Machine Learning for Many-Body Physics

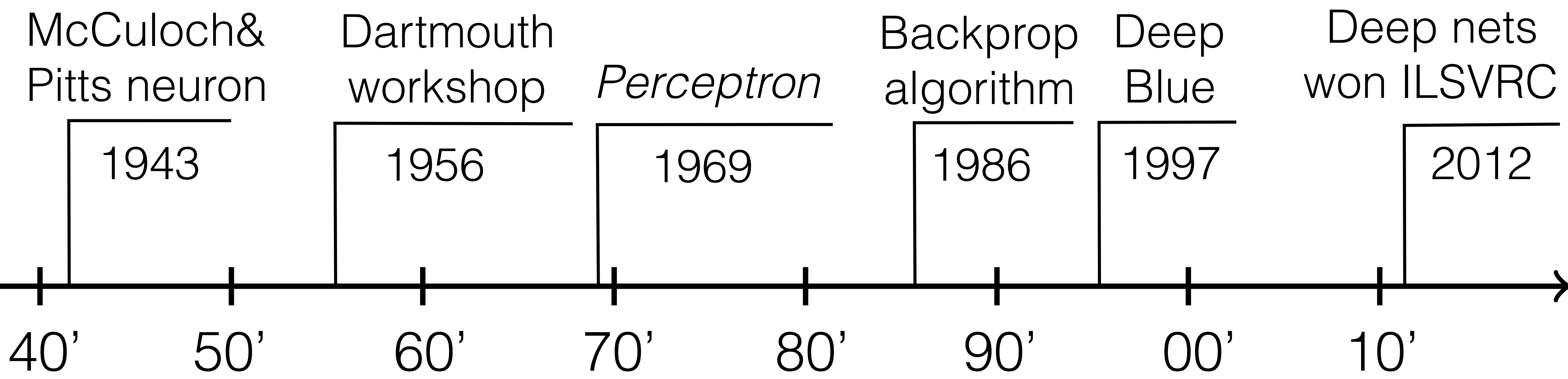
A Literature Survey

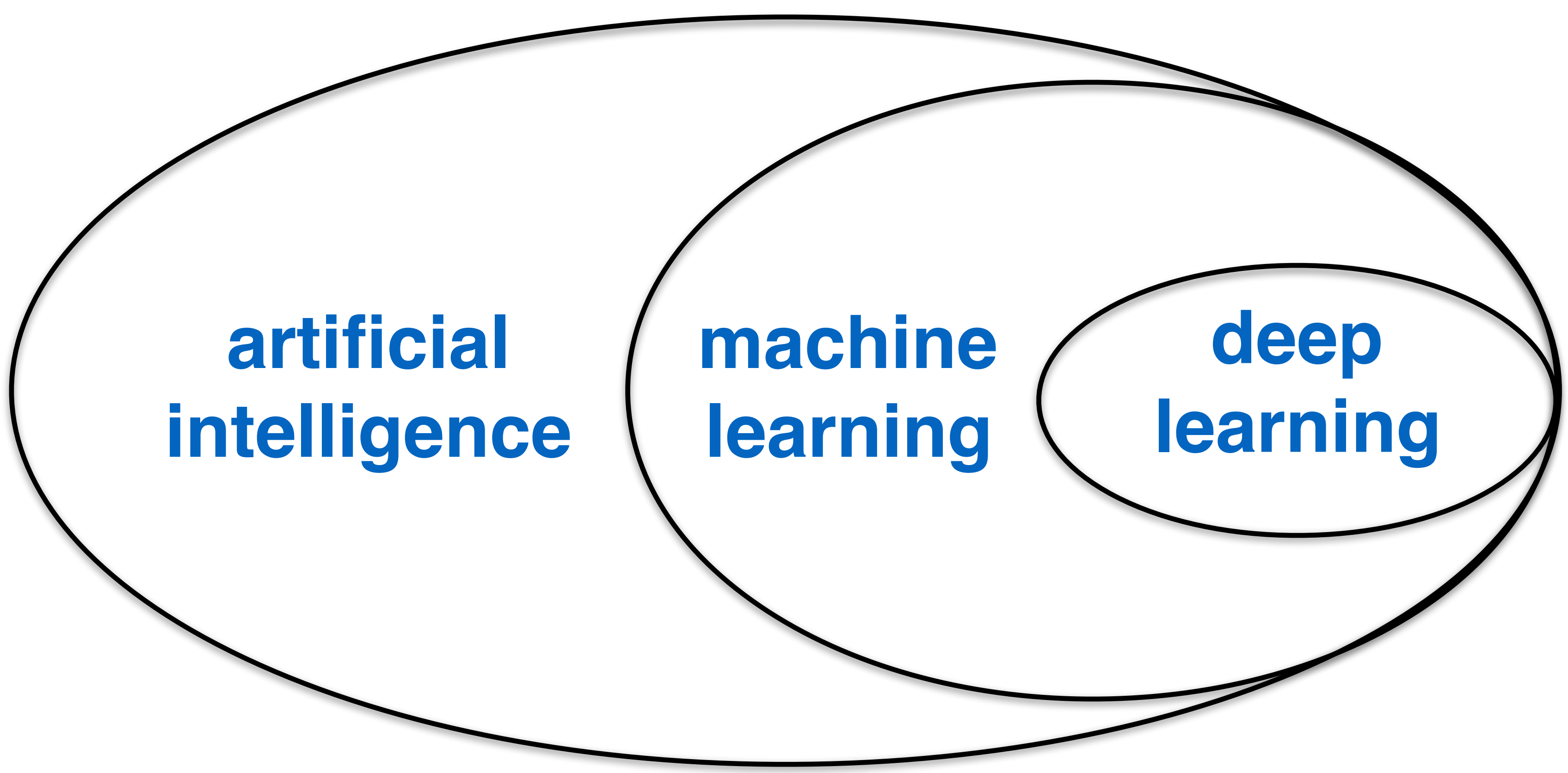
Lei Wang
Institute of Physics, CAS
https://wangleiphy.github.io

Disclamer

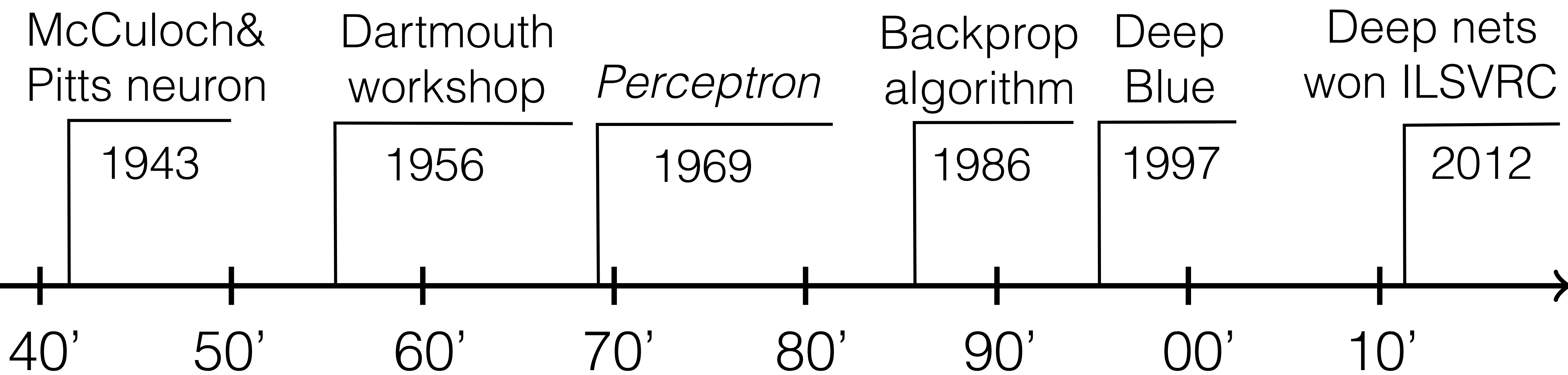
- It will be VERY subjective
- Not all important papers are reviewed
- Some are shown for negative reasons
- The purpose is to trigger discussion

Timeline of AI research





Timeline of AI research

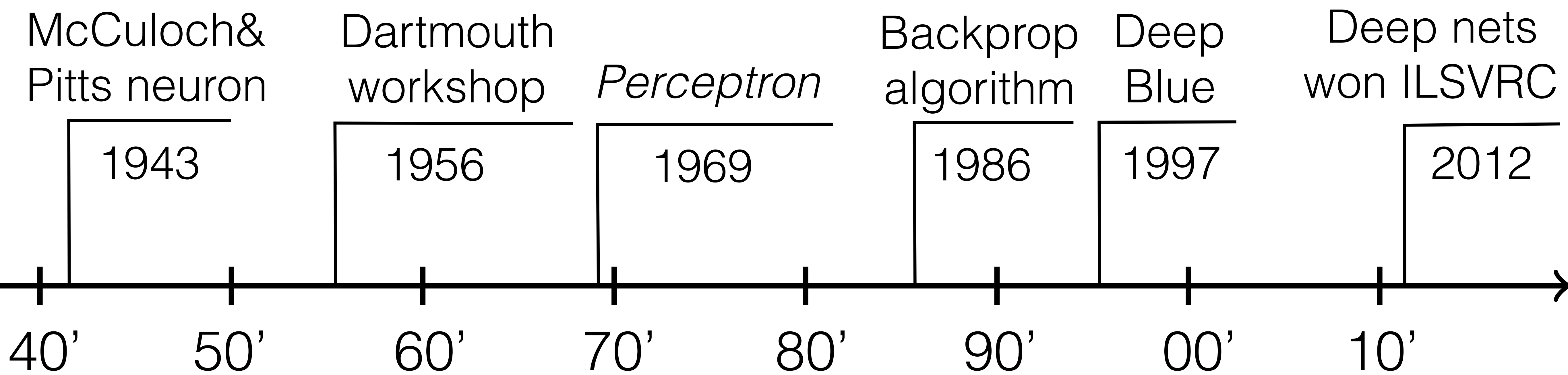


Early enthusiasm

(reasoning)

"Look, Ma, no hands"

Timeline of AI research



Early
enthusiasm

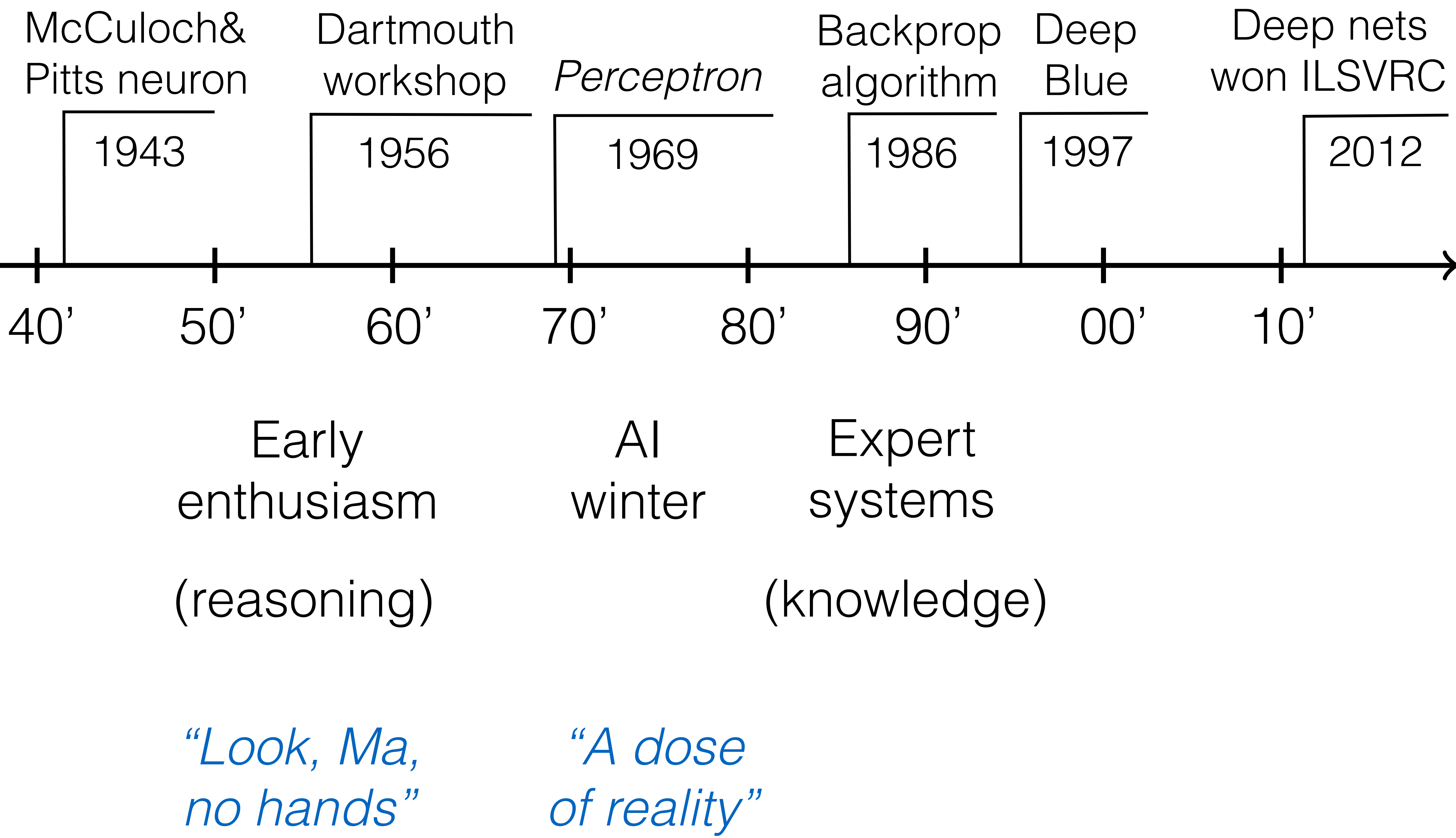
(reasoning)

AI
winter

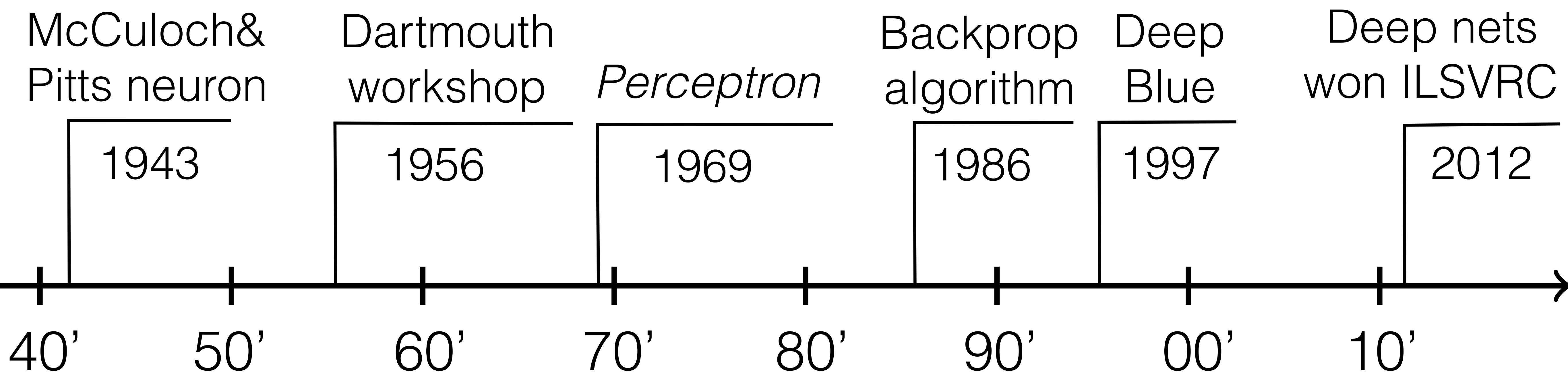
*“Look, Ma,
no hands”*

*“A dose
of reality”*

Timeline of AI research



Timeline of AI research



Early enthusiasm (reasoning)

AI winter

Expert systems (knowledge)

Deep Learning (data & representation)

“Look, Ma, no hands”

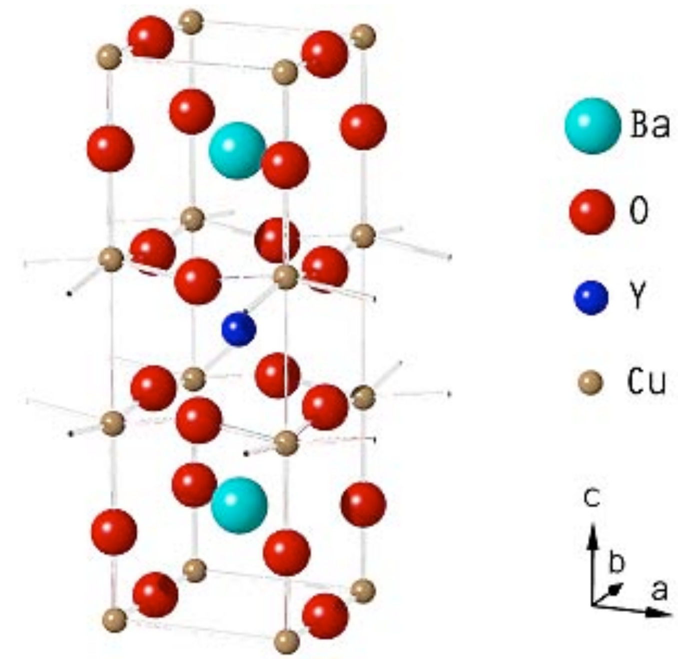
“A dose of reality”

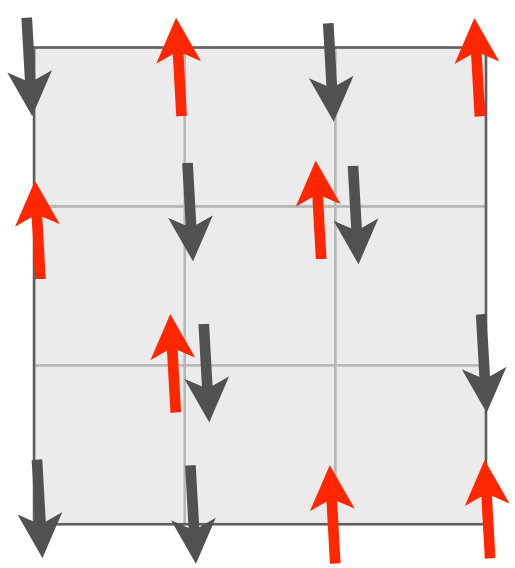
“Return of neural nets”

