Quantum-Assisted Machine Learning in Near-Term Quantum Devices

Alejandro Perdomo-Ortiz

Senior Quantum Scientist, Zapata Computing, Canada

Honorary Senior Research Associate, Computer Science Dept., UCL, UK

@aperdomoortiz, @ZapataComputing

ZAPATA

Funding:

NASA USRA UCL ZAPATA

Quantum Languages Designs and Implementation, Sept 28, 2019

Capabilities of Near-Term Quantum Devices

0. Simulation of quantum systems

1. As a discrete optimization solver:

Given $\{h_j, J_{ij}\}$, find $\{s_k = \pm 1\}$ that minimizes

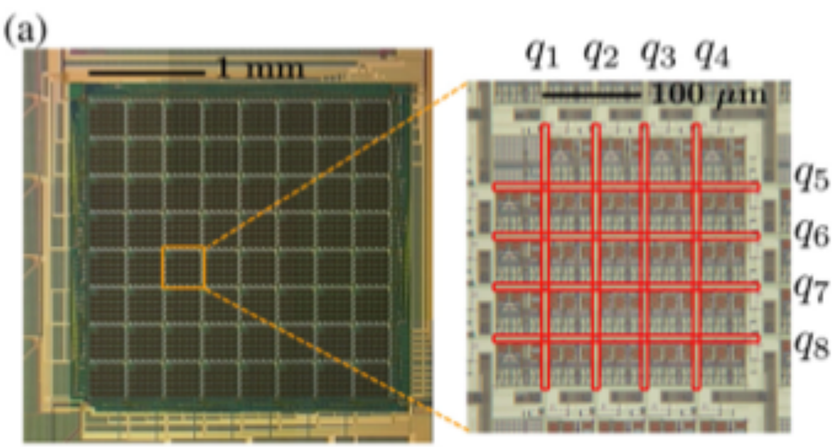
NP-hard problem

$$\xi(s_1, \ldots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E}^{N} J_{ij} s_i s_j$$

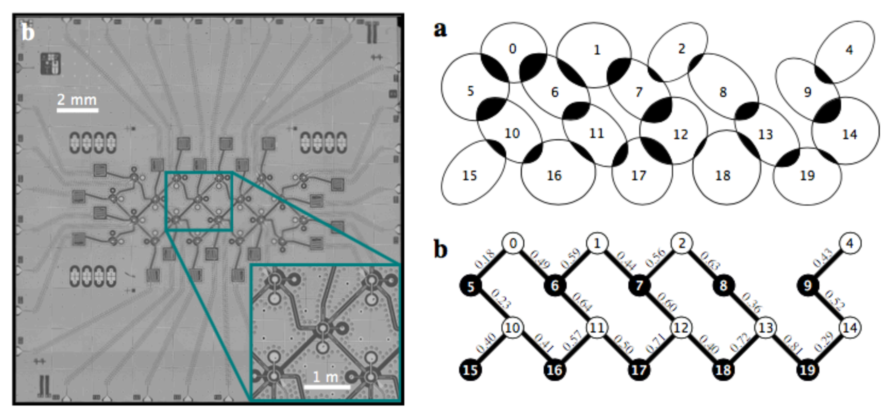
Potential applications:

- *planning*
- *scheduling*
- *fault diagnosis*
- *graph analysis*
- *communication networks, etc.*

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).



Quantum annealing (D-wave)



Example QAOA (Rigetti)

Otterbach et al. **arXiv:1712.05771**.

Capabilities of Near-Term Quantum Devices

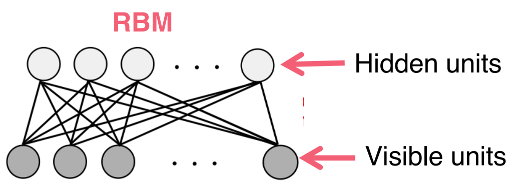
2a. As a physical device to sample from Boltzmann-like distributions:

$$P_{Boltzman} \propto exp[-\xi(s_1, ..., s_N)/T_{eff}] \longrightarrow \langle v_i h_j \rangle_{p(\boldsymbol{h},\boldsymbol{v})}$$

Computationally bottleneck

Example: Quantum annealing

Widely used in **generative unsupervised** learning



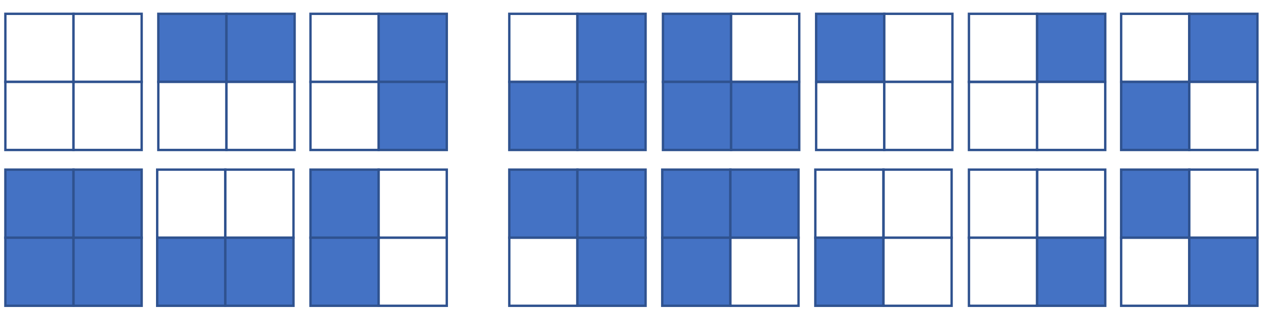
Potential applications:

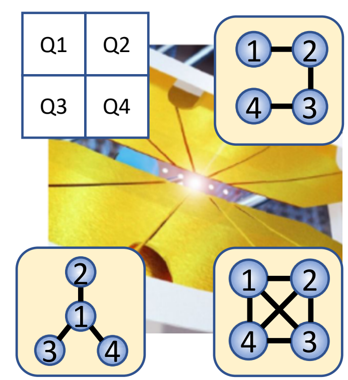
- *machine leaning (e.g., training of deep-learning networks)*

Examples for training probabilistic graphical models:
Benedetti, et al. **Phys. Rev. A 94, 022308** (2016); Benedetti, et al. **Phys. Rev. X 7, 041052 (2017).**

2b. Sampling from more broader quantum distributions

BAS patterns **Non-BAS pattern**





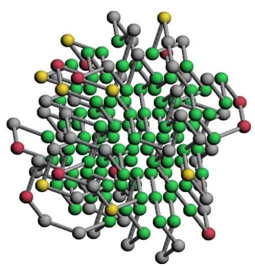
Benedetti et al. A generative modeling approach for benchmarking and training shallow quantum circuits. **npj QI, 5, 45 (2019).**

Motivation

- What are data sets and (non-obvious) real-world applications in need of quantum resources from NISQ devices?

