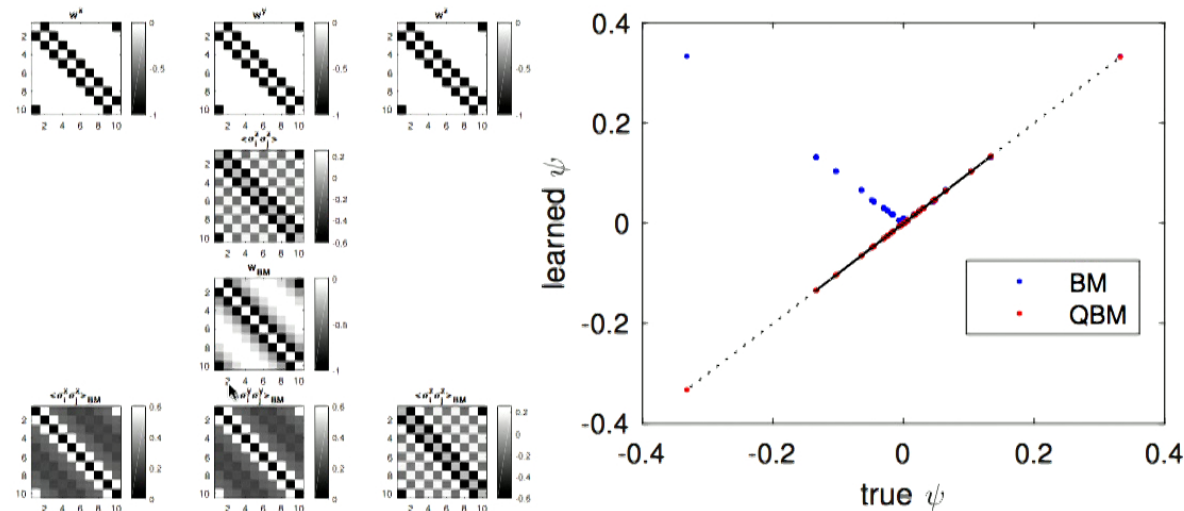
Numerical examples:

- Given H_{true} , compute the ground state wave function ψ_{true} .
- From ψ_{true} , compute the quantum statistics $\langle \sigma_i^k \rangle_\eta, \langle \sigma_i^k \sigma_j^k \rangle_\eta$.
- Train QBM and compare learned H, ψ with $H_{\text{true}}, \psi_{\text{true}}$



Anti-ferromagnetic Heisenberg model

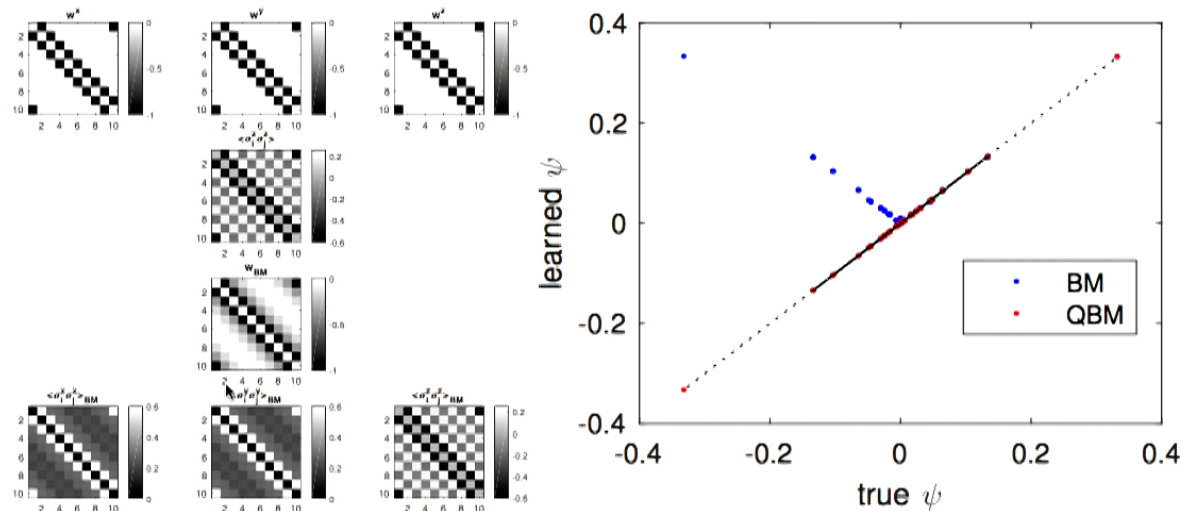
The Hamiltonian has couplings $w_{ij}^{x,y,z} = -1$ for nearest neighbours and $w_{ij}^{x,y,z} = 0$ otherwise.