Lessons

- One can define the direction of a new field
- Overhype hurts the field in the long term
- Be practical: really solve problems
- 宠辱不惊，看庭前花开花落

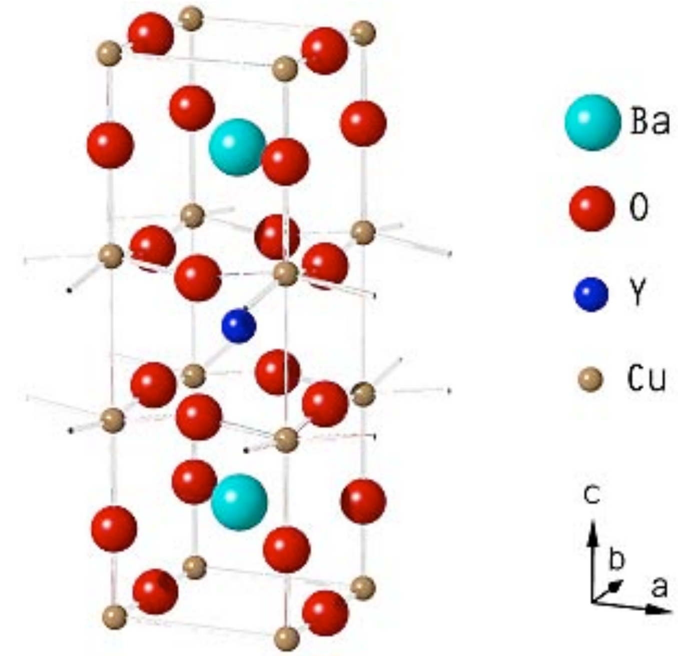
The problem we'd like to solve

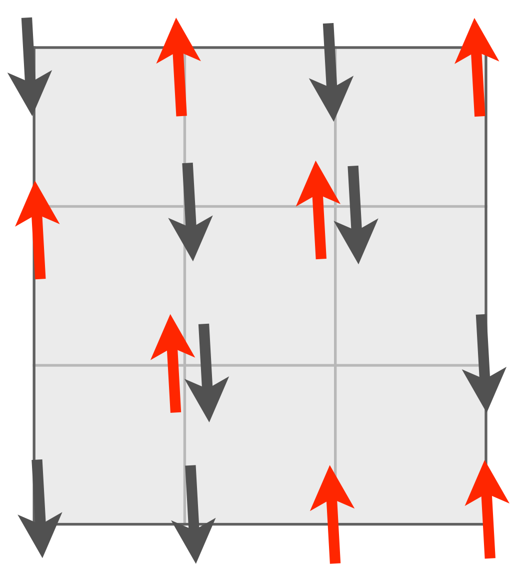




$4^{361} \approx 2.2 \times 10^{217} \gg$ # of atoms in the universe

The problem we'd like to solve





$4^{361} \approx 2.2 \times 10^{217} \gg$ # of atoms in the universe

The problem we'd like to solve

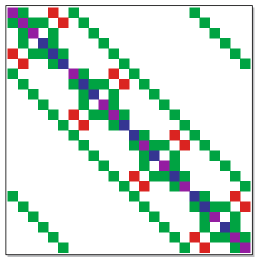
$$\boxed{i\frac{\partial}{\partial t}\Psi(\mathbf{X},t) = H\Psi(\mathbf{X},t)}$$

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \qquad N \sim 10^{23}$$

$$H = -\sum_i \frac{\nabla^2_{\mathbf{x}_i}}{2m} - \sum_{i,j} \frac{Z_j}{|\mathbf{x}_i - \mathbf{R}_j|} + \frac{1}{2}\sum_{i\neq j} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} + \ldots$$

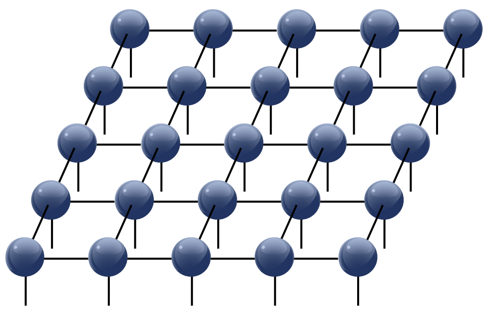
"Theory of everything" for condensed matter, chemistry and biology (including neuroscience)

Computational quantum many-body physics

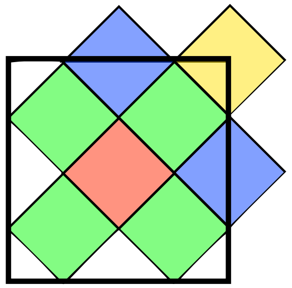


exact diagonalization

quantum Monte Carlo

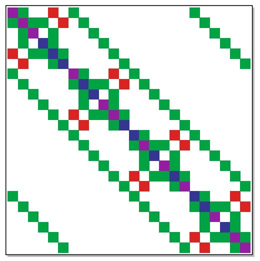


tensor network states



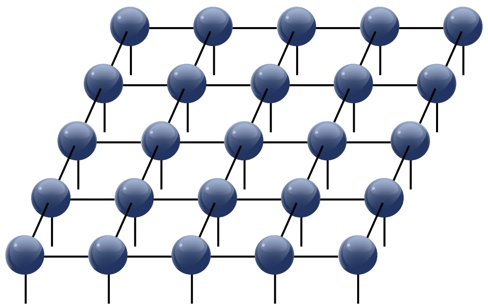
dynamical mean field theories

Computational quantum many-body physics

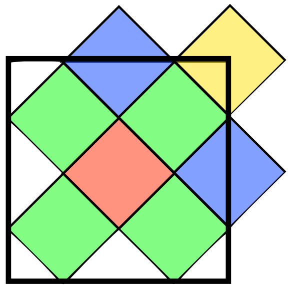


exact diagonalization

quantum Monte Carlo



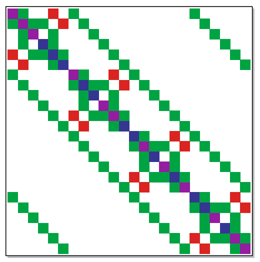
tensor network states



dynamical mean field theories

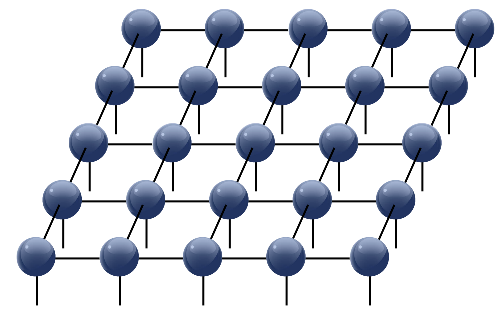
Algorithmic improvement in past 20 years outperformed Moore's law