    - **Combinatorial optimization?**

Protein Folding



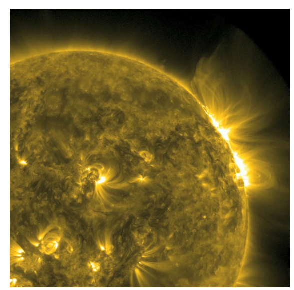
Lattice protein folding

- Perdomo-Ortiz et al. *Phys. Rev. A. 78(1):012320 (2008).*

- Perdomo-Ortiz et al. *Sci. Rep., 2, 571, (2012).*

- Kassal, et al. *Ann. Rev.Phys. Chem. 62, 185-207 (2011).*

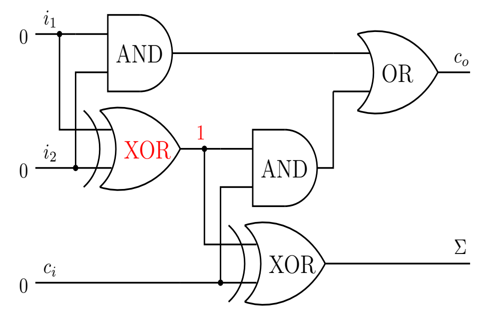
Bayesian networks



Solar Flare prediction

- O'Gorman, et al. *Eur. Phys. J. Spec. Topics. 224, 163- 188 (2015)*

Fault diagnosis Applications



- Perdomo-Ortiz et al. *arXiv:1503.01083 (2015)*

- Perdomo-Ortiz et al. *Eur. Phys. J. Spec. Topics. 224, 131-148 (2015).*

- Perdomo-Ortiz et al. *arXiv:1708.09780 (2017). Accepted in Phys. Rev. Applied.*

Motivation/Outline

- Are there data sets and (non-obvious) real-world applications in need of quantum resources from NISQ devices?
  - Combinatorial optimization? **Machine learning?**

Perspective: Perdomo-Ortiz, et al. Opportunities and Challenges in Quantum-Assisted Machine Learning in Near-term Quantum Computers. **Quantum Sci. Technol. 3, 030502 (2018).** Invited special issue on “What would you do with a 1000 qubit?”

Motivation/Outline

- Are there data sets and (non-obvious) real-world applications in need of quantum resources from NISQ devices?

    - Combinatorial optimization? **Machine learning?**

    Perspective: Perdomo-Ortiz, et al. Opportunities and Challenges in Quantum-Assisted Machine Learning in Near-term Quantum Computers. **Quantum Sci. Technol. 3, 030502 (2018).** Invited special issue on “What would you do with a 1000 qubit?”

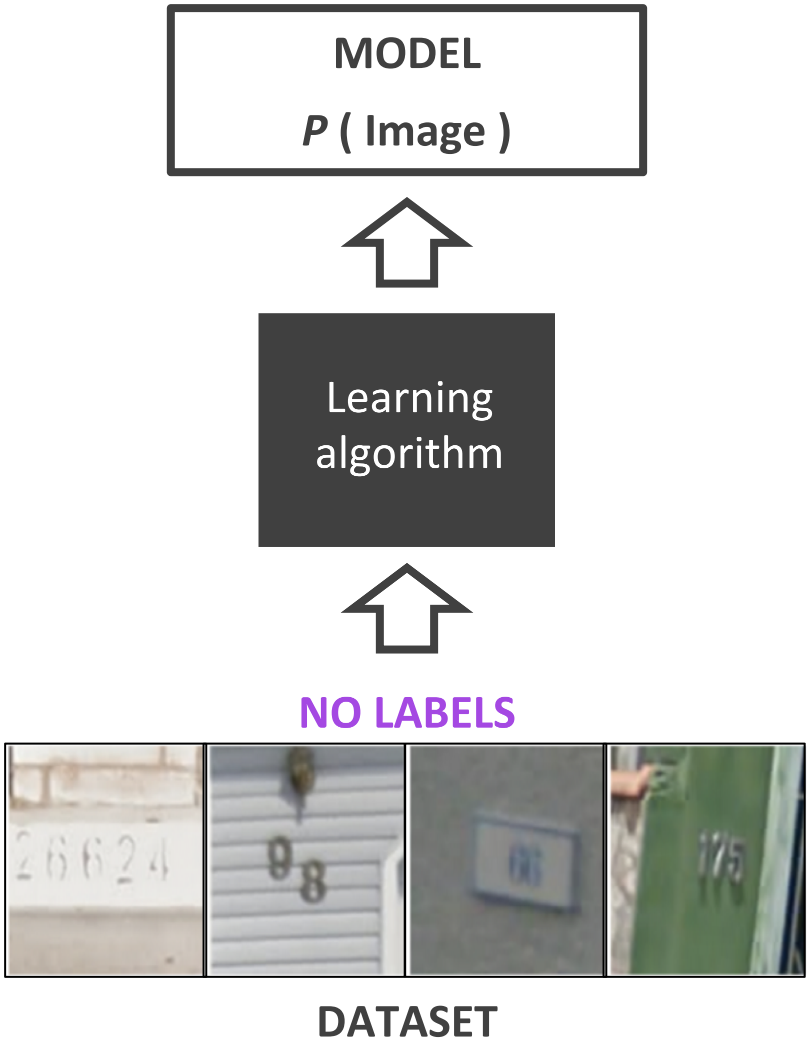
- Why and where to look for quantum advantage in quantum-assisted ML, with NISQ devices?

- Know your hybrid quantum-classical pipeline: classical optimizers, circuit ansatz, etc.

- NISQ quantum models in a real-world setting: an example from a financial application.