The classical BM does not yield a good solution: wrong couplings and wrong sign of $\sqrt{p_{bm}}$. The BM correctly captures the classical statistics but not the quantum statistics.

Anti-ferromagnetic Heisenberg model

The Hamiltonian has couplings $w_{ij}^{x,y,z} = -1$ for nearest neighbours and $w_{ij}^{x,y,z} = 0$ otherwise.



The classical BM does not yield a good solution: wrong couplings and wrong sign of $\sqrt{p_{bm}}$. The BM correctly captures the classical statistics but not the quantum statistics.

Quantum statistics in classical data

We can also apply the QBM to learn classical data. Given data $s^\mu, \mu = 1, \dots, P$:

$$q(s) = \frac{1}{P} \sum_{\mu=1}^P \delta_{s, s^\mu}$$

We construct a rank one density matrix as follows

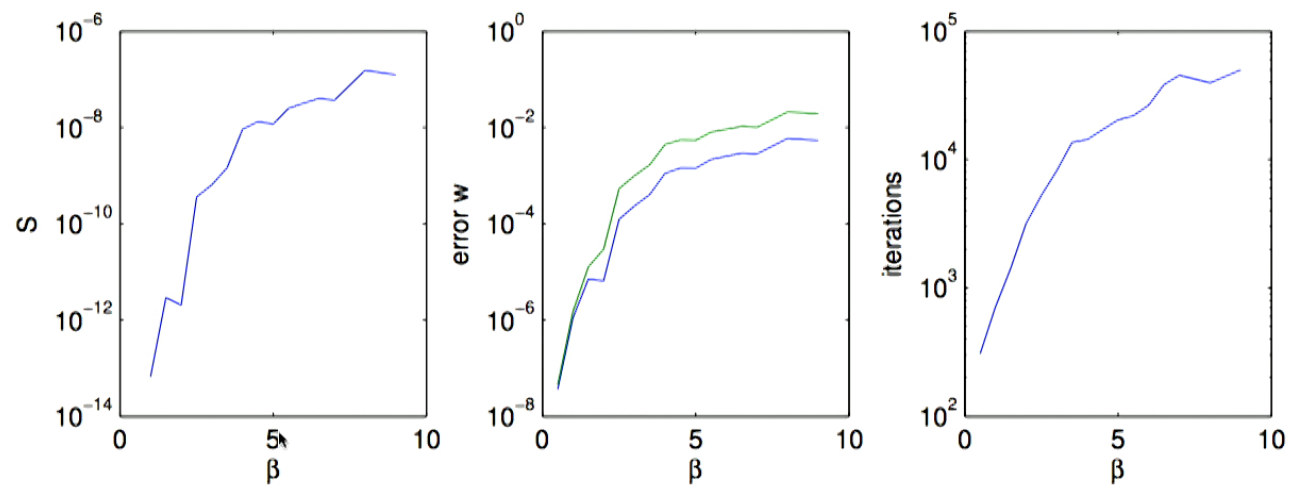
$$\eta = \psi\psi' \quad \psi(s) = \sqrt{q(s)}$$

The quantum expectations

$$\langle A \rangle_\eta = \sum_{s, s'} A(s, s') \sqrt{q(s)q(s')}$$

generalize the classical expectations and become equal for A diagonal.

Classical statistics are linear in q . The quantum statistics are quadratic in $\sqrt{q}$.