Computational quantum many-body physics

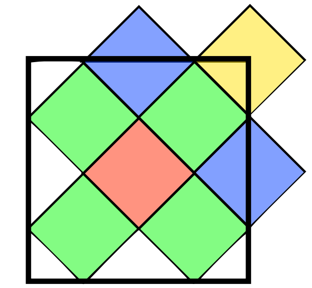


exact diagonalization

quantum Monte Carlo



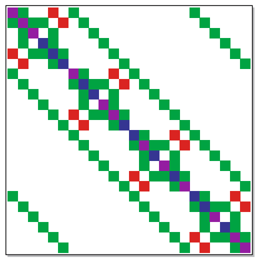
tensor network states



dynamical mean field theories

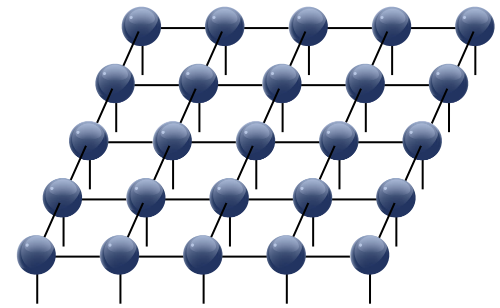
is faster than

Computational quantum many-body physics

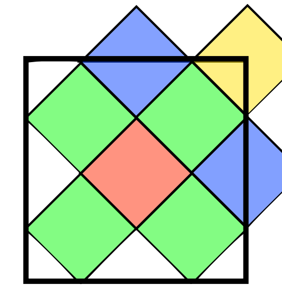


exact diagonalization

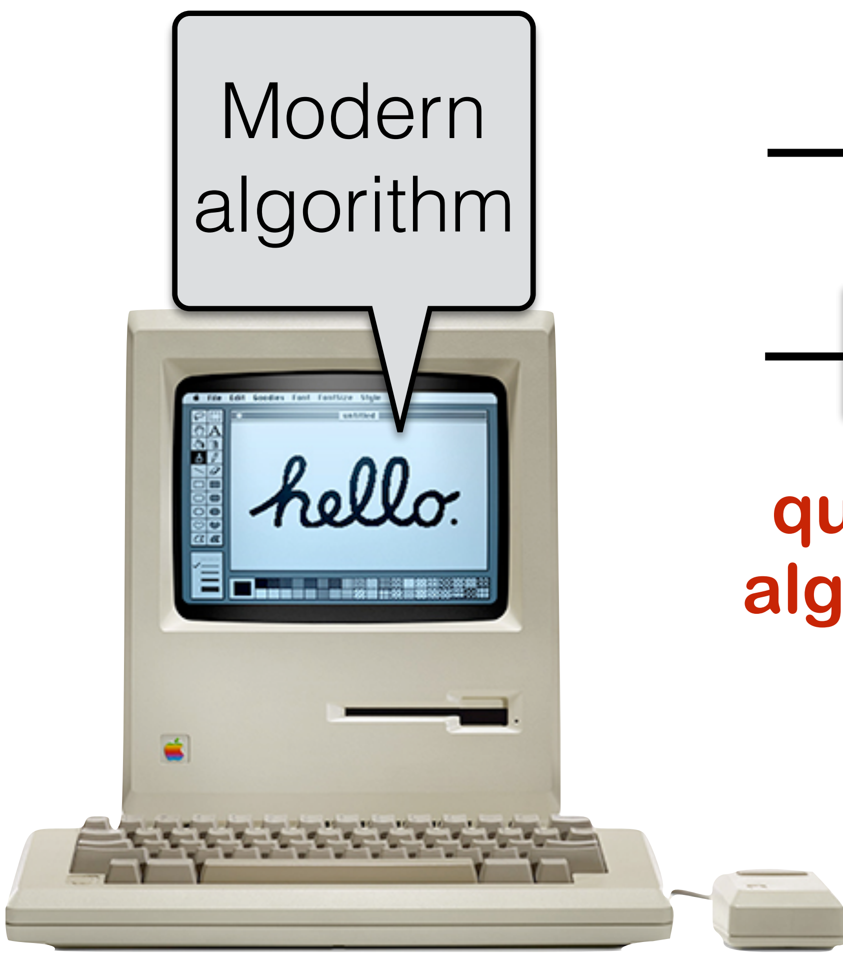
quantum Monte Carlo

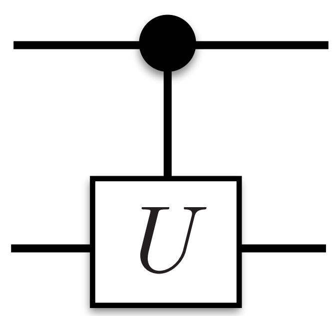


tensor network states

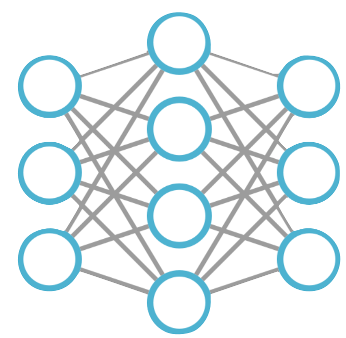


dynamical mean field theories

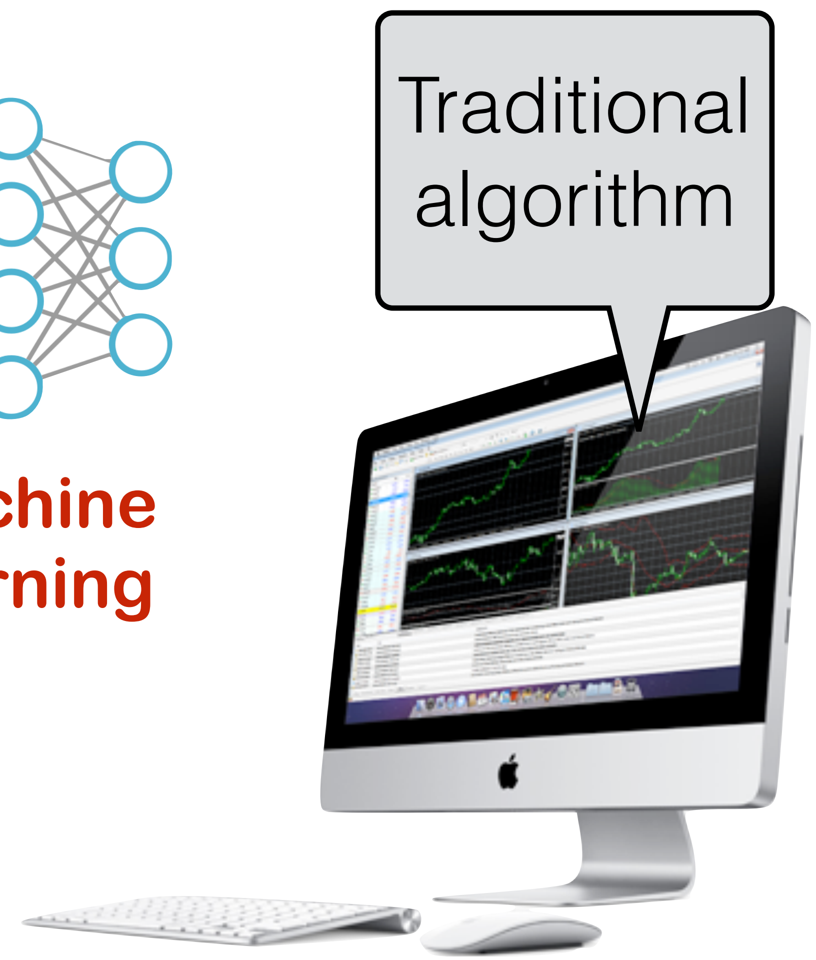




quantum algorithms

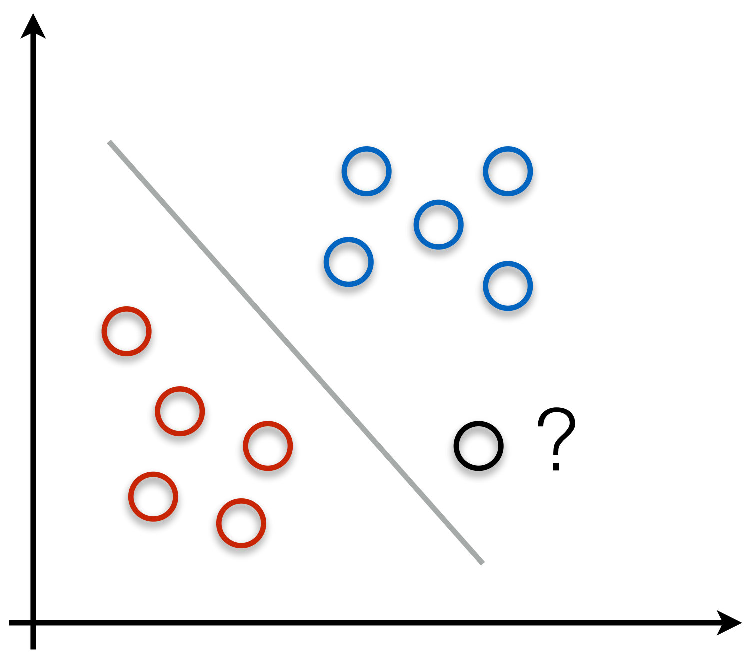


machine learning



Machine Learning 101

Supervised learning

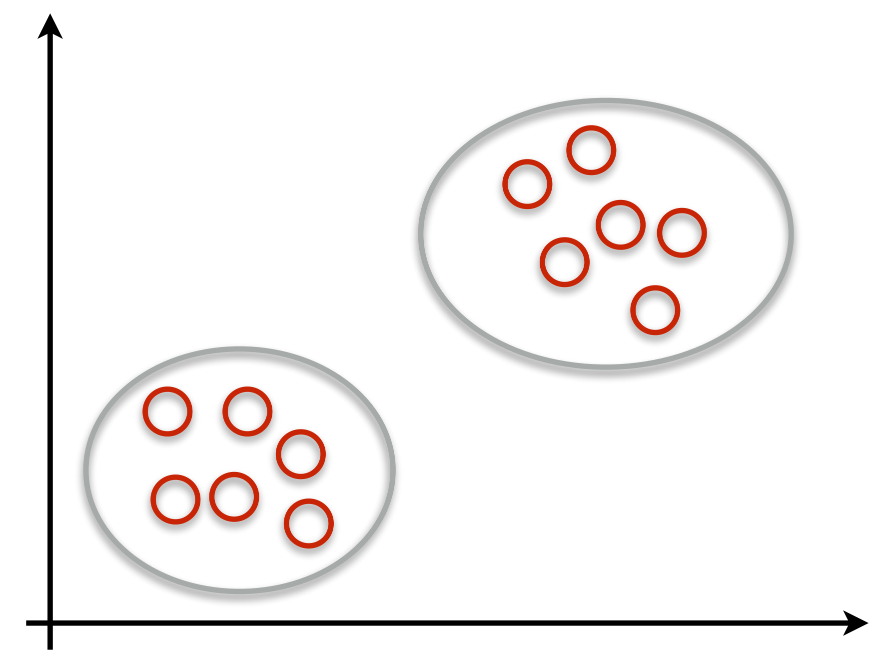


Classification

Spam detection

Image recognition

Unsupervised learning



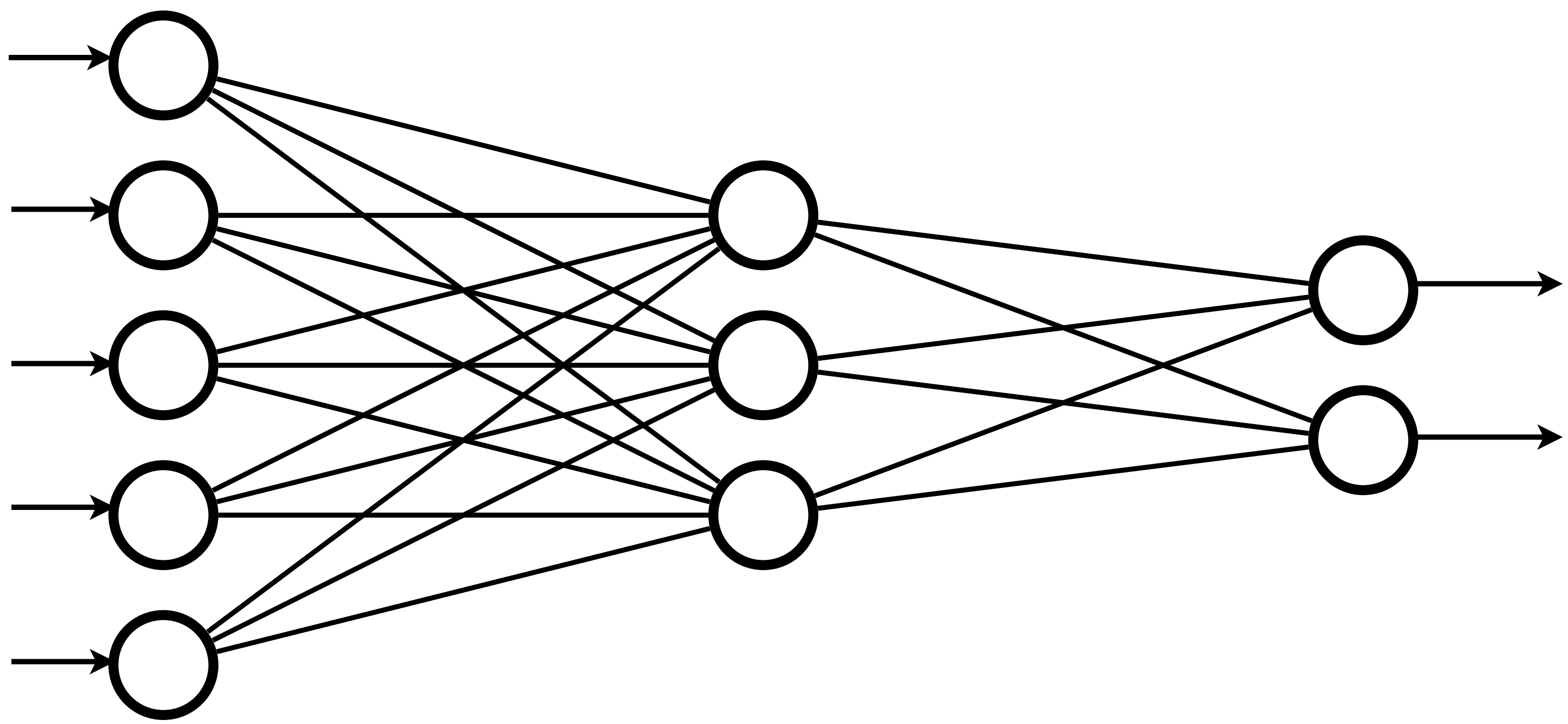
Clustering

Online advertising

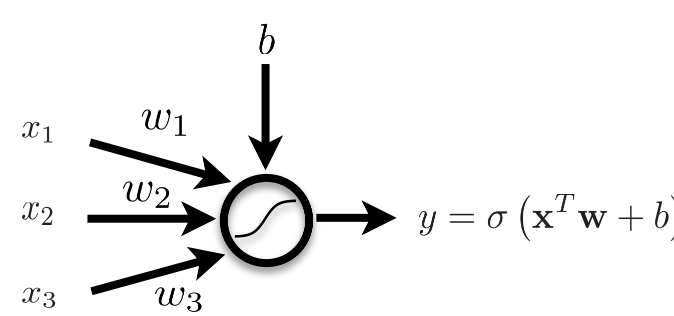
Recommender system

Machine Learning 102

neural network

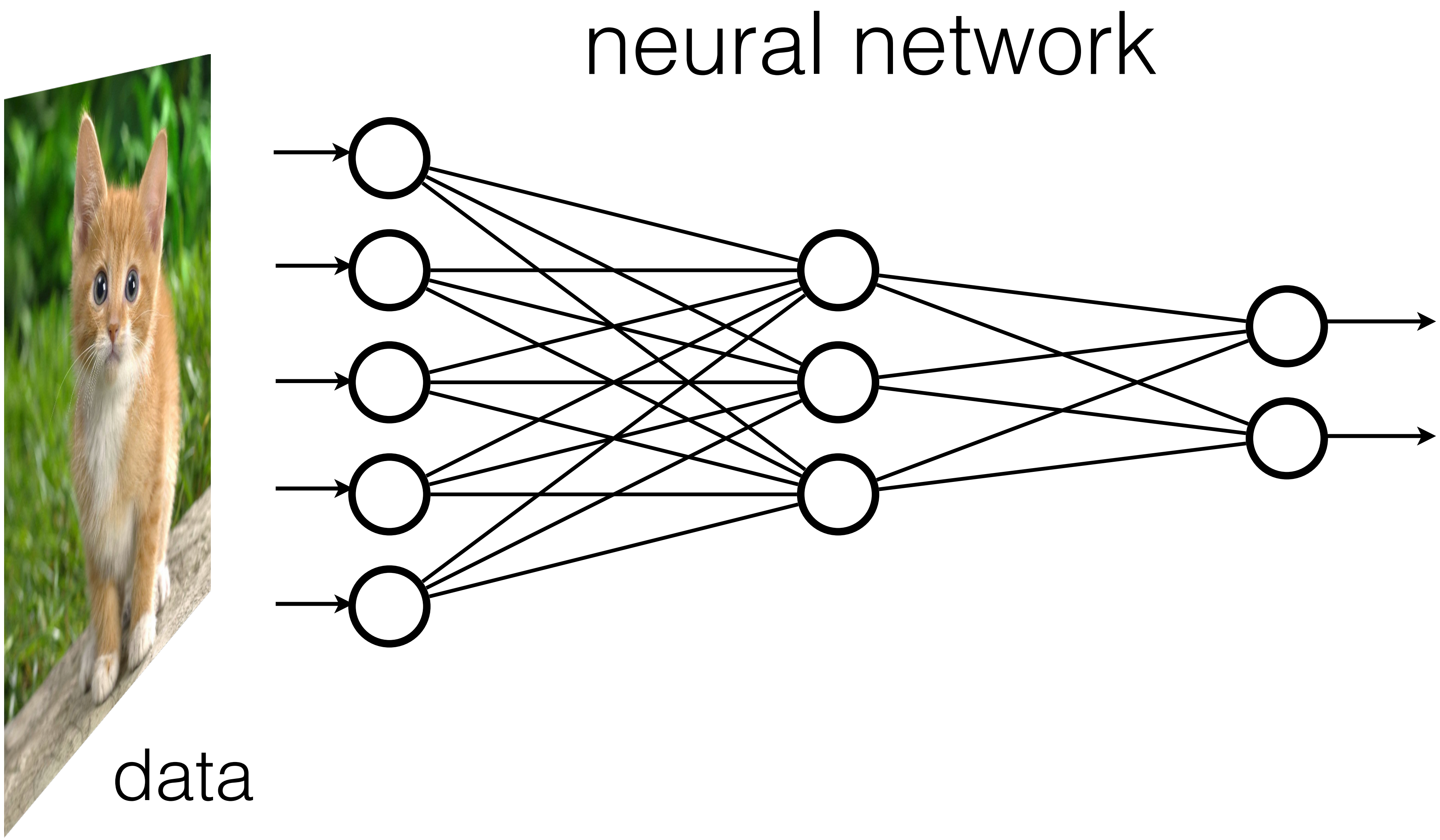


computing unit

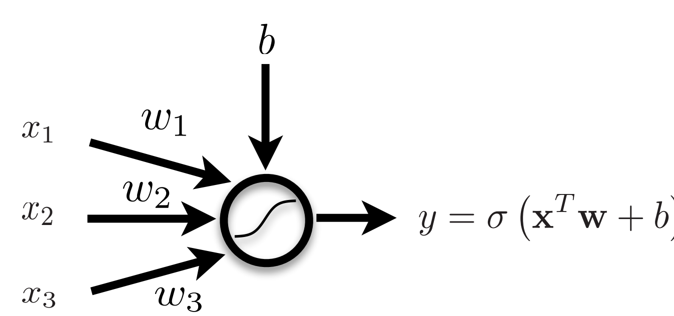


artificial neuron

Machine Learning 102

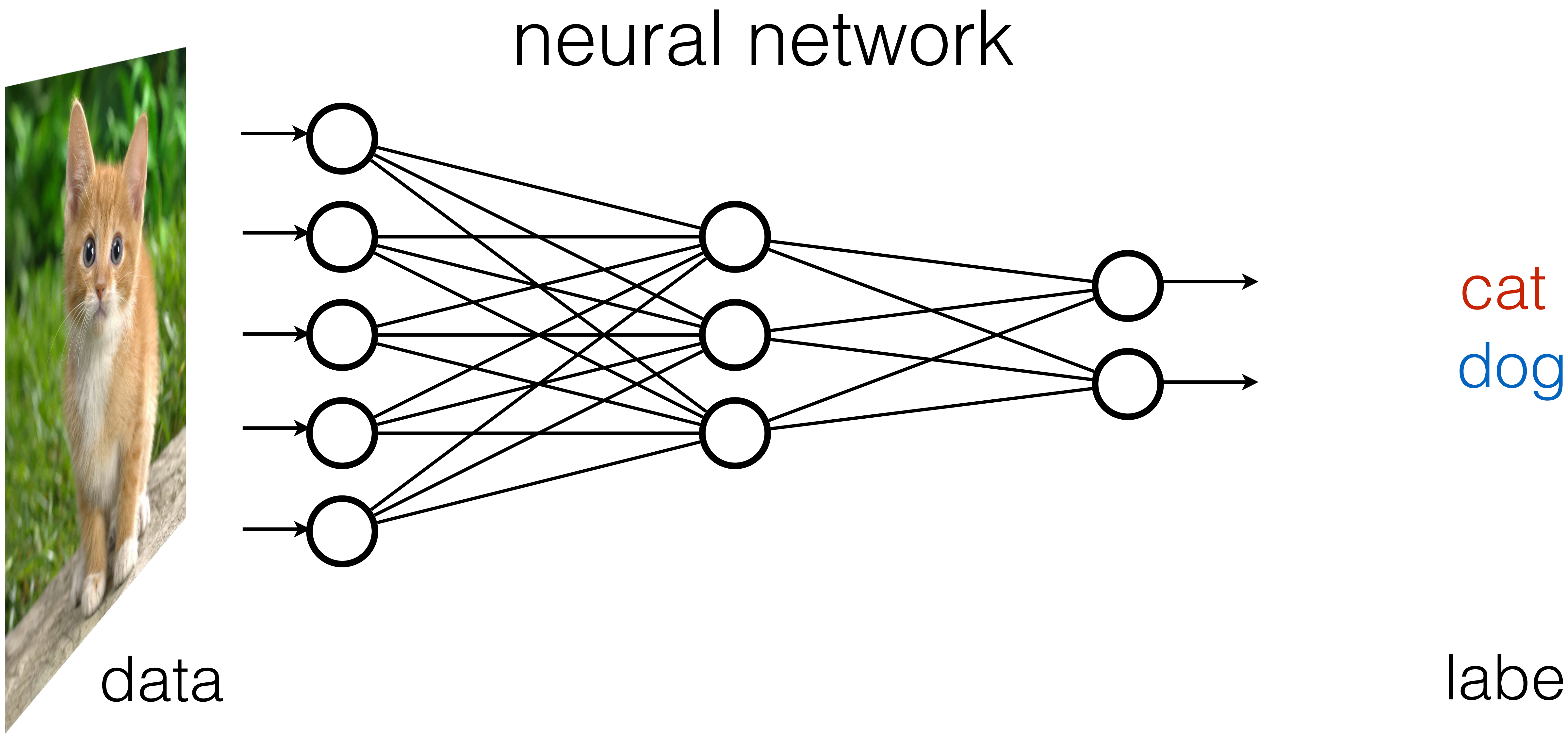


computing unit

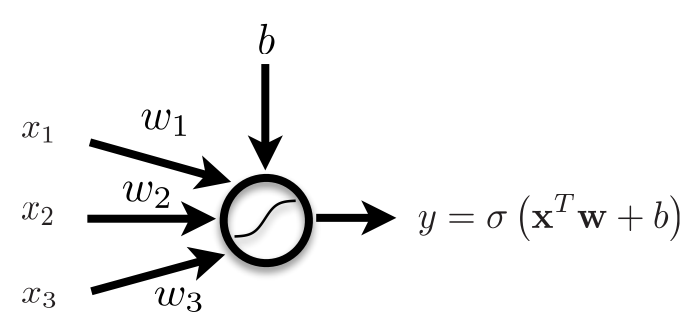


artificial neuron

Machine Learning 102

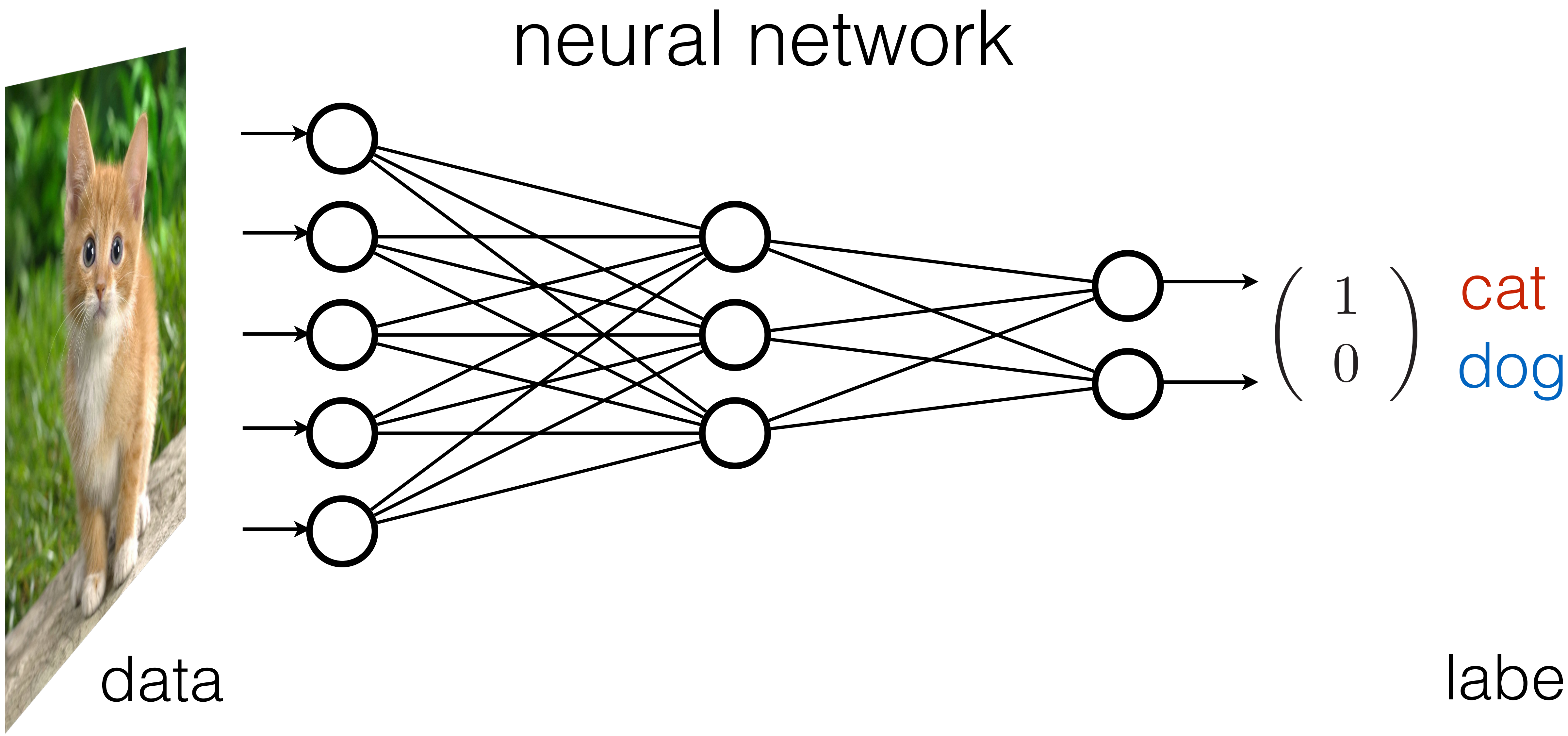


computing unit

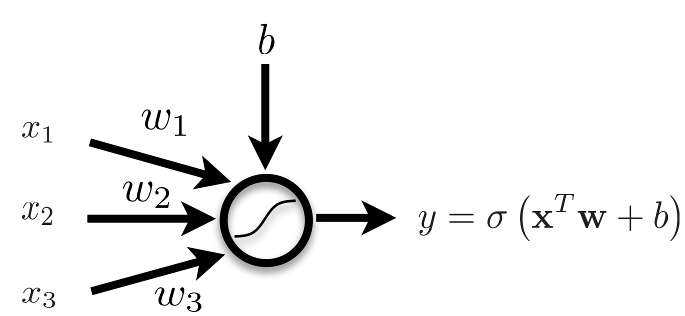


artificial neuron

Machine Learning 102

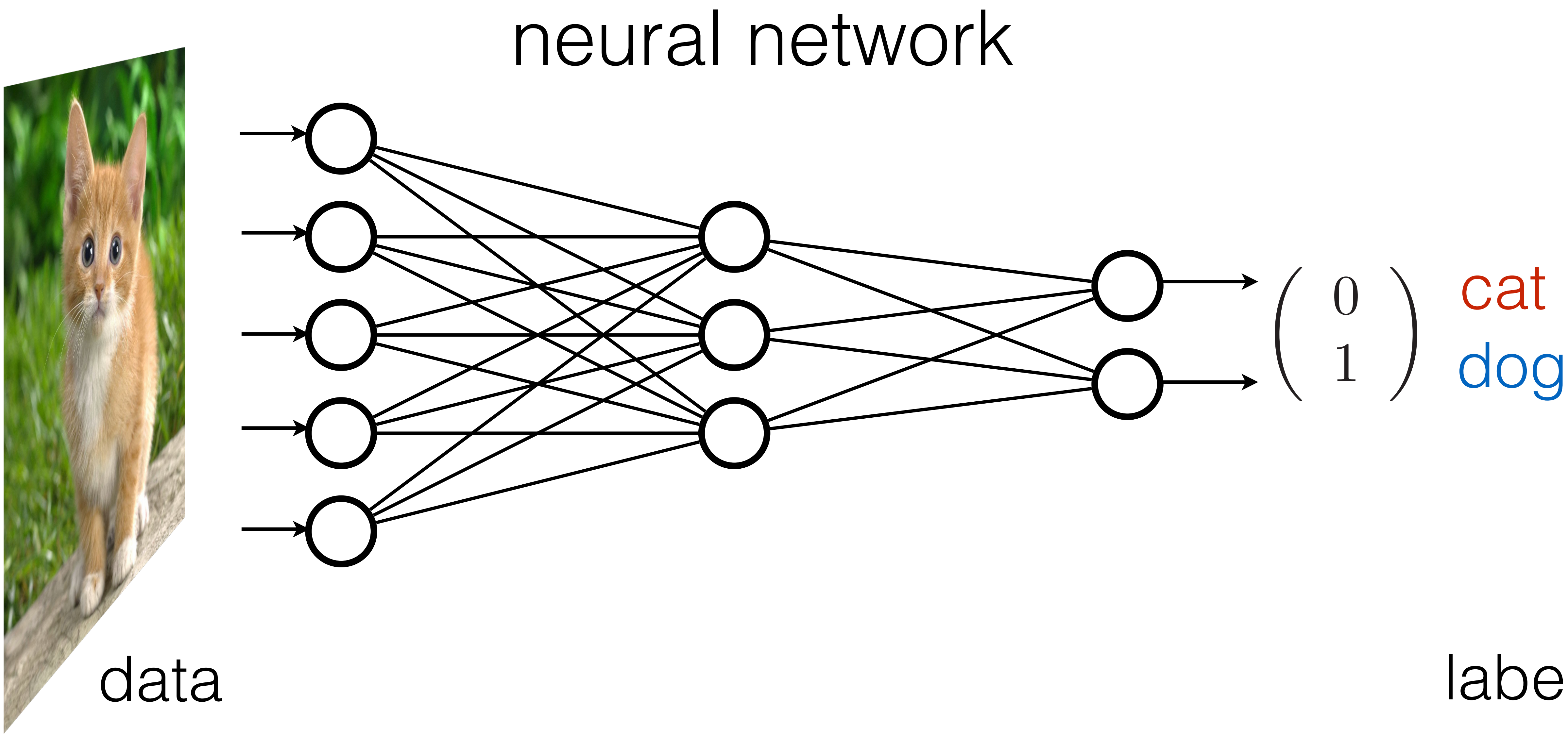


computing unit

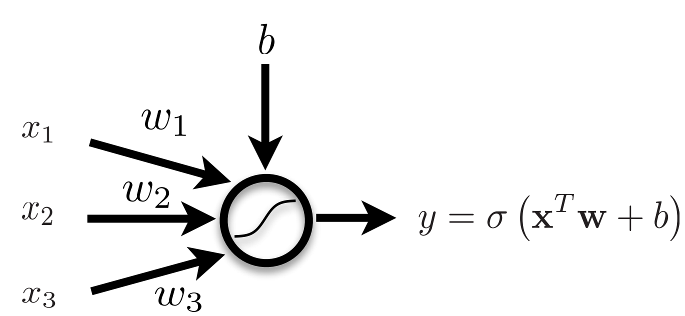


artificial neuron

Machine Learning 102



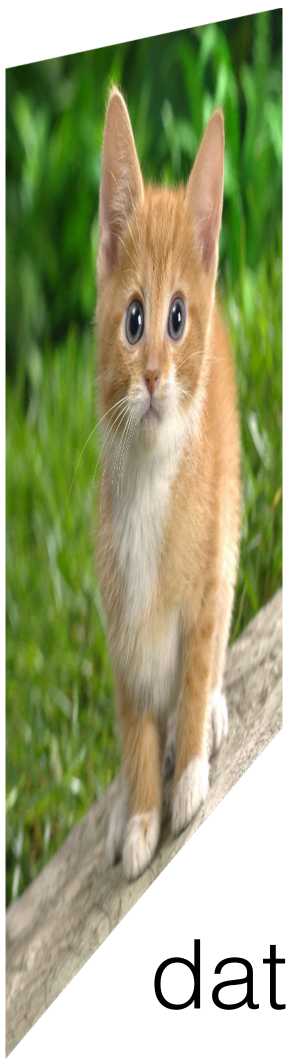
computing unit

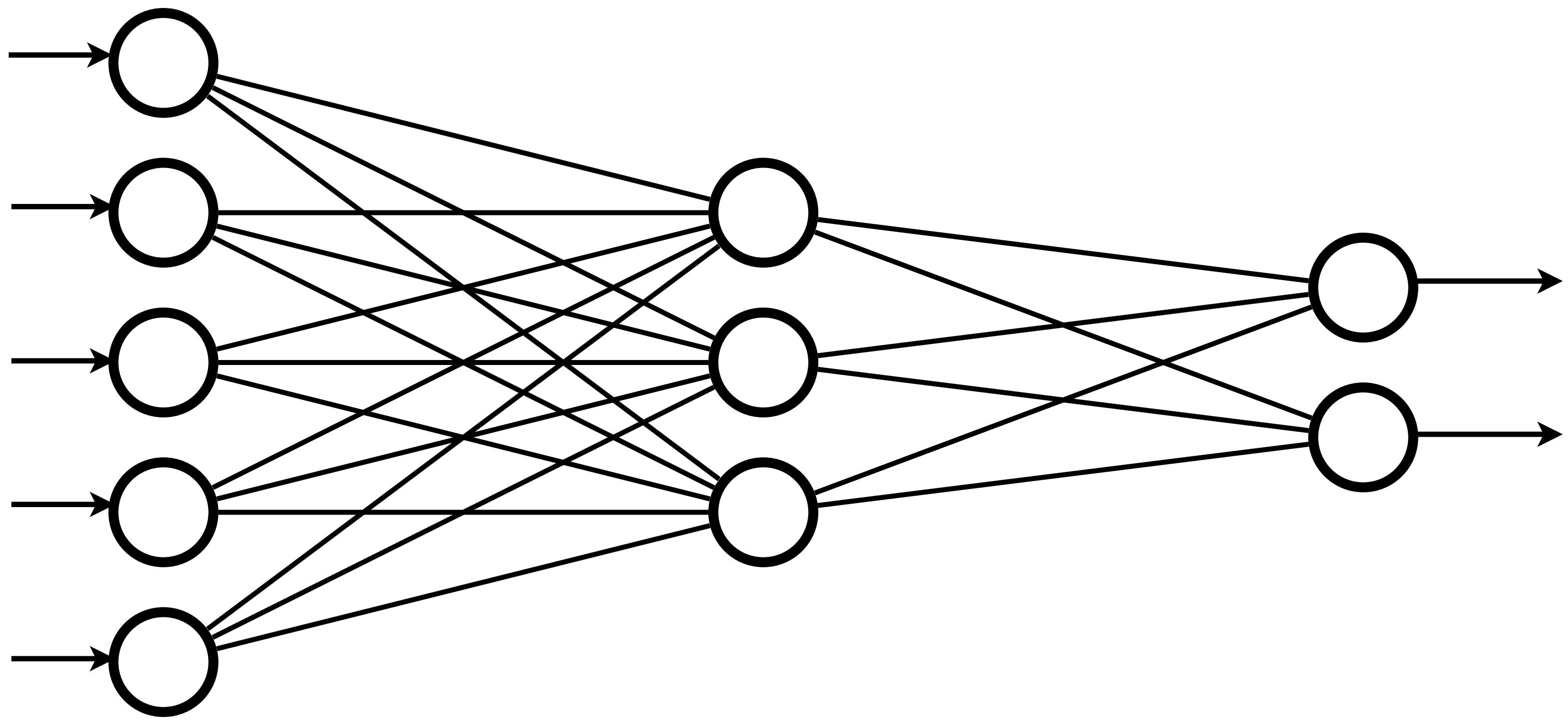


artificial neuron

Machine Learning 102

neural network





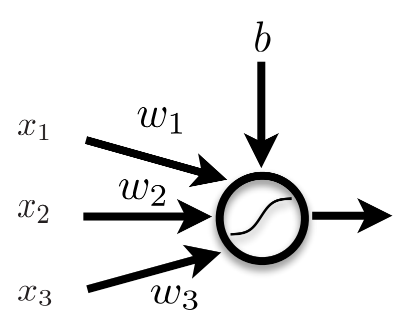
$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$ cat dog

data

universal function approximator

label

computing unit



$$y = \sigma\left(\mathbf{x}^T\mathbf{w} + b\right)$$

artificial neuron

WARNING

The following content may contain spoilers

They may spoil your fun of imagination & creation

Proceed with caution!!!

Deep Learning and RG

QUANTA MAGAZINE illuminating science

A Common Logic to Seeing Cats and Cosmos

Olena Shmahalo / Quanta Magazine

There may be a universal logic to how physicists, computers and brains tease out important features from among other irrelevant bits of data.

"An exact mapping between the Variational Renormalization Group and Deep Learning", Mehta and Schwab, 1410.3831

Deep Learning and RG

Renormalization Group

$$H(\mathbf{x}) = -\sum_i K_i x_i - \sum_{ij} K_{ij} x_i x_j - \dots$$

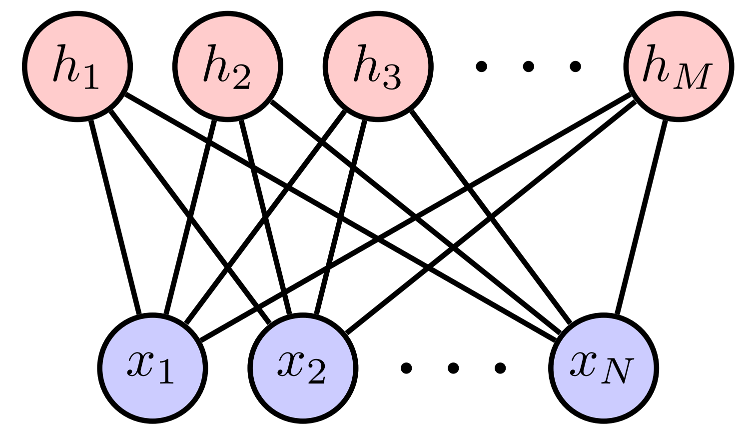
$$e^{-H_{\mathrm{RG}}(\mathbf{h})} \equiv \sum_{\mathbf{x}} e^{\boxed{T(\mathbf{x},\mathbf{h})} - H(\mathbf{x})}$$

RG Transformation

$$H_{\mathrm{RG}}(\mathbf{h}) = -\sum_i \tilde{K}_i h_i - \sum_{ij} \tilde{K}_{ij} h_i h_j - \dots$$

Restricted Boltzmann Machine

$$E(\mathbf{x},\mathbf{h}) = -\sum_{i=1}^{N} a_i x_i - \sum_{j=1}^{M} b_j h_j - \sum_{i=1}^{N}\sum_{j=1}^{M} x_i W_{ij} h_j$$