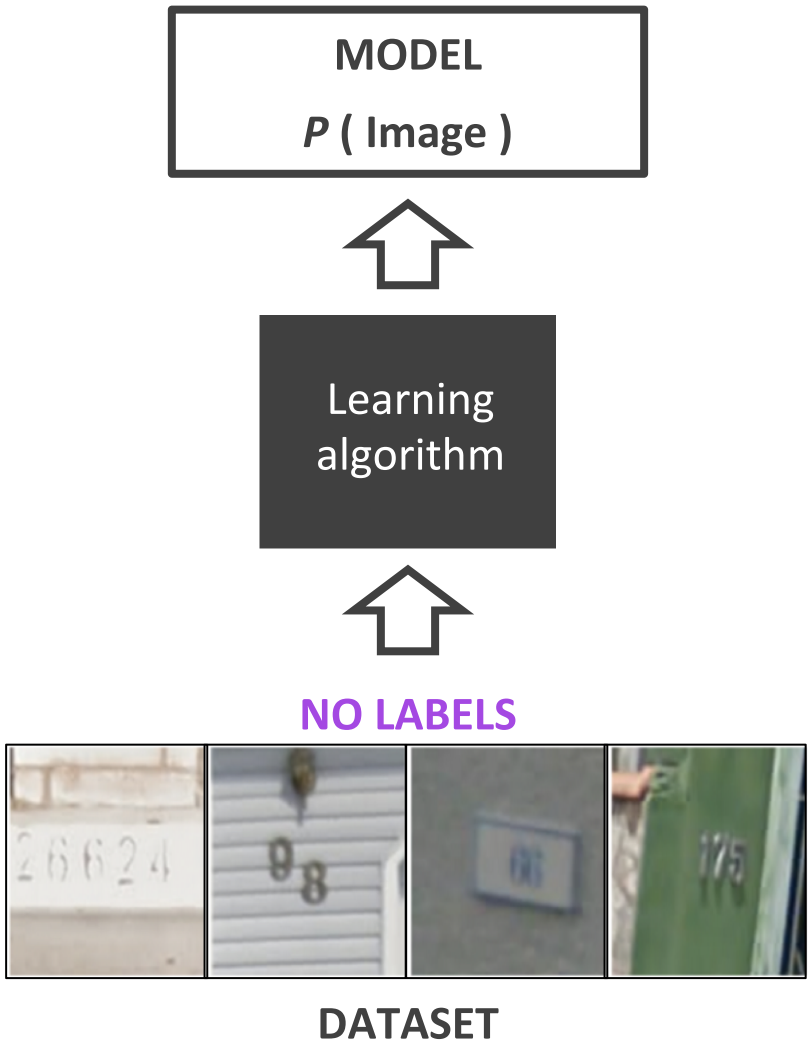
Unsupervised learning (generative models)

Learn the "best" model distribution that can generate the same kind of data

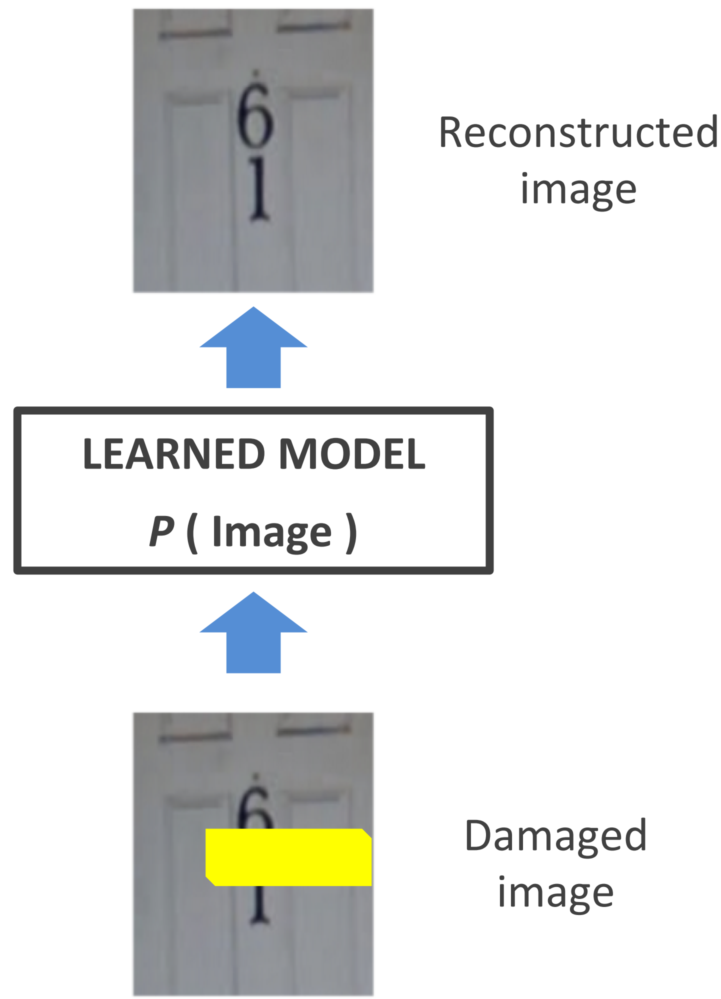


Unsupervised learning (generative models)