Quantum statistics in classical data

We can also apply the QBM to learn classical data. Given data $s^\mu, \mu = 1, \dots, P$:

$$q(s) = \frac{1}{P} \sum_{\mu=1}^P \delta_{s, s^\mu}$$

We construct a rank one density matrix as follows

$$\eta = \psi\psi' \quad \psi(s) = \sqrt{q(s)}$$

The quantum expectations

$$\langle A \rangle_\eta = \sum_{s, s'} A(s, s') \sqrt{q(s)q(s')}$$

generalize the classical expectations and become equal for A diagonal.

Classical statistics are linear in q . The quantum statistics are quadratic in $\sqrt{q}$.

Quantum statistics in classical data

For the QBM:

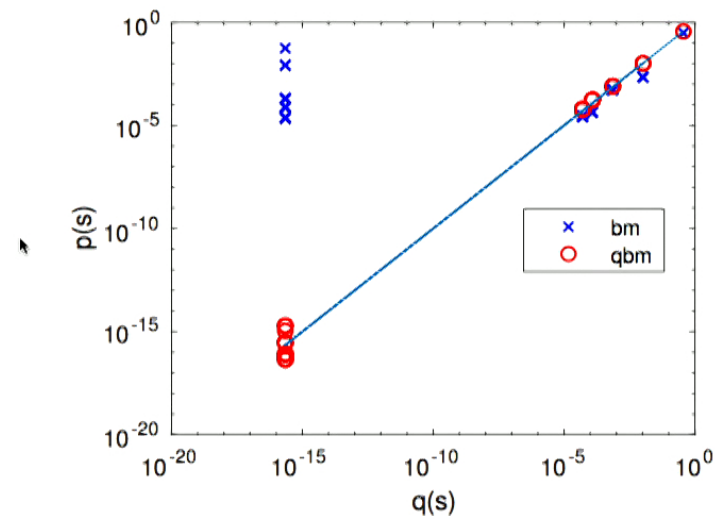
$$\begin{aligned}\langle \sigma_i^x \rangle_\eta &= \sum_s \sqrt{q(F_i s) q(s)} & \langle \sigma_i^x \sigma_j^x \rangle_\eta &= \sum_s \sqrt{q(F_i F_j s) q(s)} \\ \langle \sigma_i^y \rangle_\eta &= 0 & \langle \sigma_i^y \sigma_j^y \rangle_\eta &= - \sum_s s_i s_j \sqrt{q(F_i F_j s) q(s)} \\ \langle \sigma_i^z \rangle_\eta &= \langle s_i \rangle_q & \langle \sigma_i^z \sigma_j^z \rangle_\eta &= \langle s_i s_j \rangle_q\end{aligned}$$

The quantum statistics measure similarity between patterns in the data.

Parity problem

Data are generated as $q(s_{1:n}) = q(s_{1:n-1})q(s_n|s_{1:n-1})$ with

$$q(s_{1:n-1}) = \frac{1}{Z} \exp\left(\frac{1}{n-1} \sum_{i < j} s_i s_j\right) \quad q(s_n|s_{1:n-1}) = \delta_{s_n, \prod_{i=1}^{n-1} s_i}$$

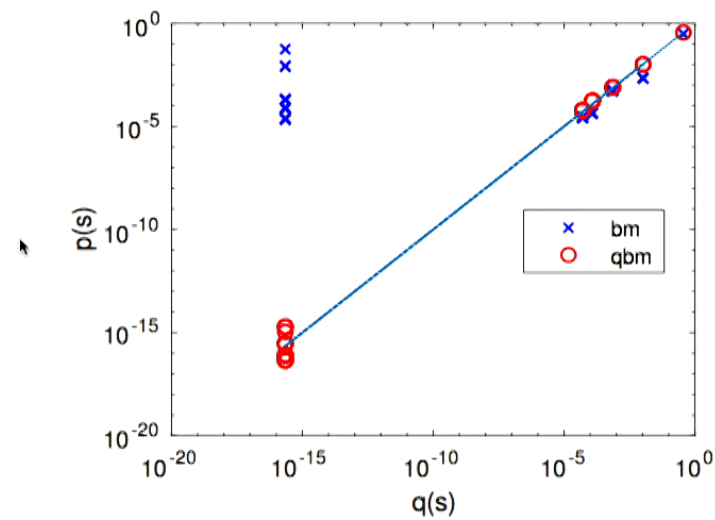


BM and QBM solutions. $S(\eta, \rho_{\text{bm}}) = 2.735$ and $S(\eta, \rho_{\text{qbm}}) = 0.404$.