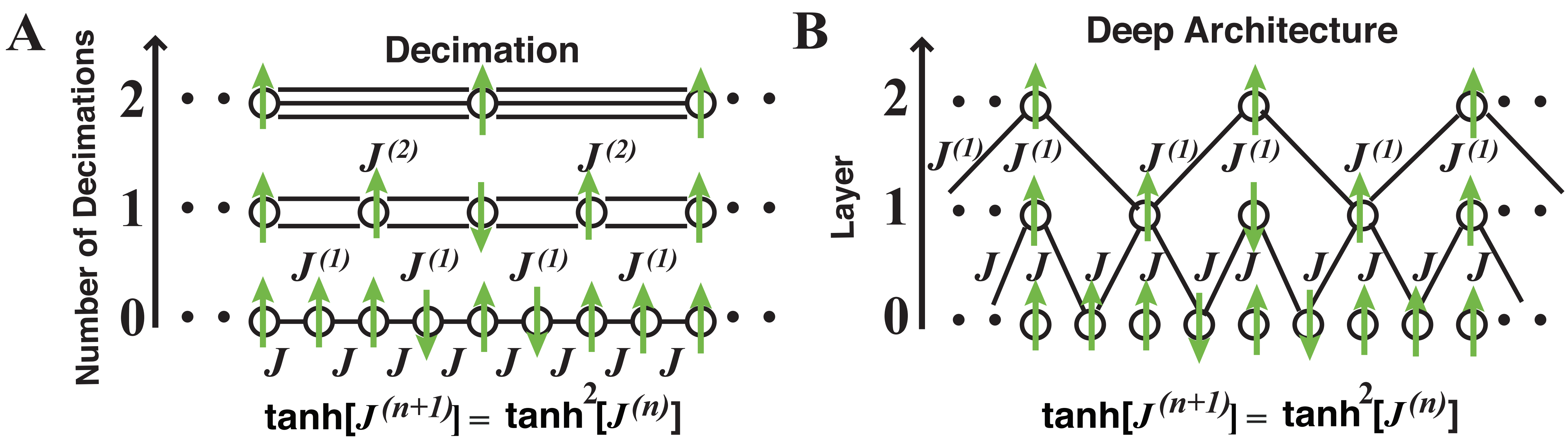


$$e^{-H_{\mathrm{RBM}}(\mathbf{h})} \equiv \sum_{\mathbf{x}} e^{-E(\mathbf{x},\mathbf{h})}$$

$$= e^{\sum_{j=1}^{M} b_j h_j} \prod_{i=1}^{N} \left(1 + e^{a_i + \sum_{j=1}^{M} W_{ij} h_j}\right)$$

“An exact mapping between the Variational Renormalization Group and Deep Learning”, Mehta and Schwab, 1410.3831

Deep Learning and RG



1D Ising model: Block spin RG vs deep neural nets

cf. “Deep learning and the renormalization group”, Bény, 1301.3124
“PCA meets RG”, Braddea and Bialek, 1610.09733

Dictionary: RG vs deep learning

Property	Variational RG	Deep Belief Networks
How input distribution is defined	Hamiltonian defining P(v)	Data samples drawn from P(v)
How interactions are defined	T(v,h)	E(v,h)
Exact transformation	$Tr_{\mathbf{h}} e^{T(\mathbf{v},\mathbf{h})} = 1$	KL divergence between P(v) and variational distribution is zero
Approximations	Minimize or bound free energy differences	Minimize the KL divergence
Method	Analytic (mostly)	Numerical
What happens under coarse-graining	Relevant operators grow/irrelevant shrink	New features emerge

From Schwab's talk at PI: http://pirsa.org/displayFlash.php?id=16080006

Comment on the quantum magazine website

Noah says:
December 26, 2014 at 9:54 am

I just spend an hour reading Mehta-Schwab paper from the beginning to end. Let me say that "A Common Logic to Seeing Cats and Cosmos" is a sensationalist article about a trivial paper, which will have no impact whatsoever. The whole M-S paper is based on the fact that couplings of two systems appear in more than one context and that distributions can sometimes appear as marginal distributions on product spaces. There is no one-to-one mappings between renormalization group (RG) scheme of Kadanoff and Restricted Boltzmann Machines (RBM) in Deep Neural Networks (DNN) in their paper. What they show is that RBM can be represented as a RG scheme with a very specific choice of coupling function T in equation (18). Conveniently, this coupling function depends on the Hamiltonian of the spin system, which it normally should not. Equivalence in equations (8) and (9) is also not correct. Condition (9) of course implies that the scheme is exact, but not the other way around, unless the authors make some implicit assumptions about coupling function T not mentioned in the paper. The paper contains no non-trivial ideas, it does not "open up a door to something very exciting", and I will not hold my breath expecting new breakthroughs because of this connection.

https://www.quantamagazine.org/20141204-a-common-logic-to-seeing-cats-and-cosmos/

Deep learning and physics

MIT
Technology
Review

Computing

The Extraordinary Link Between Deep Neural Networks and the Nature of the Universe

Nobody understands why deep neural networks are so good at solving complex problems. Now physicists say the secret is buried in the laws of physics.

"Why does deep and cheap learning work so well?"

Lin and Tegmark, 1608.08225

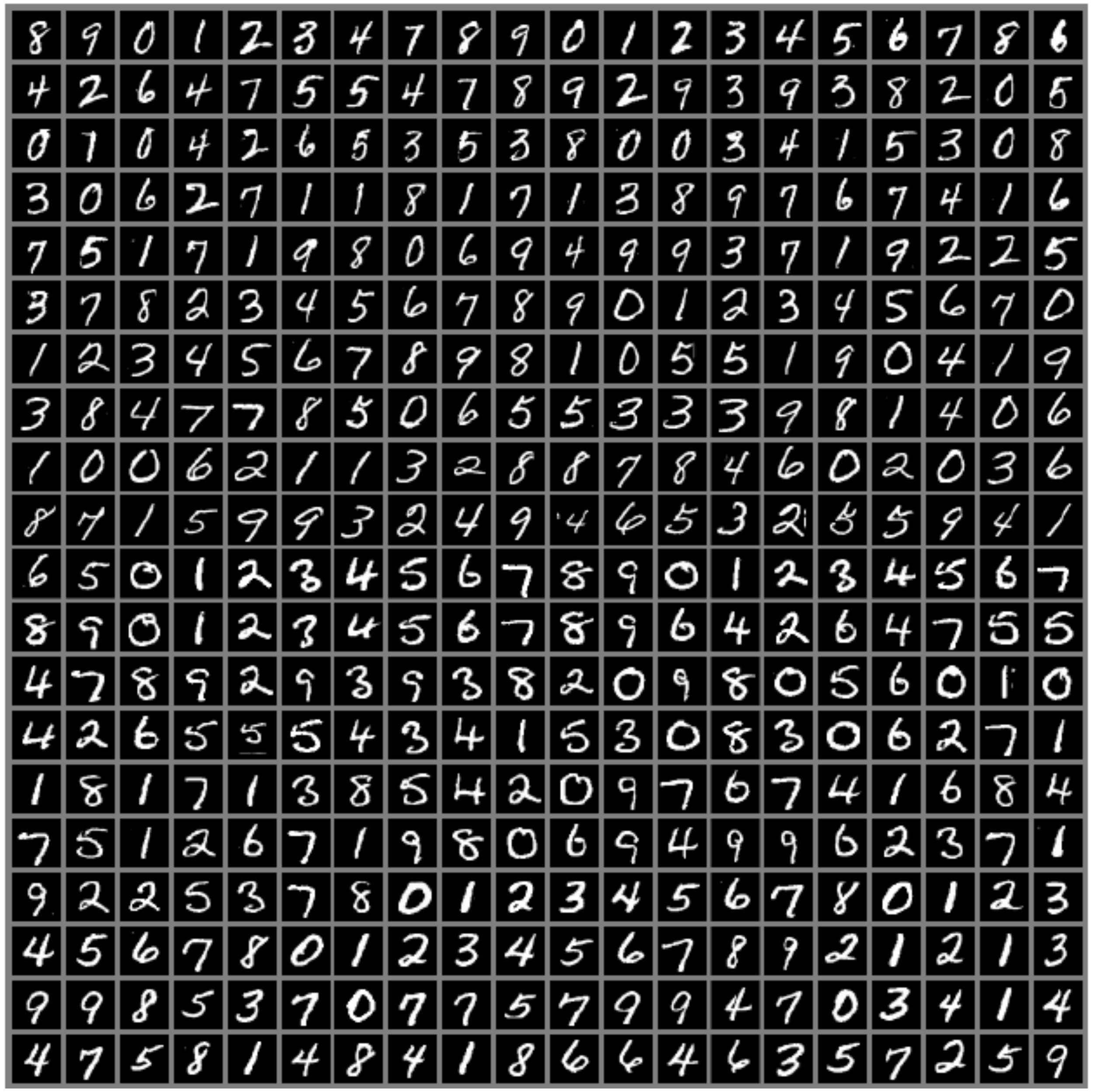
Why deep learning works ?