Learn the "best" model distribution that can generate the same kind of data

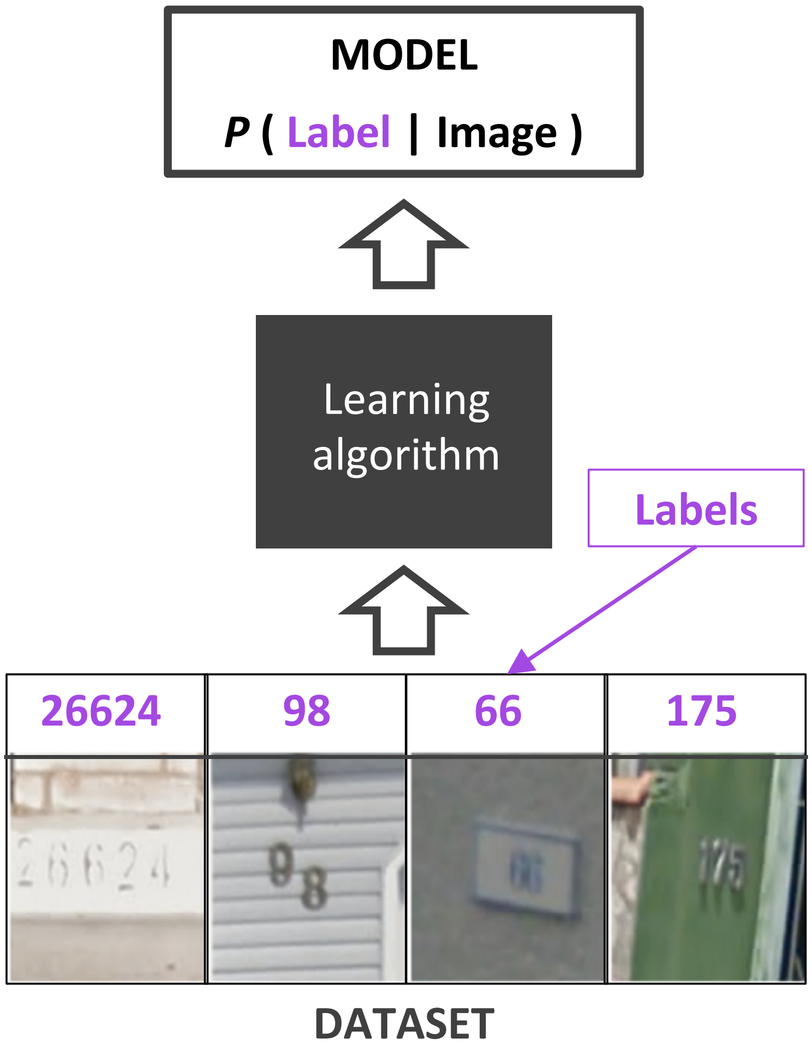


Example application:
Image reconstruction

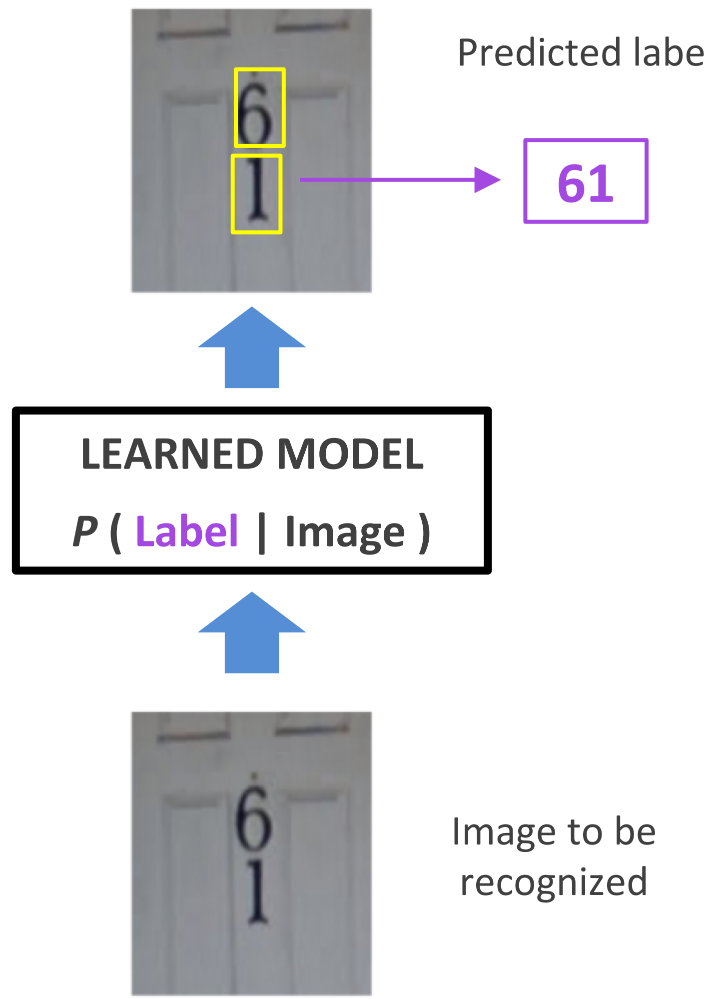


Supervised learning (discriminative models)

Learn the “best” model that can perform a specific task



Example application:
Image recognition



A near-term approach for quantum-enhanced machine learning

Insight 1: Work on intractable problems of interest to ML experts (e.g., **generative models in unsupervised learning**). **Quantum advantage in near term.**

*"Unsupervised learning [... has] been overshadowed by the successes of purely supervised learning. [... We] expect **unsupervised learning to become far more important in the longer term**. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object."*

LeCun, Bengio, Hinton, *Deep Learning*, **Nature 2015**

Perdomo-Ortiz, Benedetti, Realpe-Gomez, Biswas. **Quantum Sci. Technol. 3, 030502 (2018).**

A near-term approach for quantum-enhanced machine learning

Insight 1: Work on intractable problems of interest to ML experts (e.g., **generative models in unsupervised learning**). **Quantum advantage in near term.**

"In the context of the deep learning approach to undirected modeling, it is rare to use any approach other than Gibbs sampling. ***Improved sampling techniques are one possible research frontier.****"*

Goodfellow, Bengio, Courville, *Deep Learning*, book in preparation for MIT Press, **2016**

A near-term approach for quantum-enhanced machine learning

Insight 1: Work on intractable problems of interest to ML experts (e.g., **generative models in unsupervised learning**). **Quantum advantage in near term.**

*"Most of the previous work in **generative models** has focused on variants of **Boltzmann Machines** [...] While these models **are very powerful**, each iteration of **training requires a computationally costly step of MCMC** to approximate derivatives of an intractable partition function (normalization constant), making it **difficult to scale** them **to large datasets**."*

Mansimov, Parisotto, Ba, **Salakhutdinov**, ICLR **2016**

Perdomo-Ortiz, Benedetti, Realpe-Gomez, Biswas. **Quantum Sci. Technol. 3, 030502 (2018).**