Parity problem

Data are generated as $q(s_{1:n}) = q(s_{1:n-1})q(s_n|s_{1:n-1})$ with

$$q(s_{1:n-1}) = \frac{1}{Z} \exp\left(\frac{1}{n-1} \sum_{i < j} s_i s_j\right) \quad q(s_n|s_{1:n-1}) = \delta_{s_n, \prod_{i=1}^{n-1} s_i}$$



BM and QBM solutions. $S(\eta, \rho_{\text{bm}}) = 2.735$ and $S(\eta, \rho_{\text{qbm}}) = 0.404$.

Entanglement

The entanglement between variables in subset A and its complement B is defined as the quantum mutual information

$$I_{AB} = \mathcal{H}(\rho_A) + \mathcal{H}(\rho_B) - \mathcal{H}(\rho) \quad \mathcal{H}(\rho) = -\text{Tr}(\rho \log \rho)$$

with $\mathcal{H}(\rho_A), \mathcal{H}(\rho_B)$ the entropies of the marginal density matrices

$$\rho_A = \text{Tr}_B(\rho) \quad \rho_B = \text{Tr}_A(\rho)$$

Entanglement

The entanglement between variables in subset A and its complement B is defined as the quantum mutual information

$$I_{AB} = \mathcal{H}(\rho_A) + \mathcal{H}(\rho_B) - \mathcal{H}(\rho) \quad \mathcal{H}(\rho) = -\text{Tr}(\rho \log \rho)$$

with $\mathcal{H}(\rho_A), \mathcal{H}(\rho_B)$ the entropies of the marginal density matrices

$$\rho_A = \text{Tr}_B(\rho) \quad \rho_B = \text{Tr}_A(\rho)$$

The entropies satisfy the Akari-Lieb inequality

$$\mathcal{H}(\rho) \geq |\mathcal{H}(\rho_A) - \mathcal{H}(\rho_B)|$$

For a pure state $\rho = \psi\psi'$

$$\mathcal{H}(\rho) = 0 \quad \mathcal{H}(\rho_A) = \mathcal{H}(\rho_B) > 0$$

This would be classically forbidden. The entanglement is $I_{AB} = 2\mathcal{H}(\rho_A)$.

Bell inequality

Consider three stochastic binary variables a, b, c with values ± 1 that satisfy a joint probability distribution $p(a, b, c)$. Then,

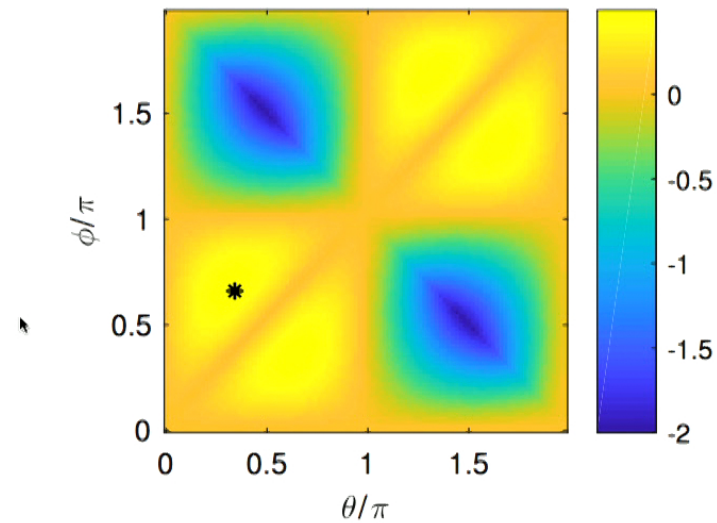
$$|\langle ab \rangle - \langle ac \rangle| + \langle bc \rangle \leq 1 \quad (\text{Bell inequality})$$

Violation of this inequality implies that a, b, c cannot be described simultaneously by a classical distribution.