- It is not only a math, but also a physics question
- The class of functions of practical interests (natural scenes, drawings etc) can be approximated through "deep and cheap learning" because they follow the laws of physics
- Symmetry, locality, compositionality and polynomial log-probability

"Why does deep and cheap learning work so well?"

Lin and Tegmark,1608.08225

MNIST database



random images

from the “Deep Learning” book by Goodfellow, Bengio, Courville

https://www.deeplearningbook.org/

Physics-ML dictionary

Physics	Machine learning
Hamiltonian	Surprisal $-\ln p$
Simple H	Cheap learning
Quadratic H	Gaussian p
Locality	Sparsity
Translationally symmetric H	Convnet
Computing p from H	Softmaxing
Spin	Bit
Free energy difference	KL-divergence
Effective theory	Nearly lossless data distillation
Irrelevant operator	Noise
Relevant operator	Feature

"Why does deep and cheap learning work so well?"

Lin and Tegmark,1608.08225

Scott Aaronson’s comment

Several people wrote in to tell me about a recent paper by Henry Lin and Max Tegmark, which tries to use physics analogies and intuitions to explain why deep learning works as well as it does. To my inexpert eyes, the paper seemed to contain a lot of standard insights from computational learning theory (for example, the need to exploit symmetries and regularities in the world to get polynomial-size representations), but expressed in a different language. What confused me most was the paper’s claim to prove “no-flattening theorems” showing the necessity of large-depth neural networks—since in the sense I would mean, such a theorem couldn’t possibly be proved without a major breakthrough in computational complexity (e.g., separating the levels of the class TC^0). Again, anyone who understands what’s going on is welcome to share in the comments section.

http://www.scottaaronson.com/blog/?p=2918

More discussions

- Comment on "Why does deep and cheap learning work so well?" Schwab and Mehta, 1609.03541
- Why Deep Neural Networks? Liang and Srikant, 1610.04161
- Why and When Can Deep -- but Not Shallow -- Networks Avoid the Curse of Dimensionality: a Review, Poggio et al, 1611.00740
- Understanding Deep Neural Networks with Rectified Linear Units, Arora et al, 1611.01491

Why machine learning for many-body physics ?

- Conceptual connections: a new and natural way to think about (quantum) many-body systems

- Data driven approach: making scientific discovery based on big datasets

- Techniques: neural networks, kernel methods, feature extraction, dimensional reduction, clustering analysis, probabilistic modeling, deep learning, hardware acceleration…

Ideas

Ideas

Conceptual connections to RG

A general way to do fitting and interpolations

Solving inverse problems

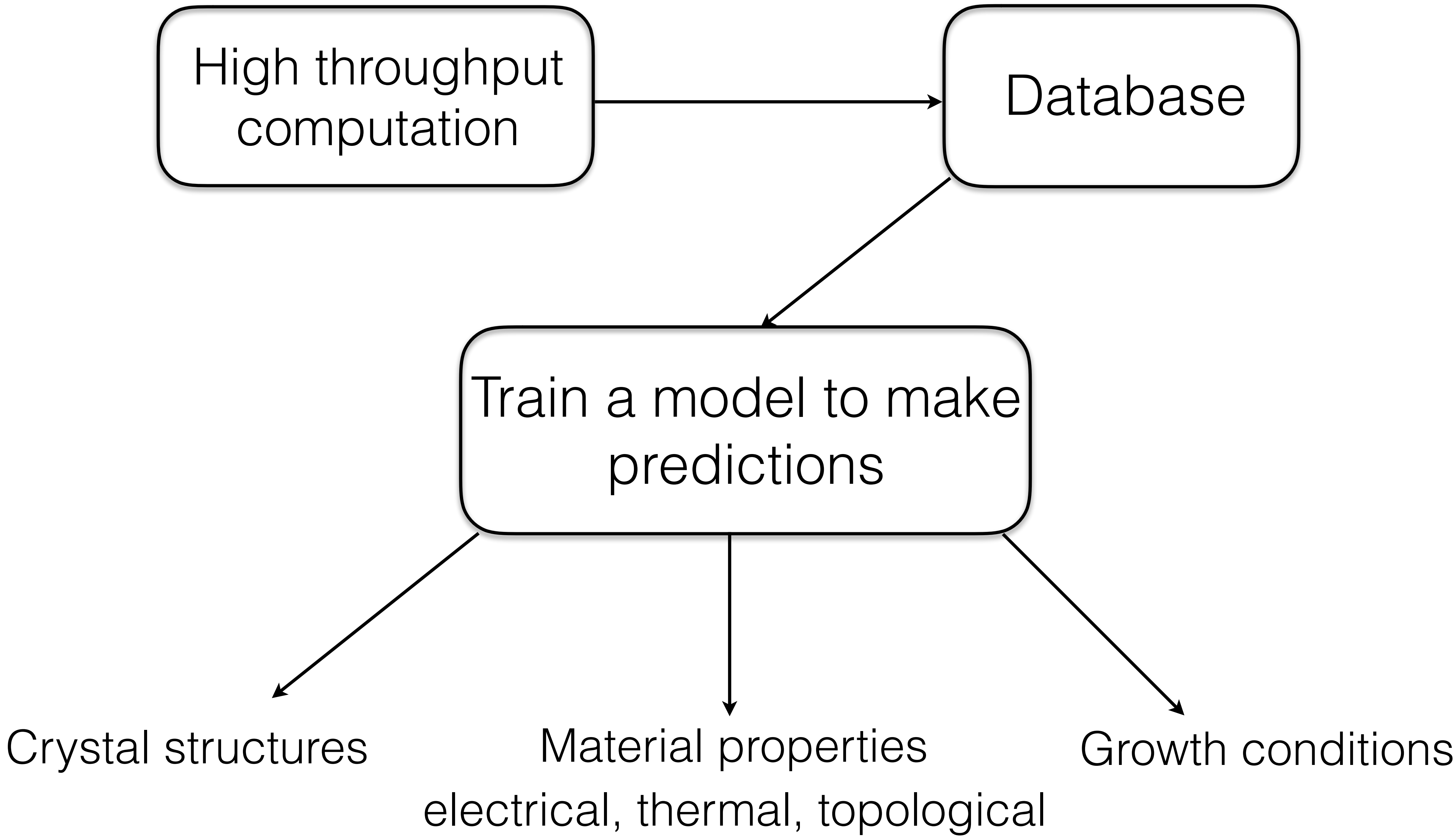
Variational wave functions

Quantum error correction

Classification/discovery phases of matter

Algorithmic development

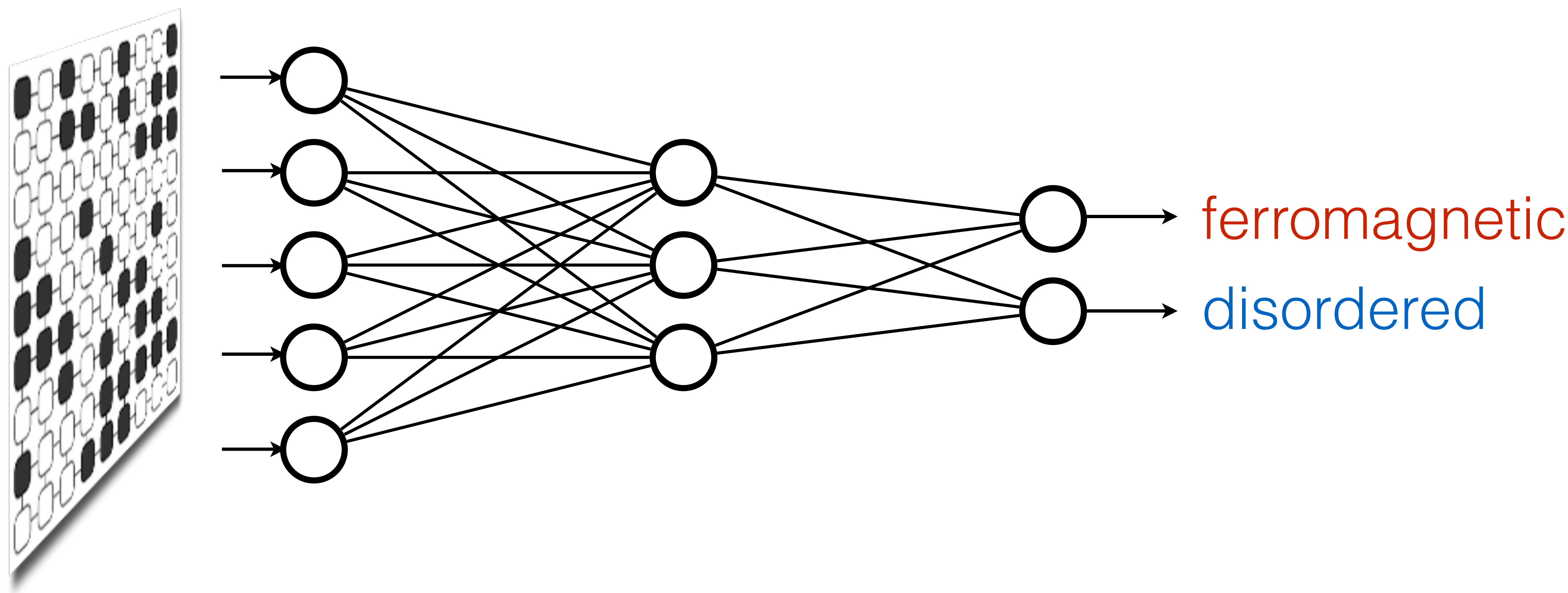
Material Discovery



cf. Hongbin Ren's talk tomorrow

Classify phases of matter

Ising configurations

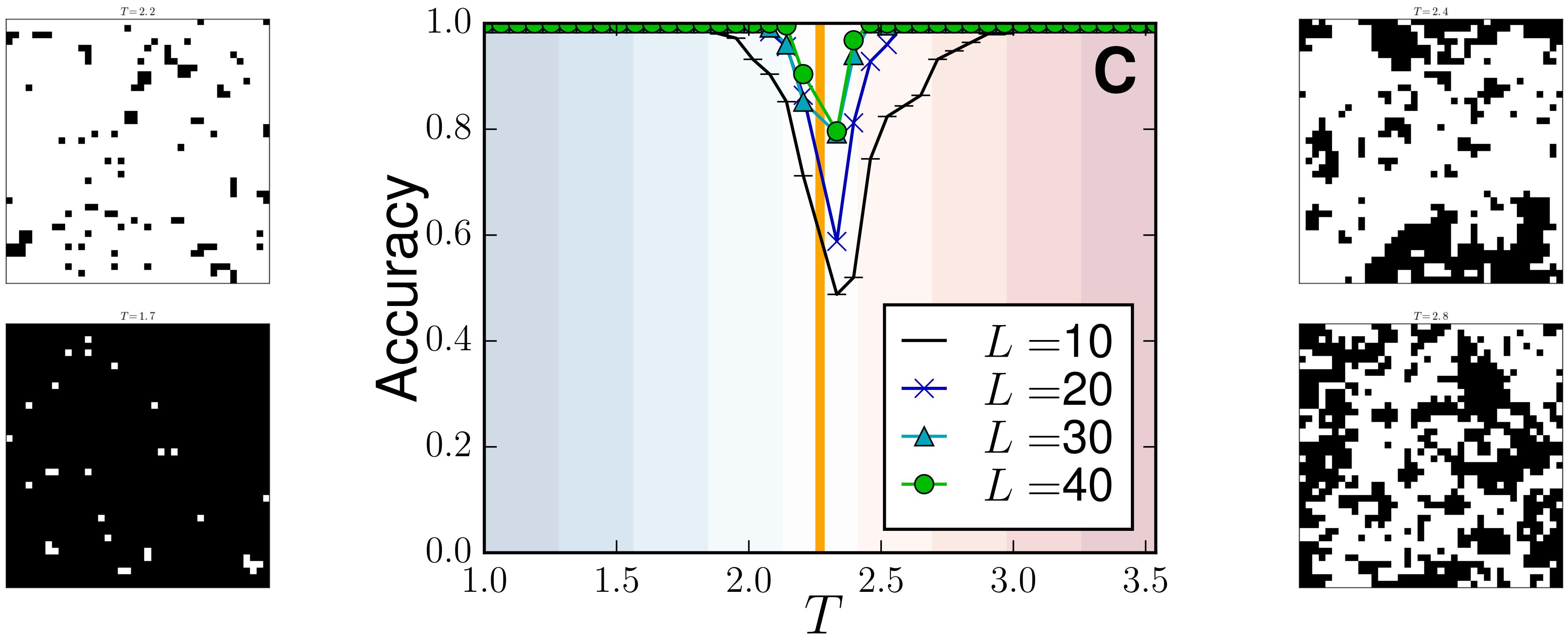


data

label

“Machine Learning Phase of Matter”
Carrasquilla and Melko, 1605.01735

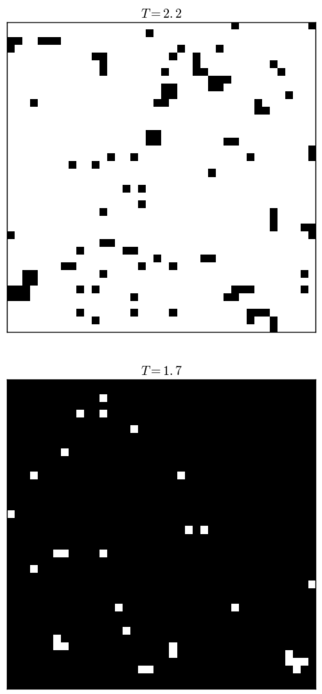
Classify phases of matter

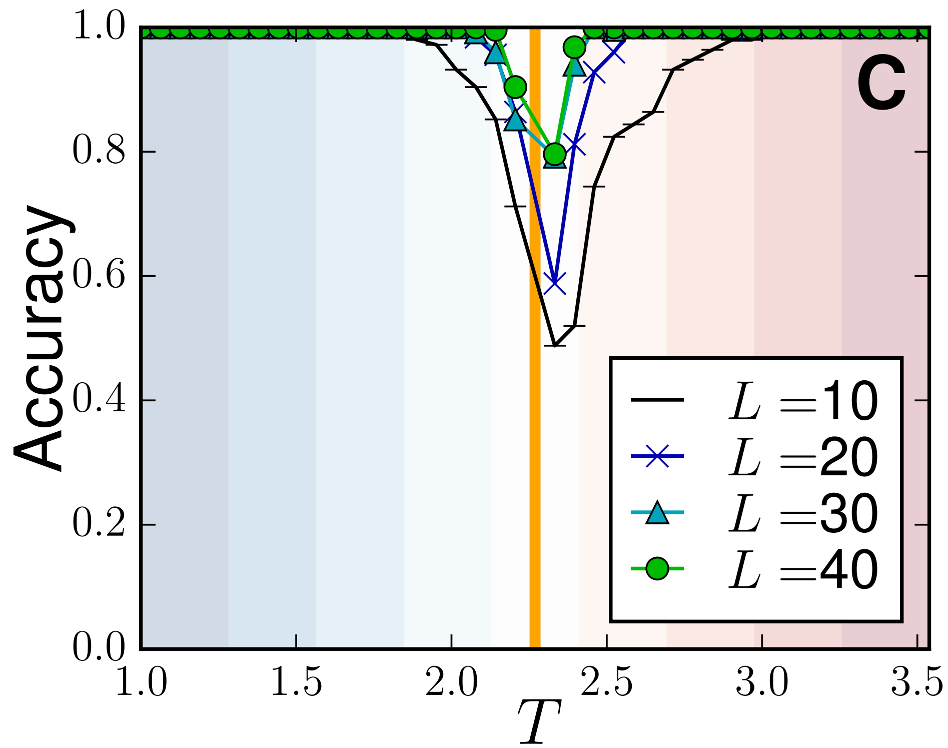


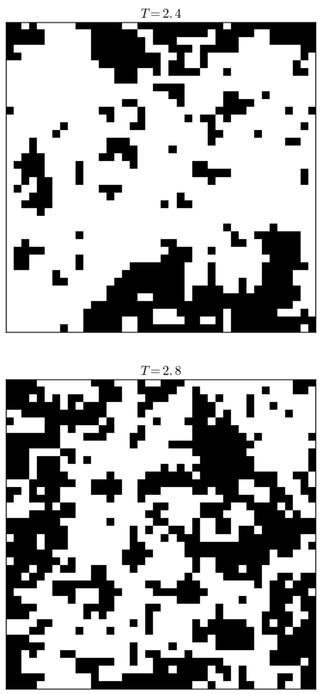
The net computes |total magnetization| for discrimination