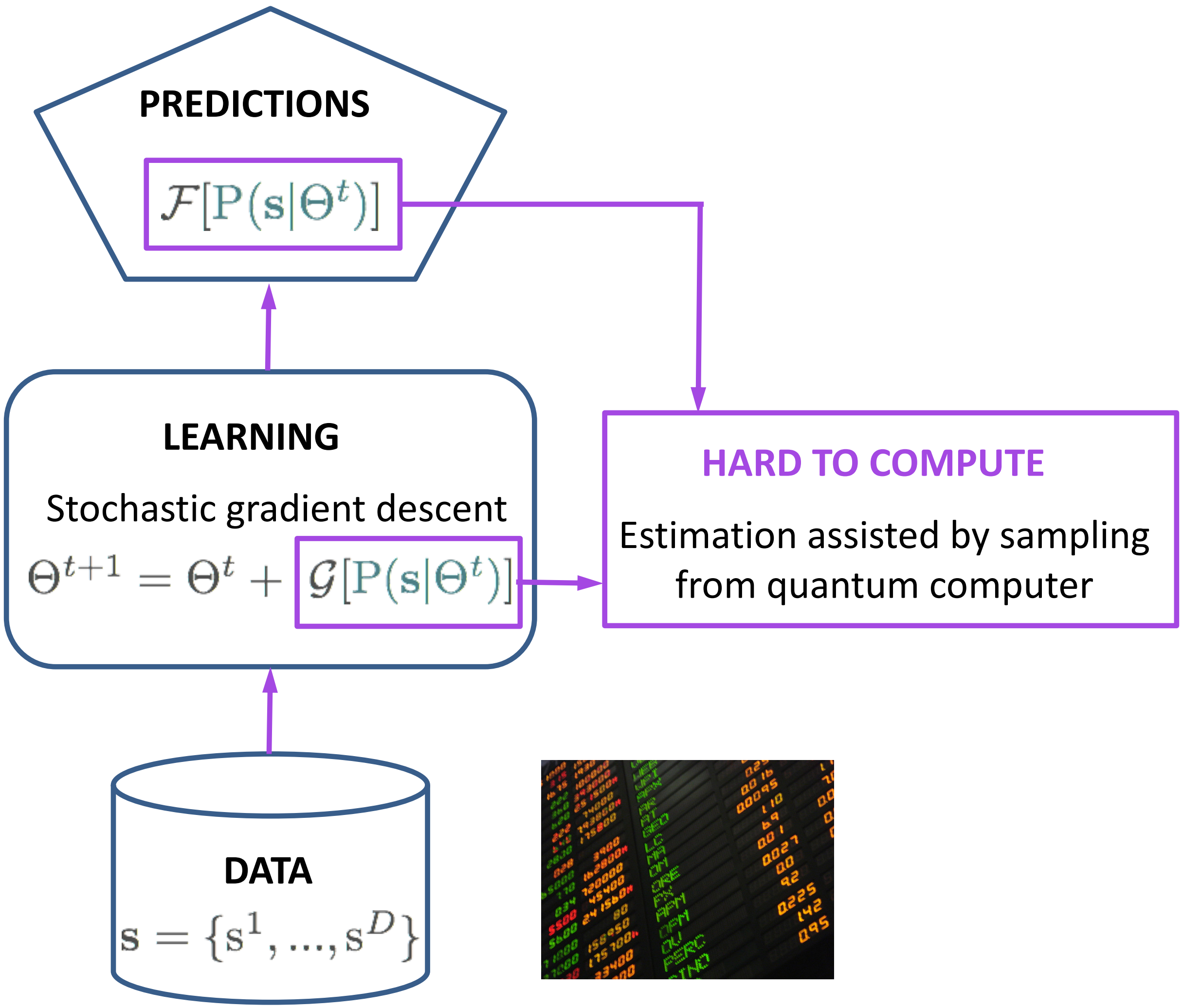
A near-term approach for quantum-enhanced machine learning

Insight 2: Focus on hybrid quantum-classical approaches.
Cope with hardware constrains and available quantum resources

Perdomo-Ortiz, Benedetti, Realpe-Gomez, Biswas. **Quantum Sci. Technol. 3, 030502 (2018).**

A near-term approach for quantum-enhanced machine learning

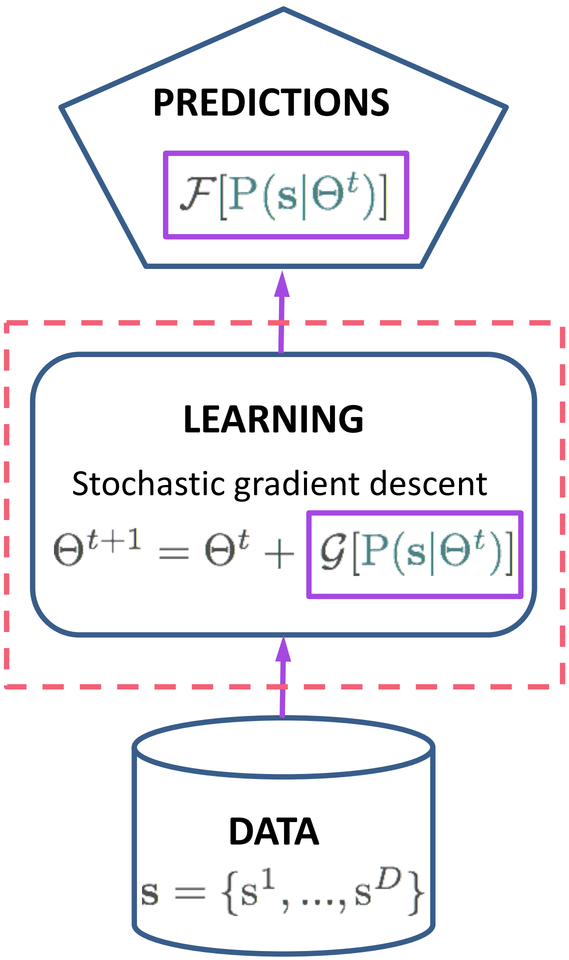
Insight 2: Focus on <u>hybrid</u> quantum-classical approaches.
Cope with hardware constrains and available quantum resources



Perdomo-Ortiz, Benedetti, Realpe-Gomez, Biswas. **Quantum Sci. Technol. 3, 030502 (2018).**

A near-term approach for quantum-enhanced machine learning

Insight 2: Focus on hybrid quantum-classical approaches.
Cope with hardware constrains and available quantum resources