Bell inequality

The Bell inequality is

$$|\cos \theta - \cos \phi| - 1 + \cos(\theta - \phi) \leq 0$$



The conclusion is that the pair-wise (quantum) correlations of a, b, c cannot be reproduced by any classical probability distribution.

Intractability

Forward problem: $H \rightarrow \langle \dots \rangle_\rho$ is intractable.

Inverse problem (learning): $\langle \dots \rangle_\eta \rightarrow H$ is solved as a sequence of forward problems:

$$w \rightarrow H(w) \rightarrow \langle \dots \rangle_\rho \quad \Delta w \propto \langle \dots \rangle_\eta - \langle \dots \rangle_\rho$$

Low rank approximation $\rho \approx \sum_{i=1}^L \lambda_i v_i v_i'$ reduces $O(2^{2n})$ to $O(L2^n)$.

Intractability

Forward problem: $H \rightarrow \langle \dots \rangle_\rho$ is intractable.

Inverse problem (learning): $\langle \dots \rangle_\eta \rightarrow H$ is solved as a sequence of forward problems:

$$w \rightarrow H(w) \rightarrow \langle \dots \rangle_\rho \quad \Delta w \propto \langle \dots \rangle_\eta - \langle \dots \rangle_\rho$$

Low rank approximation $\rho \approx \sum_{i=1}^L \lambda_i v_i v_i'$ reduces $O(2^{2n})$ to $O(L2^n)$.

Possible accelerations in software

- Finite temperature: diffusion Monte Carlo, mean field methods, quantum belief propagation
- Zero temperature: variational approximation for the ground state wave function (Carleo Troyer)

Intractability

Forward problem: $H \rightarrow \langle \dots \rangle_\rho$ is intractable.

Inverse problem (learning): $\langle \dots \rangle_\eta \rightarrow H$ is solved as a sequence of forward problems:

Forward problem: $H \rightarrow \langle \dots \rangle_\rho$ is intractable.

Low rank approximation $\rho = \sum_{i=1}^L \lambda_i v_i v_i'$ reduces $O(2^{2n})$ to $O(L2^n)$.

Inverse problem (learning): $\langle \dots \rangle_\eta \rightarrow H$ is solved as a sequence of forward problems:

Possible accelerations in software

• Finite temperature: diffusion Monte Carlo, mean field methods, quantum belief propagation

• Zero temperature: variational approximation for the ground state wave function (Carleo Troyer)

$$w \rightarrow H(w) \rightarrow \langle \dots \rangle_\rho \quad \Delta w \propto \langle \dots \rangle_\eta - \langle \dots \rangle_\rho$$

Quantum hardware

• use dedicated hardware for the forward problem (Korenkevych 2016 for BMs)

Low rank approximation $\rho \approx \sum_{i=1}^L \lambda_i v_i v_i'$ reduces $O(2^{2n})$ to $O(L2^n)$.

Possible accelerations in software

- Finite temperature: diffusion Monte Carlo, mean field methods, quantum belief propagation
- Zero temperature: variational approximation for the ground state wave function (Carleo Troyer)

Quantum hardware

- use dedicated hardware for the forward problem (Korenkevych 2016 for BMs)



Bert Kappen

QBM 2018 43

Discussion

Quantum likelihood: a maximum likelihood framework for density matrices.

Discussion

Quantum likelihood: a maximum likelihood framework for density matrices.

- recovers Hamiltonian (ie. everything!) from statistics
- elegant but expensive
- extends classical statistical description of data
 - quantum statistics
 - entanglement
- Opportunities for quantum hardware

Kappen, Hilbert J. "Learning quantum models from quantum or classical data."
arXiv preprint arXiv:1803.11278 (2018).