"Machine Learning Phase of Matter"
Carrasquilla and Melko, 1605.01735

Classify phases of matter







The net computes |total magnetization| for discrimination

✓ Train on square lattice, predict on triangle lattice

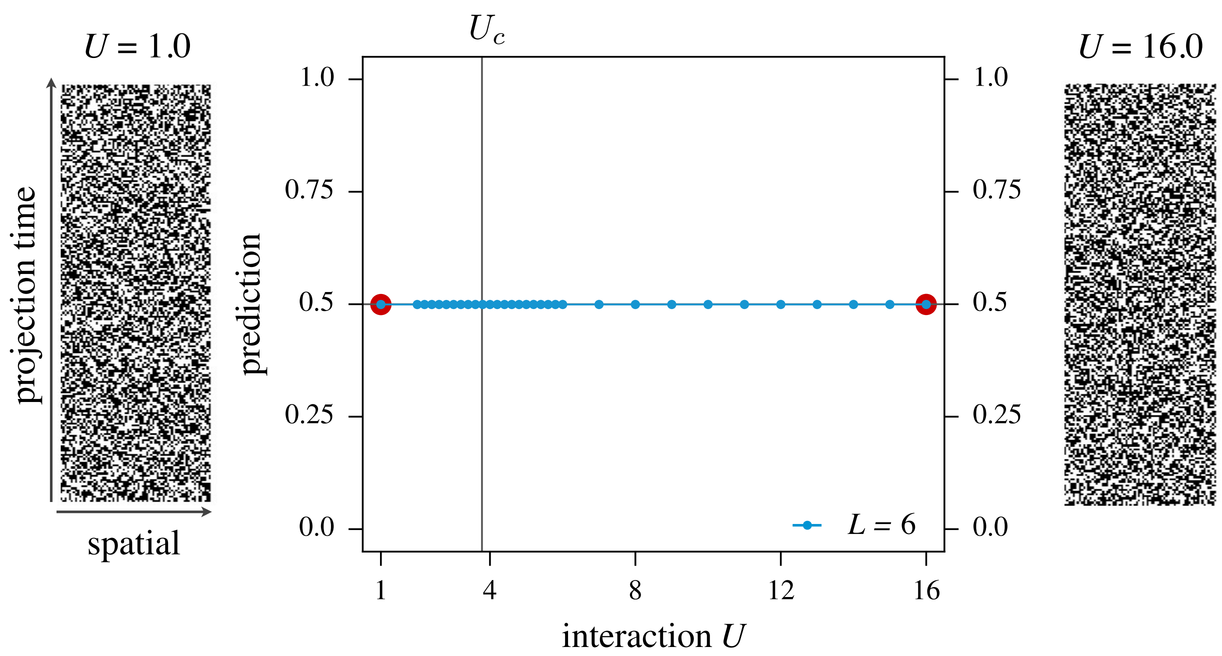
✗ Train for ferromagnets, predict on anti-ferromagnets

✗ Also fails to learn the topology of Ising gauge fields*

"Machine Learning Phase of Matter"
Carrasquilla and Melko, 1605.01735

*The convnet learns about local constraints but not the topological invariance

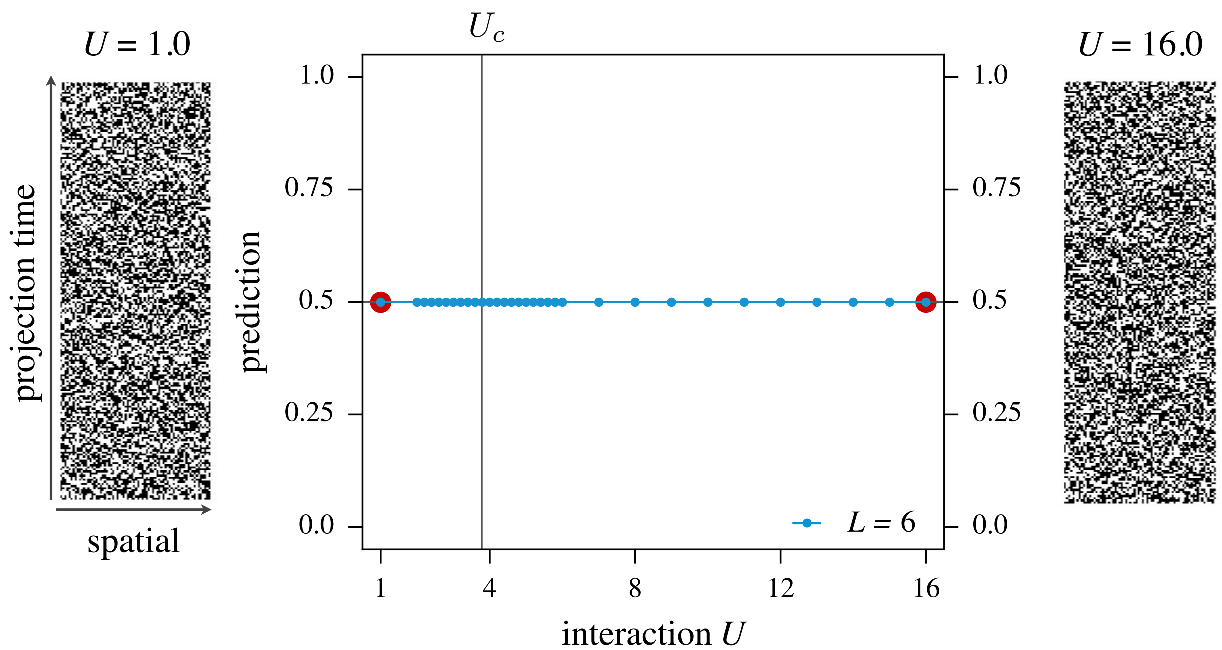
Sign problem in the Hubbard model



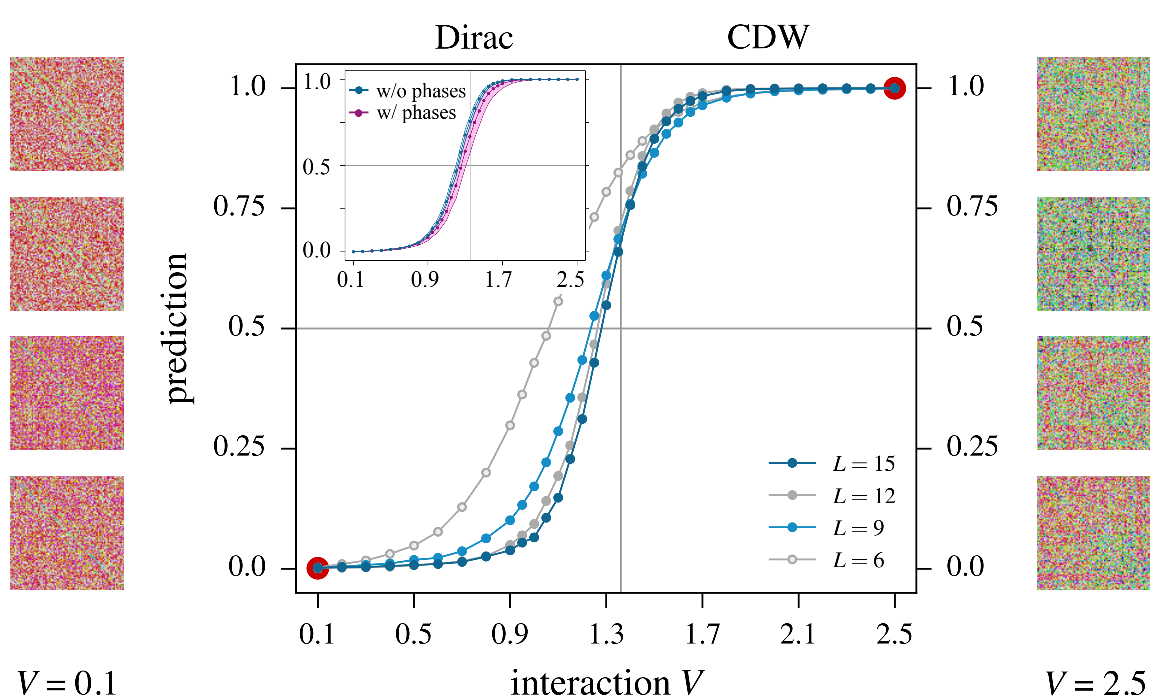
Fails to detect the phase transition in the auxiliary fields

"Machine learning quantum phases of matter beyond the fermion sign problem", Broecker, Carrasquilla, Melko, Trebst, 1608.07848

Sign problem in the Hubbard model



Fails to detect the phase transition in the auxiliary fields



Seems to work with the Green's function

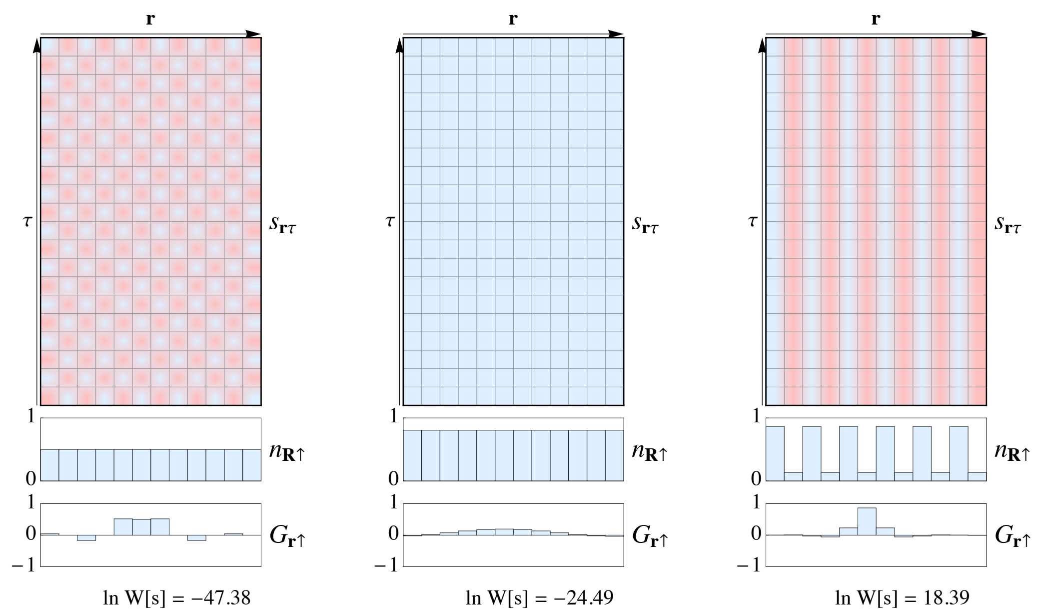
$$\langle c_i c_j^\dagger \rangle_\alpha$$

"Machine learning quantum phases of matter beyond the fermion sign problem", Broecker, Carrasquilla, Melko, Trebst, 1608.07848

Problems with the paper

- **The sign-ignoring model does not have to show phase transition (at the right place) at all**

- Hubbard-Stratonovich transformation in a wrong channel



(a) Fermi-liquid-like configuration (b) Ferromagnetic (c) Antiferromagnetic

Duchon, Loh, Trivedi, 1311.0543

Hirsch PRB, 28, 4059 (1983)

$$\langle S_i(\tau)S_j(0)\rangle = (1-e^{-\Delta\tau U})^{-1}\langle \sigma_i(\tau)\sigma_j(0)\rangle$$

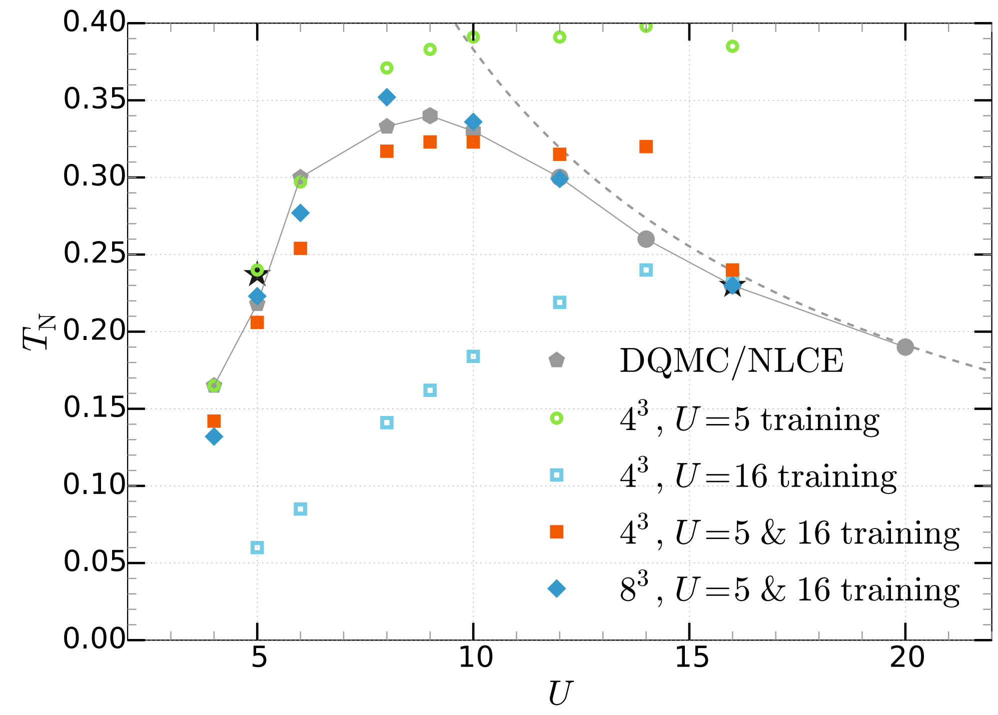
- Isn't it feature engineering to use the Green's function ?

- Quantum problem is in 2+1 d spacetime: a 3d CNN is more appropriate

3D Hubbard model

Can recognize AF phase transition from the auxiliary field configurations

However, has to train both for U=5 and U=16



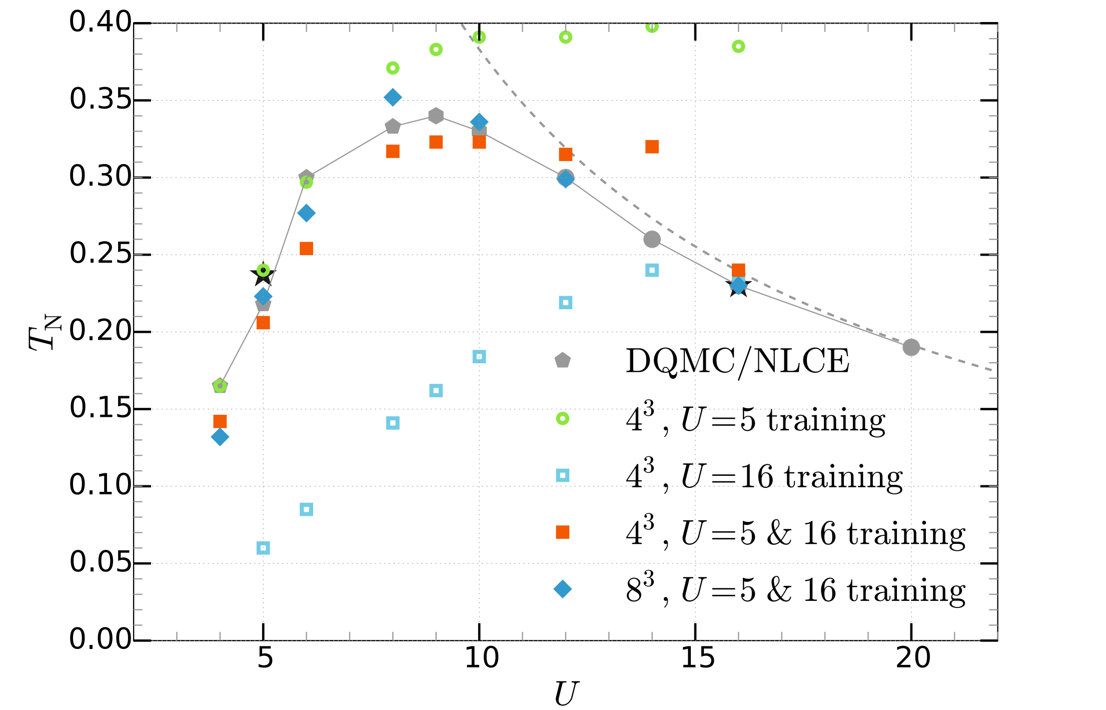
"Machine Learning Phases of Strongly Correlated Fermions"

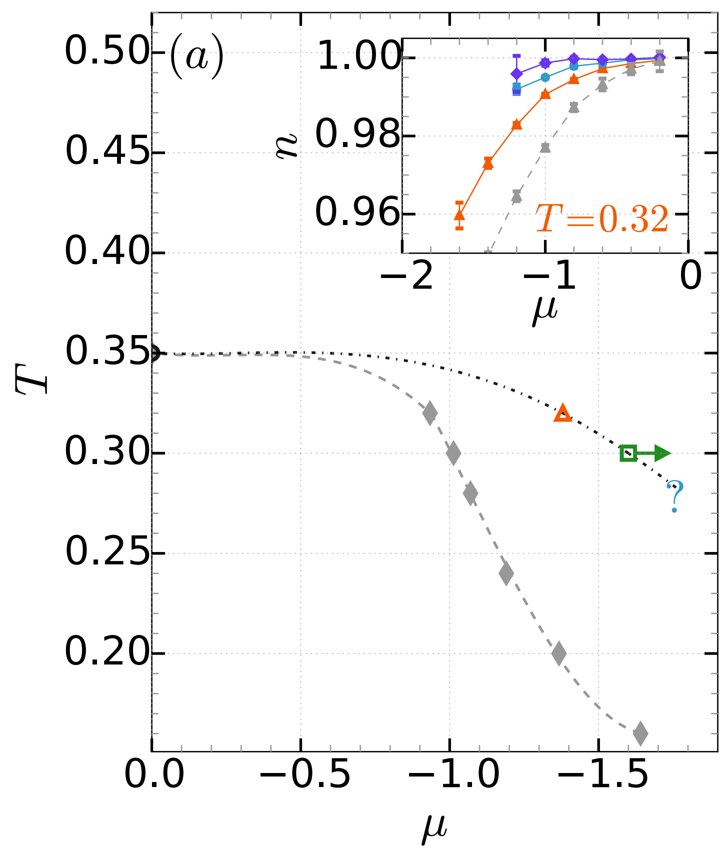
Ch'ng, Carrasquilla, Melko, Khatami, 1609.02552