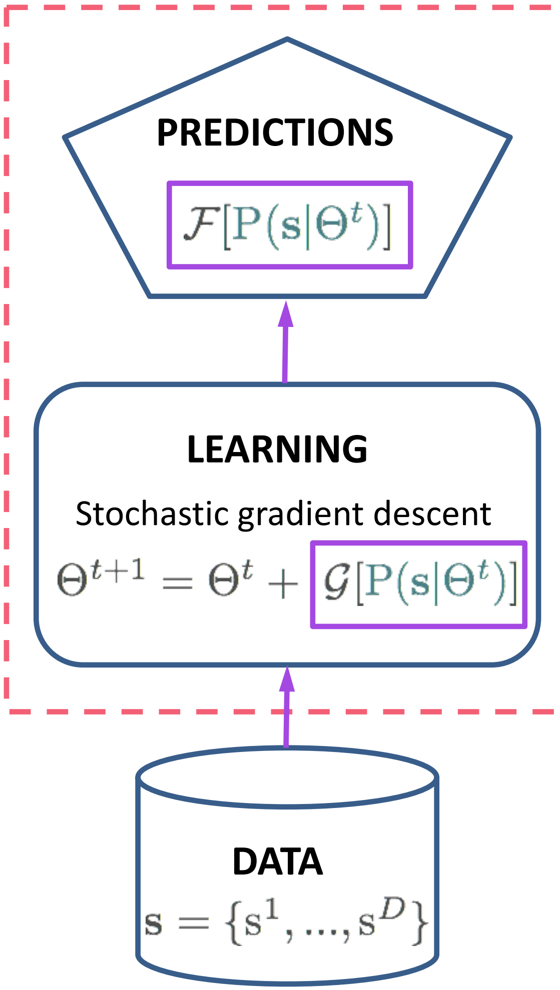
Challenges solved:

1. Perdomo-Ortiz, et al. **Opportunities and Challenges** in Quantum-Assisted Machine Learning in Near-term Quantum Computers. *Quantum Sci. Tecnol. 3, 030502 (2018)*.

2. Benedetti, et al. **Estimation of effective temperatures** in quantum annealers for sampling applications: A case study with possible applications in deep learning. *Phys. Rev. A 94, 022308 (2016).*

A near-term approach for quantum-enhanced machine learning

Insight 2: Focus on hybrid quantum-classical approaches.
Cope with hardware constrains and available quantum resources



Challenges solved:

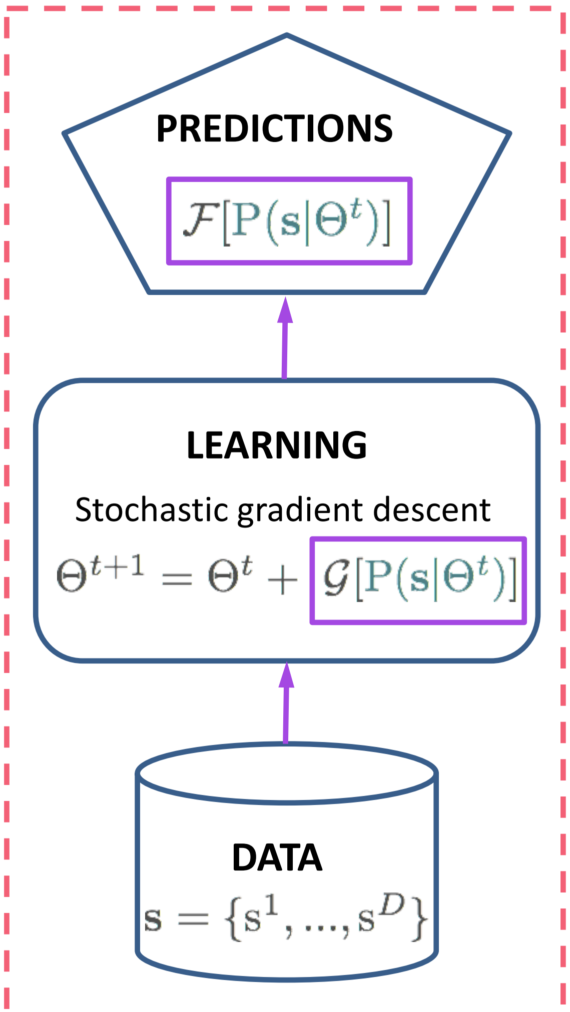
1. Perdomo-Ortiz, et al. **Opportunities and Challenges** in Quantum-Assisted Machine Learning in Near-term Quantum Computers. *Quantum Sci. Tecnol. 3, 030502 (2018)*.

2. Benedetti, et al. **Estimation of effective temperatures** in quantum annealers for sampling applications: A case study with possible applications in deep learning. *Phys. Rev. A 94, 022308 (2016).*

3. Benedetti, et al. Quantum-assisted learning of **hardware-embedded** probabilistic graphical models. *Phys. Rev. X 7, 041052 (2017).*

A near-term approach for quantum-enhanced machine learning

Insight 2: Focus on <u>hybrid</u> quantum-classical approaches.
Cope with hardware constrains and available quantum resources



Challenges solved:

1. Perdomo-Ortiz, et al. **Opportunities and Challenges** in Quantum-Assisted Machine Learning in Near-term Quantum Computers. *<u>Quantum Sci. Tecnol</u>. 3, 030502 (2018)*.

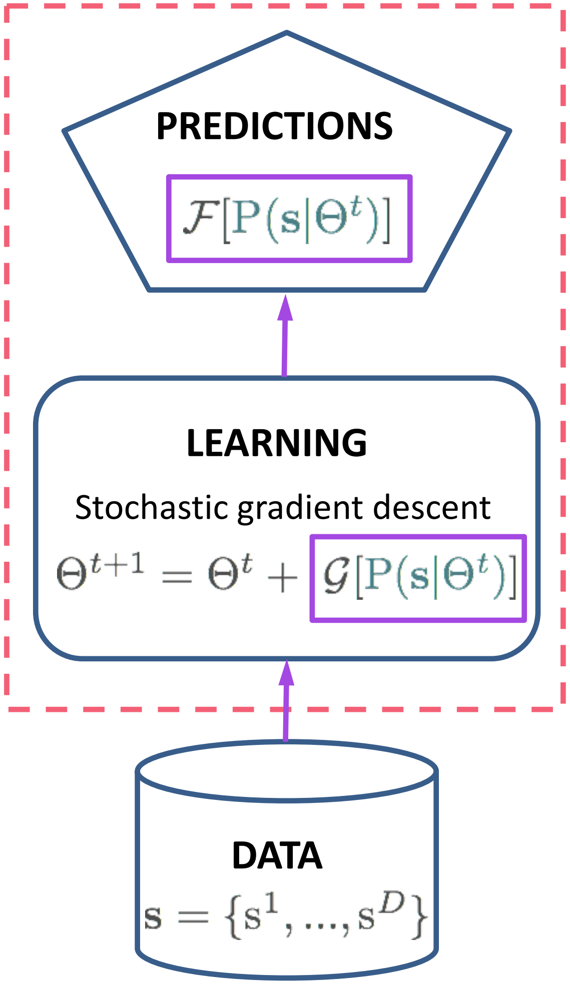
2. Benedetti, et al. **Estimation of effective temperatures** in quantum annealers for sampling applications: A case study with possible applications in deep learning. *<u>Phys. Rev. A</u> 94, 022308 (2016).*

3. Benedetti, et al. Quantum-assisted learning of **hardware-embedded** probabilistic graphical models. *<u>Phys. Rev. X</u> 7, 041052 (2017).*

4. Benedetti, et al. Quantum-assisted Helmholtz machines: A quantum-classical deep learning framework for **industrial datasets in near-term devices**. *<u>Quantum Sci. Technol</u>. 3, 034007 (2018).*

A near-term approach for quantum-enhanced machine learning

Insight 2: Focus on hybrid quantum-classical approaches.
Cope with hardware constrains and available quantum resources



Challenges solved:

5. Benedetti, et al. A generative modeling approach for **benchmarking and training shallow quantum circuits**. *npj QI, 5, 45 (2019)*.

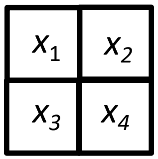
6. Zhu et al. **Training** of Quantum Circuits on a Hybrid Quantum Computing. *arXiv:1812.08862 (2018).*

7. Leyton-Ortega, et al. **Robust** Implementation of Generative Modeling with Parametrized Quantum Circuits. *arXiv:1901.08047 (2019).*

8. Leyton-Ortega, et al. **Benchmarking Optimizers** for Hybrid Quantum-Classical Algorithms. *To appear soon in arXiv.*

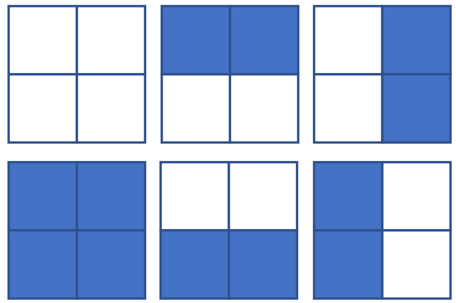
9. Alcazar et al. Classical **versus** Quantum Models in ML: Insights from a Finance Application. *arXiv:To appear soon in arXiv.*

Unsupervised generative modeling with NISQ devices



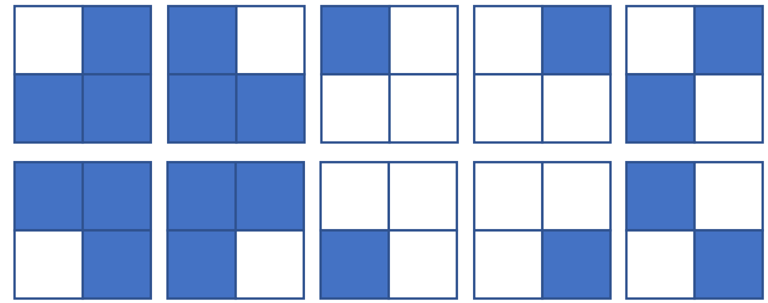
$x_1x_2x_3x_4$

BAS patterns



{0000,1100, 0101,
1111,0011,1010}

Non-BAS pattern



Ludwig Boltzmann (1844-1906)

Boltzmann machines

$$P_{Boltzman} \propto exp[-\xi(s_1, ..., s_N)/T_{eff}]$$

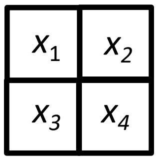
Max Born (1882-1970)

Quantum Circuit Born machines (QCBM) $P_{\boldsymbol{\theta}}(\mathbf{x}) = |\langle \mathbf{x} | \psi(\boldsymbol{\theta}) \rangle|^2.$

Generative modeling with NISQ devices, beyond Boltzmann

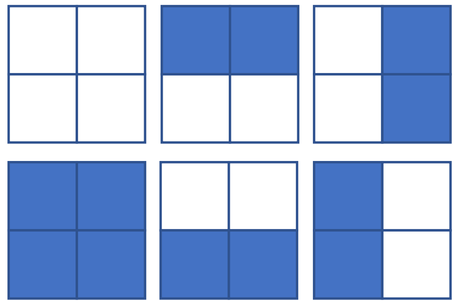
QCBMs: Benedetti, Garcia-Pintos, Perdomo, Leyton-Ortega, Nam, and Perdomo-Ortiz. A generative modeling approach for benchmarking and training shallow quantum circuits. **npj QI, 5, 45 (2019).**

Unsupervised generative modeling with NISQ devices



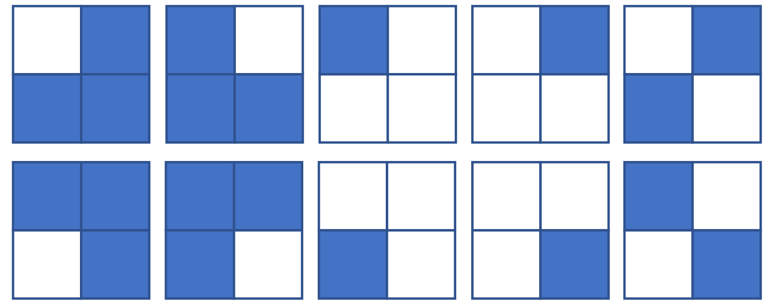
$x_1x_2x_3x_4$

BAS patterns



{0000,1100, 0101, 1111,0011,1010}

Non-BAS pattern



Ludwig Boltzmann (1844-1906)

Boltzmann machines

$$P_{Boltzman} \propto exp[-\xi(s_1, ..., s_N)/T_{eff}]$$

Max Born (1882-1970)

Born machines $P_{\boldsymbol{\theta}}(\mathbf{x}) = |\langle \mathbf{x}|\psi(\boldsymbol{\theta})\rangle|^2.$

Tensor networks: Chen, Chen, Wang. **arXiv:1712.04144**.