3D Hubbard model

Can recognize AF phase transition from the auxiliary field configurations

However, has to train both for U=5 and U=16





Transfer learning: train for the half-filled Hubbard model, make predictions for the doped cases

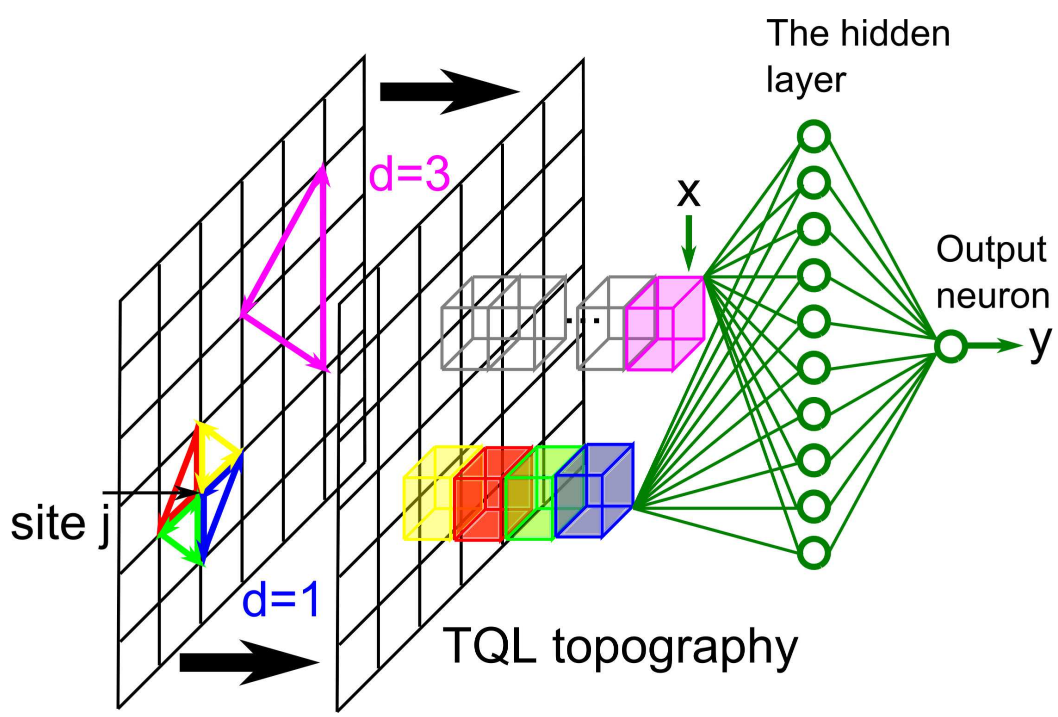
However, the predictions are ambiguous

Also, no theoretical justifications of why it should work

“Machine Learning Phases of Strongly Correlated Fermions”

Ch'ng, Carrasquilla, Melko, Khatami, 1609.02552

Topological (Chern) insulators



Predict topological character from the variational Monte Carlo snapshot (requires hand-crafted feature)

$$\tilde{P}_{jk}\tilde{P}_{kl}\tilde{P}_{lj} \qquad \text{where } \tilde{P}_{jk} \equiv \left\langle c_j^\dagger c_k \right\rangle_\alpha$$

It works because Kitaev 2006

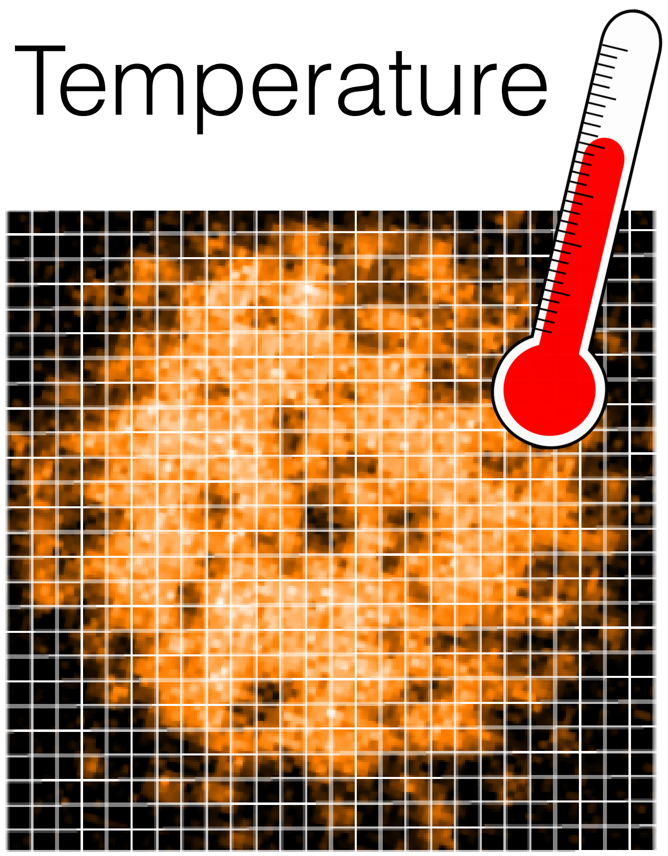
$$\sigma_{xy} = \frac{e^2}{h} \cdot \frac{1}{N} \sum 4\pi i P_{jk} P_{kl} P_{lj} S_{\triangle jkl}$$

"Triangular Quantum Loop Topography for Machine Learning"

Zhang, Kim, 1611.01518

Learn from experimental data

Temperature



“Quantum gas microscope thermometer”

LW, Zi Cai, Unpublished

“Measuring quantum entanglement, machine learning and wave function tomography: Bridging theory and experiment with the quantum gas microscope”, Tubman,1609.08142

Reflections

- How useful is it in the end ?
  - Needs labels for supervised learning
  - Hand-crafted features are superfluous
- The way out: unsupervised feature learning

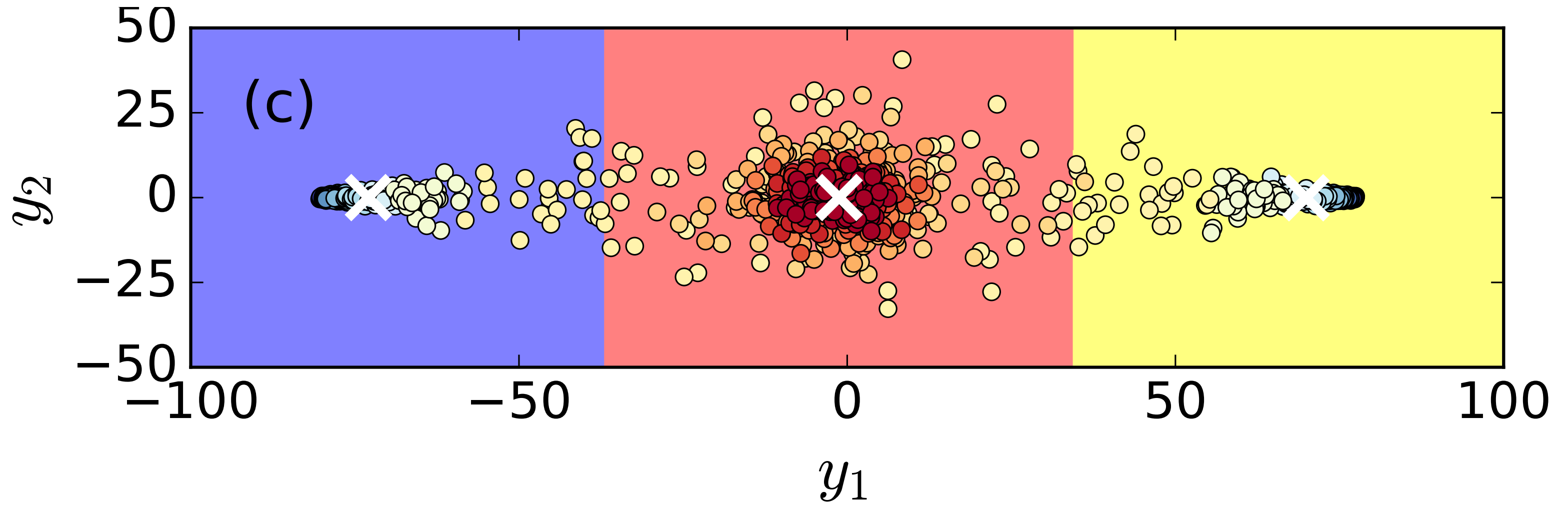
Reflections

- How useful is it in the end ?
  - Needs labels for supervised learning
  - Hand-crafted features are superfluous

- The way out: unsupervised feature learning

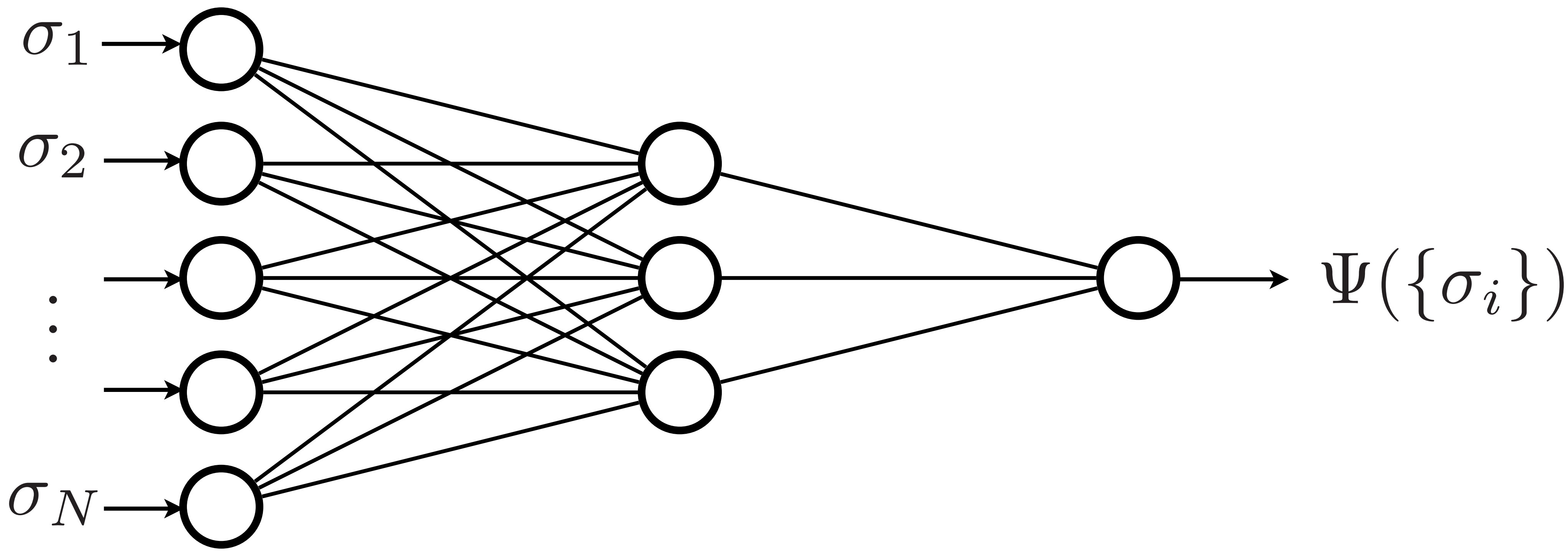
"Discovering Phase Transitions with Unsupervised Learning"

LW, 1606.00318



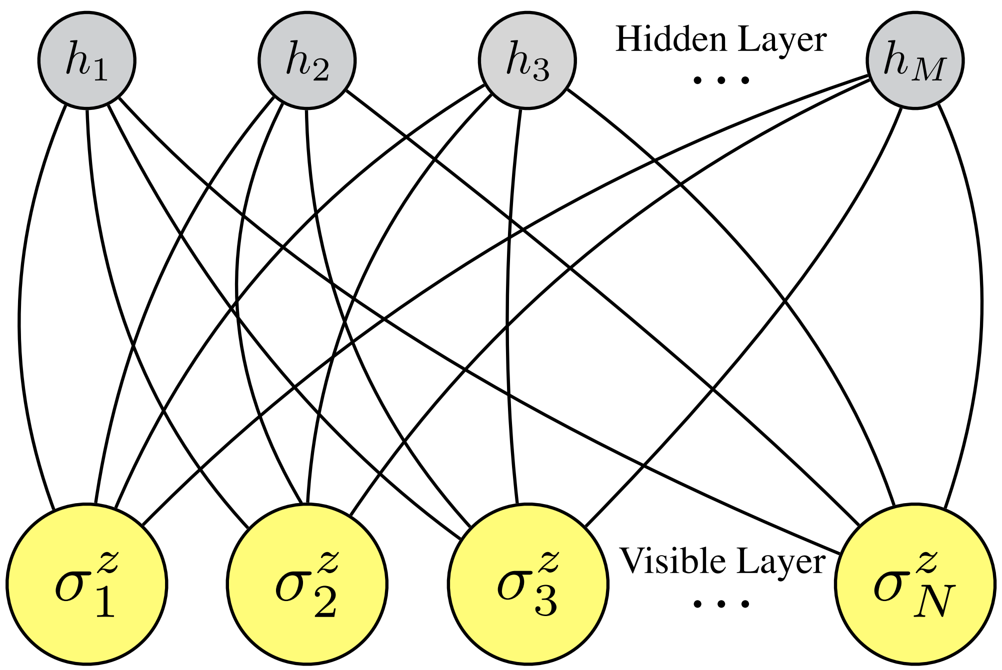
cf Pan Zhang's talk tomorrow about unsupervised learning and clustering

Variational wave functions



- Use neural net as a many-body wave function
- Abstraction power from the deep hierarchical structure
- Hardware (GPU) accelerations
- Software library & framework built by the industry (Tensorflow etc)

RBM variational wave function



$$\Psi_M(\mathcal{S};\mathcal{W}) = \sum_{\{h_i\}} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z}$$

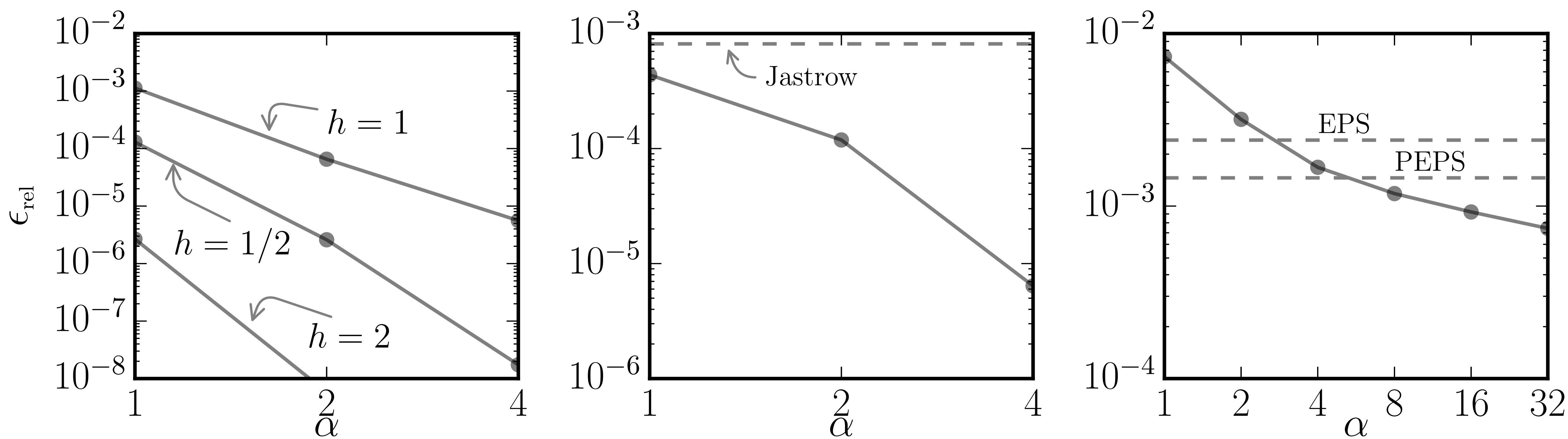
$$= e^{\sum_i a_i \sigma_i^z} \times \Pi_{i=1}^{M} F_i(\mathcal{S})$$

where $\quad F_i(\mathcal{S}) \;=\; 2\cosh\left[b_i + \sum_j W_{ij}\sigma_j^z\right]$

The remaining is standard
variational Monte Carlo calculation

“Solving the Quantum Many-Body Problem with Artificial Neural Networks”, Carleo and Troyer, 1606.02318

Accuracy of the RBM wfs



1d TFIM
L=80, PBC

1d Heisenberg
L=80, PBC

2d Heisenberg
L=10, PBC

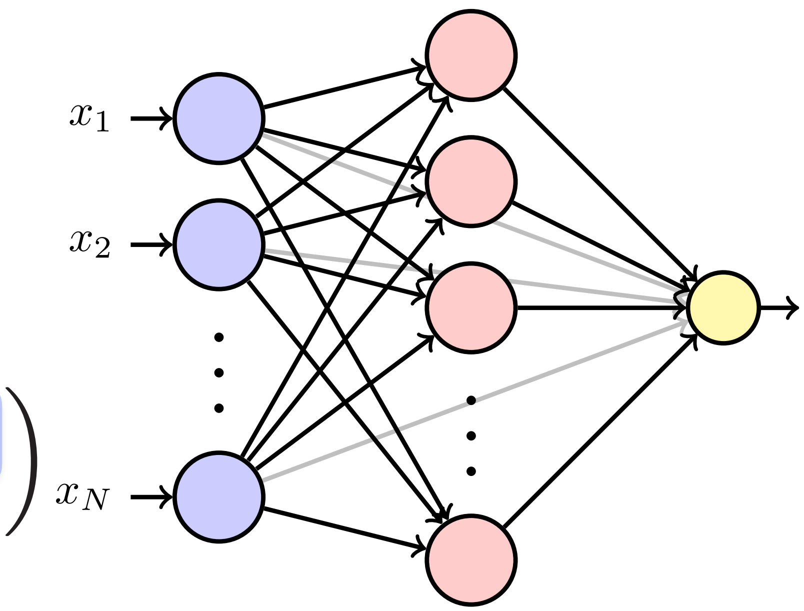
"Solving the Quantum Many-Body Problem with Artificial Neural Networks", Carleo and Troyer, 1606.02318

cf Xiaopeng Li and Zi Cai's talk in the afternoon

RBM as a recommender engine for QMC

Supervised learning of RBM

$$\ln p\left(\mathbf{x}\right) \sim \ln p_{\mathrm{RBM}}\left(\mathbf{x}\right)$$
$$= \sum_{i=1}^{N} a_i x_i + \sum_{j=1}^{M} \ln\left(1 + e^{b_j + \sum_{i=1}^{N} x_i W_{ij}}\right)$$



Generative sampling of RBM

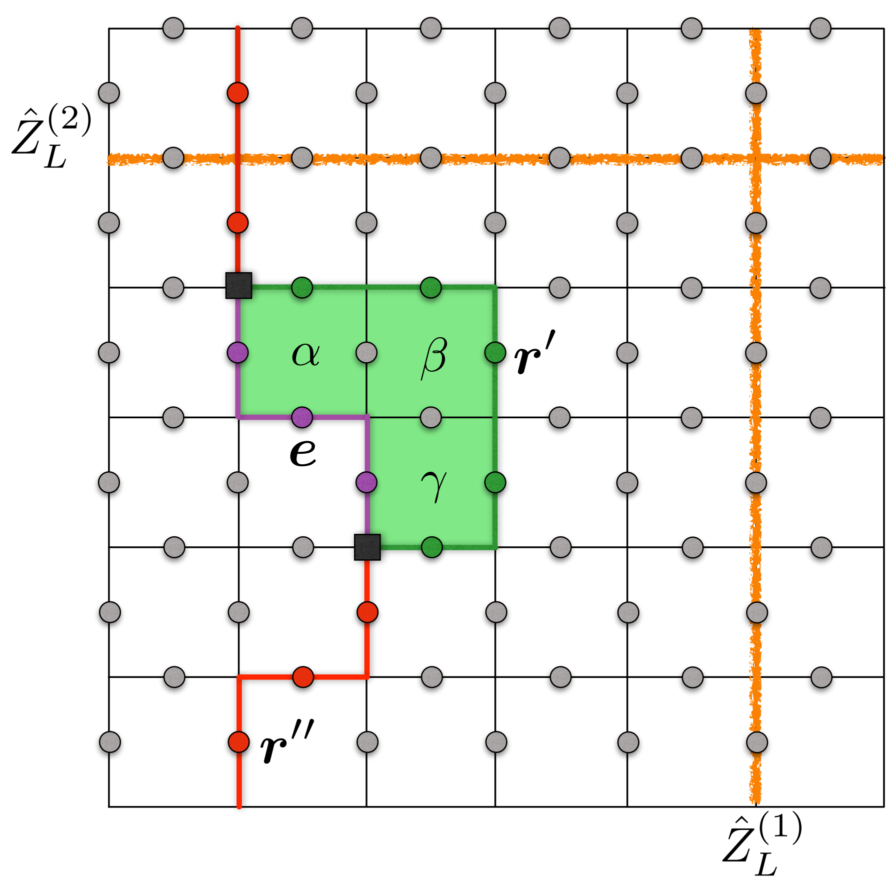
$$A(\mathbf{x} \to \mathbf{x}') = \min\left[1, \frac{p_{\mathrm{RBM}}(\mathbf{x})}{p_{\mathrm{RBM}}(\mathbf{x}')} \cdot \frac{p(\mathbf{x}')}{p(\mathbf{x})}\right]$$

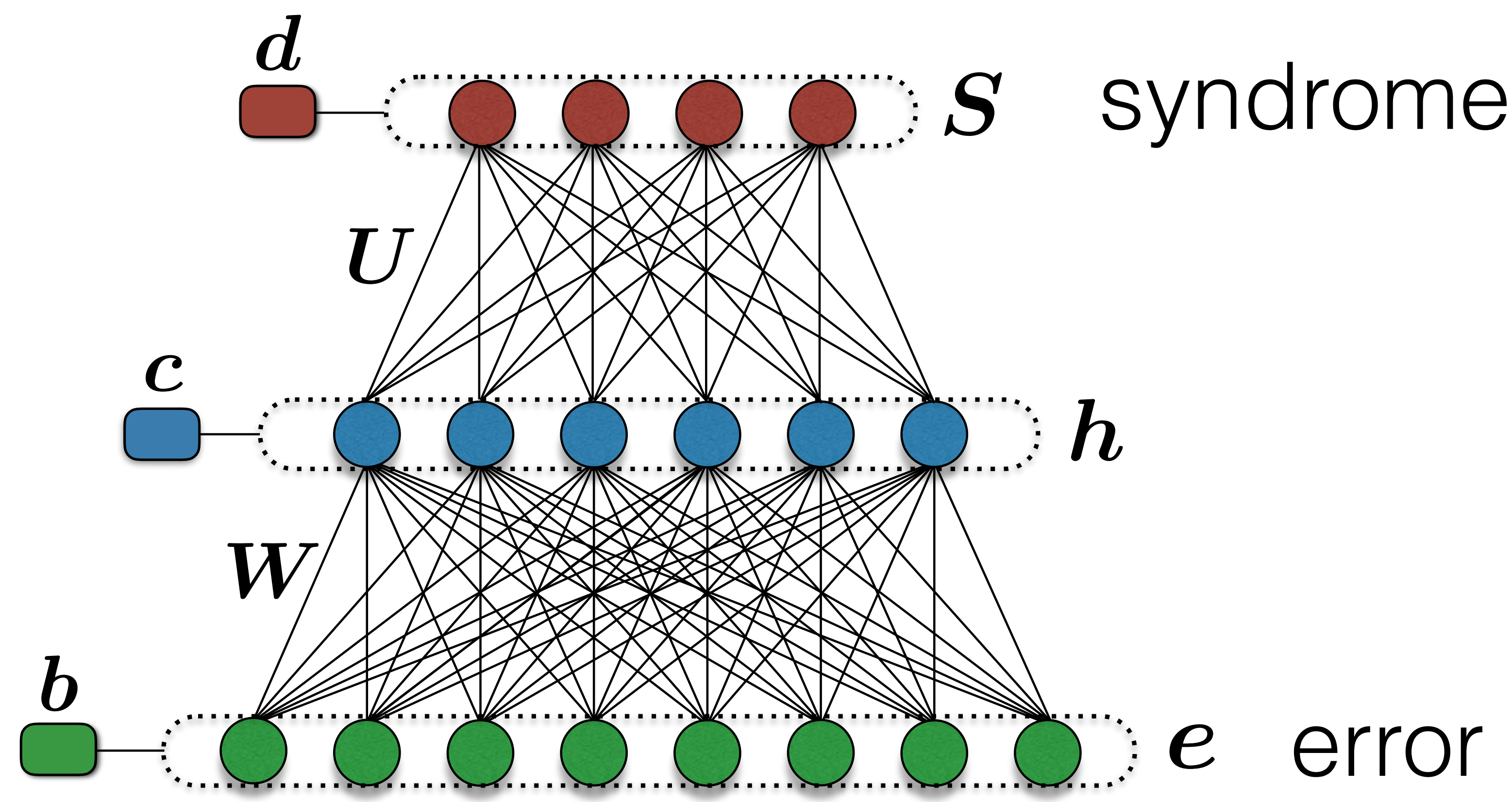
Huang and LW, 1610.02746

Liu, Qi, Meng, Fu, 1610.03137

cf Ziyang Meng and Xiaoyan Xu's talks tomorrow

Quantum Error Correction



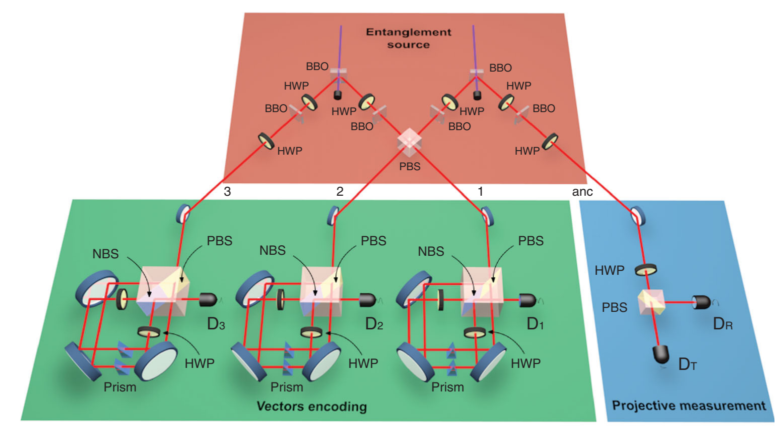


Approximate the probability p(S|e) with an RBM

"A Neural Decoder for Topological Codes"
Torlai and Melko, 1610.04238

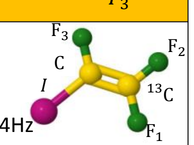
Quantum Machine Learning

- Use a quantum computer to speed up classical ML subroutines
    - Optimization
    - Linear algebra
    - Sampling
    - Clustering
    - Support vector machine
    - Principle component analysis



Cai et al, PRL **114**, 110504 (2015)

	^{13}C	F_1	F_2	F_3
^{13}C	15479.9Hz			
F_1	-297.7Hz	-33130.1Hz		
F_2	-275.7Hz	64.6Hz	-42681.4Hz	
F_3	39.1Hz	51.5Hz	-129.0Hz	-56443.5Hz
T_2^*	1.22s	0.66s	0.63s	0.61s
T_2	7.9s	4.4s	6.8s	4.8s



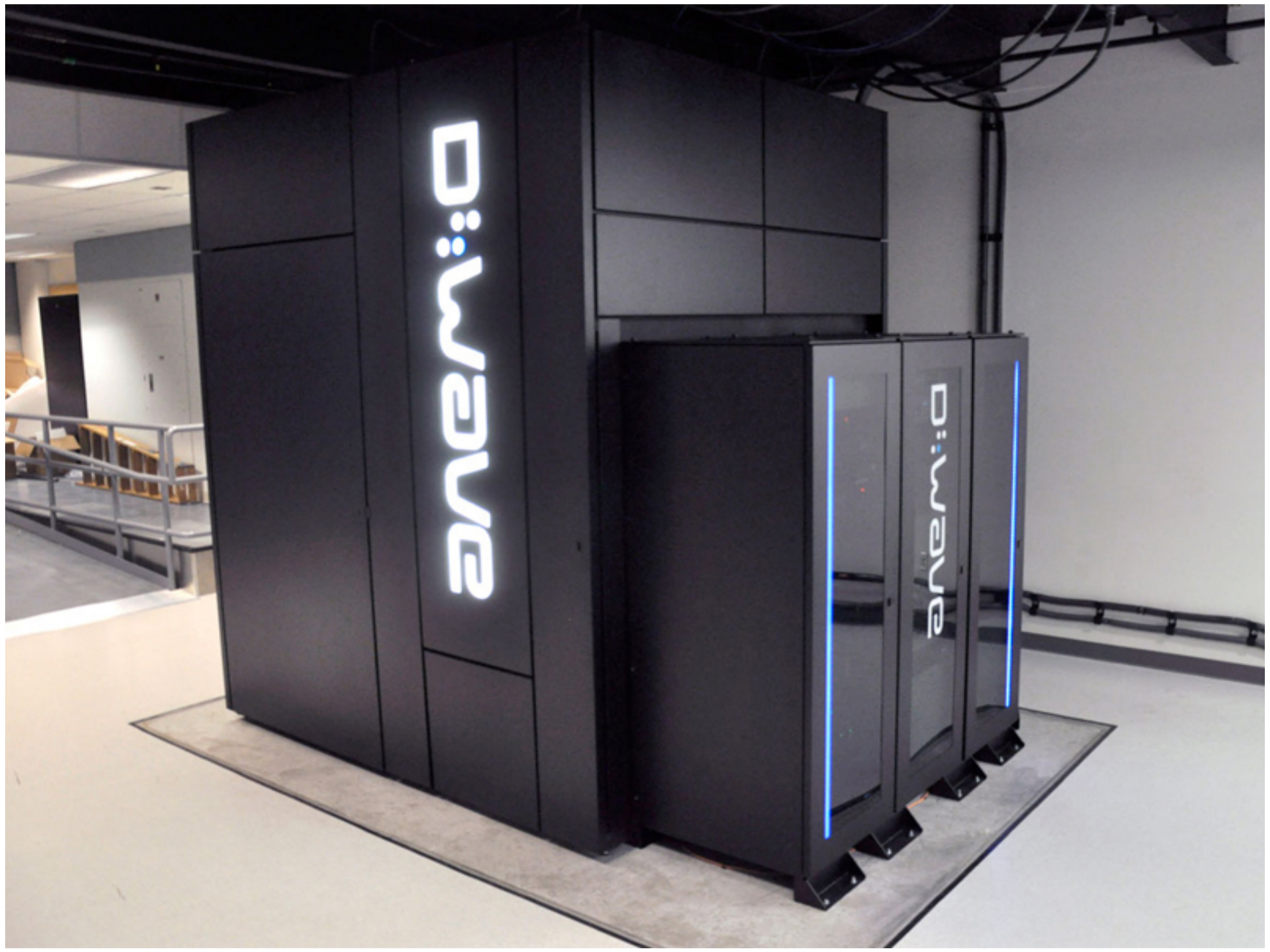
Li et al, PRL **114**, 140504 (2015)

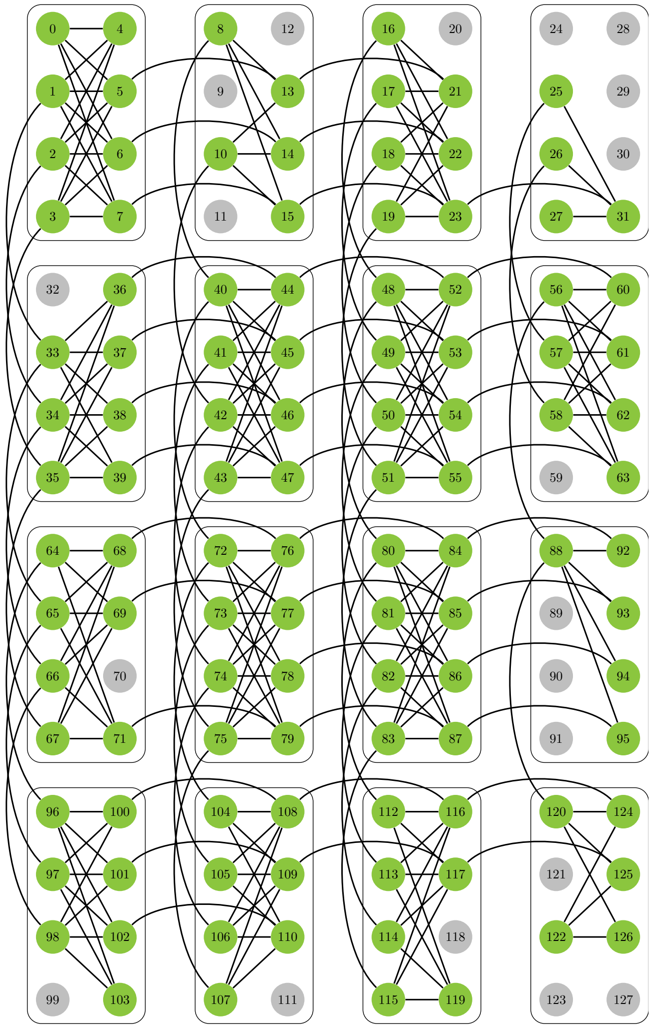
- Quantum data and quantum architecture

“Advances in quantum machine learning”, Adcock et al, 1512.02900
“Quantum machine learning”, Biamonte et al, 1611.09347

Quantum Boltzmann Machine

$15 million "quantum" Ising machine



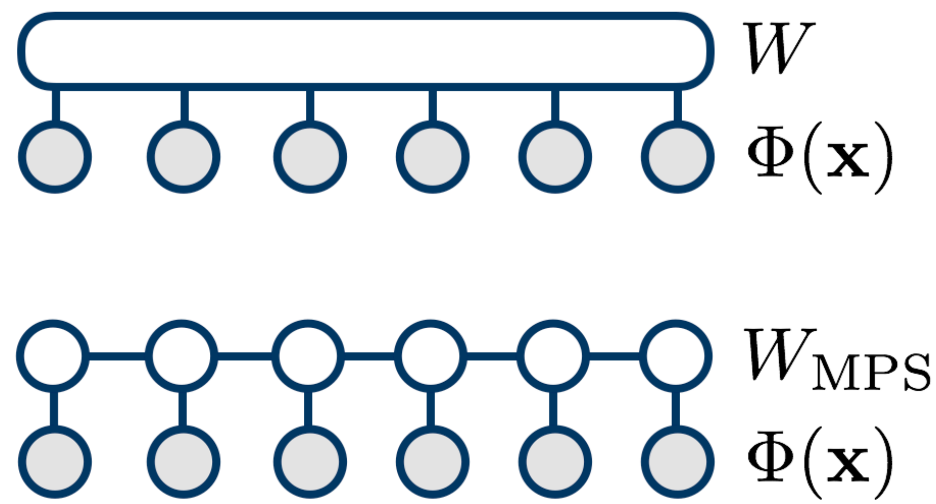


"Quantum Boltzmann Machine"
Amin et al, 1601.02036

Physics -> ML

use MPS for pattern recognition

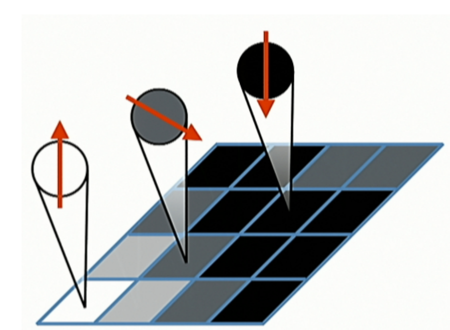
$f(\mathbf{x}) =$



$\approx$

$$f(\mathbf{x}) = \sum_{\{s\}} W_{s_1 s_2 \cdots s_N} \phi^{s_1}(x_1) \otimes \phi^{s_2}(x_2) \otimes \cdots \phi^{s_N}(x_N)$$

$$\phi^{s_j}(x_j) = \left[\cos\left(\frac{\pi}{2} x_j\right),\ \sin\left(\frac{\pi}{2} x_j\right)\right]$$



%99.03 accuracy on MNIST dataset*

* bond dimension 120
images scaled to 14*14

“Supervised Learning With Quantum-Inspired Tensor Networks”, Stoudenmire and Schwab, 1605.05775   cf Novikov et al, 1605.03795