Generative Modeling.

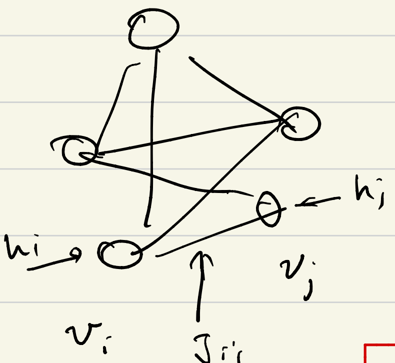
Dataset:

$\{\vec{v}^{(i)}\}_{i=1}^{D}$ D samples that are i.i.d. from $p(\vec{v})$
↳ data dist.

Find $\vec{\theta}$: $q(\vec{v}|\theta) \simeq p(\vec{v})$ → learning

Boltzmann Machines (BMs)

Energy PGM



$$\vec{v}^{(i)} = v_1^{(i)}\, v_2^{(i)} \ldots v_N^{(i)}$$

$\vec{v}$ = visible units

$$E(\vec{v}, \{\vec{h}, \vec{J}\}) = \sum_i h_i v_i + \sum_{ij} J_{ij} v_i v_j$$

↓

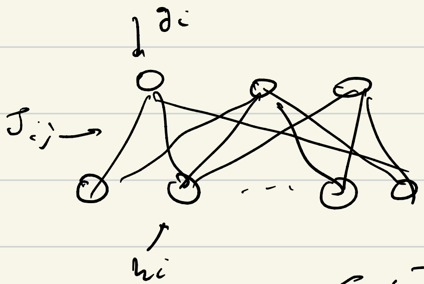
$$E(\vec{v}|\theta)$$

FVBM

$$q_\theta(\vec{v}) = \frac{e^{-E(\vec{v}|\theta)/T}}{Z(\theta)}$$

$$Z(\theta) = \sum_{\vec{v}} e^{-E(\vec{v}|\theta)/T}$$

Restricted Boltzmann Machine (RBM)



$\vec{u}$ = unseen or hidden variable

$\vec{v}$ = visible units

$$E(\vec{v},\vec{u}|\theta) = \sum_i h_i v_i + \sum_i g_i u_i + \sum_{ij} J_{ij} v_i u_j$$

$$\theta = \{\vec{h}, \vec{g}, \vec{J}\}$$

$$q_\theta(\vec{v},\vec{u}) = \frac{e^{-E(\vec{v},\vec{u}|\theta)/T}}{Z(\theta)}$$

$$\boxed{q_\theta(\vec{v}) = \sum_{\vec{u}} q_\theta(\vec{v},\vec{v})}$$

Likelihood and NLL(θ):

$$\mathcal{L}(\theta) = \prod_{i=1}^{D} q_\theta(v^{(i)})$$

$$NLL(\theta) = -\log \mathcal{L}(\theta) = -\sum_{i=1}^{D} \log q_\theta(\vec{v}^{(i)})$$

$$\approx -D \langle \log q_\theta(\vec{v}) \rangle_{p(\vec{v})}$$

Kullback-Leibler Divergence (KL)

(cross-entropy)

$$KL(p(\vec{v}) \| q(\vec{v})) = \sum_{\vec{v}} p(\vec{v}) \log \frac{p(\vec{v})}{q(\vec{v})}$$

$KL \geq 0$ $\quad$ $KL = 0$ iff $p = q$

$$KL(p \| q_\theta) = \sum_{\vec{v}} p(\vec{v}) \log p(\vec{v}) - \sum_{\vec{v}} p(\vec{v}) \log q_\theta(\vec{v})$$

(-entropy) → 0 when taking gradient

$$\nabla_\theta KL(p \| q_\theta) = -\sum_{\vec{v}} p(\vec{v}) \nabla_\theta \log q_\theta(\vec{v})$$

$$= -\nabla_\theta \langle \log q_\theta(\vec{v}) \rangle_{p(\vec{v})}$$

Averaging!

$$\langle A(x) \rangle_{p(x)} = \sum_x p(x) A(x) \simeq \frac{1}{N} \sum_{i=1} A(x^{(i)})$$

if $\vec{x}^{(i)}$ is iid from $p(x)$

Stochastic Gradient Descent (SGD)

$$\theta_i = \theta_i - \eta \frac{\partial NLL(\theta)}{\partial \theta_i}$$

Example for FVBM:

$$E(\vec{v}|\theta) = \sum h_i v_i + \sum_{ij} J_{ij} v_i v_j$$

$$q_\theta(\vec{v}) = \frac{e^{-E(\vec{v}|\theta)/T}}{Z(\theta)} \qquad Z(\theta) = \sum_{\vec{v}} e^{-E(\vec{v}|\theta)/T}$$

$$NLL(\theta) = -\sum_{i=1}^{D} \log q_\theta(\vec{v})$$

$$= -\sum_{i=1}^{D} \left(-\frac{E(\vec{v}|\theta)}{T} - \log Z(\theta) \right)$$

$\partial \log Z(\theta)$

$$\frac{\partial NLL(\theta)}{\partial h_j} = \sum_{i=1}^{D} \frac{v_j^{(i)}}{T} + \sum_{i=1}^{D} \frac{1}{Z(\theta)} \sum_{\vec{v}} \left(\frac{-v_j}{T}\right) e^{-E(\vec{v}|\theta)/T}$$

$$= \frac{D}{T} \left(\frac{1}{D} \sum_{i=1}^{D} v_j^{(i)} - \sum_{\vec{v}} v_j q_\theta(\vec{v}) \right)$$

$$= \frac{D}{T} \left(\langle v_j \rangle_{p(\vec{v})} - \langle v_j \rangle_{q_\theta(\vec{v})} \right)$$

$$\frac{\partial NLL(\theta)}{\partial h_j} = \frac{D}{T} \left(\langle v_j \rangle_{data} - \langle v_j \rangle_{model} \right)$$

$$\frac{\partial NLL(\theta)}{\partial J_{ij}} = \frac{D}{T}\left(\langle v_i v_j \rangle_{data} - \langle v_i v_j \rangle_{model} \right)$$

MCMC.