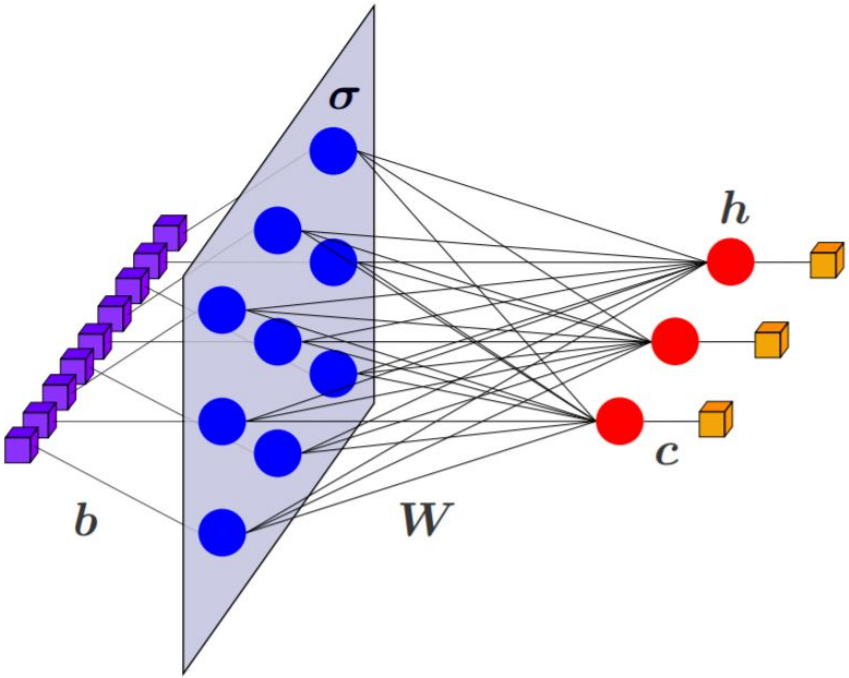
Generative Modelling for Statistical Physics

By Rajat Chandra Mishra
Supervised by Dr Chae-Yeun Park

Contents

Restricted Boltzmann Machines

- Each circle represents a node, which can take one of 2 values (0 or 1). Nodes in the same layer are not connected to each other



Energy and Probability

$$E(\mathbf{x},\mathbf{h}) = -\mathbf{h}^T\mathbf{W}\mathbf{x} - \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{h}$$

$$p(\mathbf{x},\mathbf{h}) = \frac{\exp(-E)}{Z}$$

Conditional Probability

$$p(\mathbf{h}|\mathbf{x}) = \Pi_j p(h_j|\mathbf{x})$$

$$p(h_j = 1|\mathbf{x}) = sigm(b_j + \mathbf{W}_{j,:}\mathbf{x})$$

$$p(\mathbf{x}|\mathbf{h}) = \Pi_k p(x_k|\mathbf{h})$$

$$p(x_k = 1|\mathbf{h}) = sigm(c_k + \mathbf{h}^T\mathbf{W}_{:,k})$$

Free Energy/Effective Visible Energy

$$p(\mathbf{x}) = \Sigma_h \exp(-E(\mathbf{x},\mathbf{h}))/Z$$

$$p(\mathbf{x}) = \exp(\mathbf{c}^T\mathbf{x} + \Sigma_j \log[1 + \exp(b_j + \mathbf{W}_{j,:}\mathbf{x})])/Z$$

$$we\ can\ define\ a\ free\ energy,$$

$$F(\mathbf{x}) = -\mathbf{c}^T\mathbf{x} - \Sigma_j \log(1 + \exp(b_j + \mathbf{W}_{j,:}\mathbf{x}))$$

$$Then,\ p(\mathbf{x}) = \exp(-F(\mathbf{x})/Z$$

$$softplus(x) = \log(1 + \exp(x))$$

$$F(\mathbf{x}) = -\mathbf{c}^T\mathbf{x} - \Sigma_j softplus(b_j + \mathbf{W}_{j,:}\mathbf{x})$$

Rows of j can be understood as ‘features’. If x has the same features, then the probability of x is higher

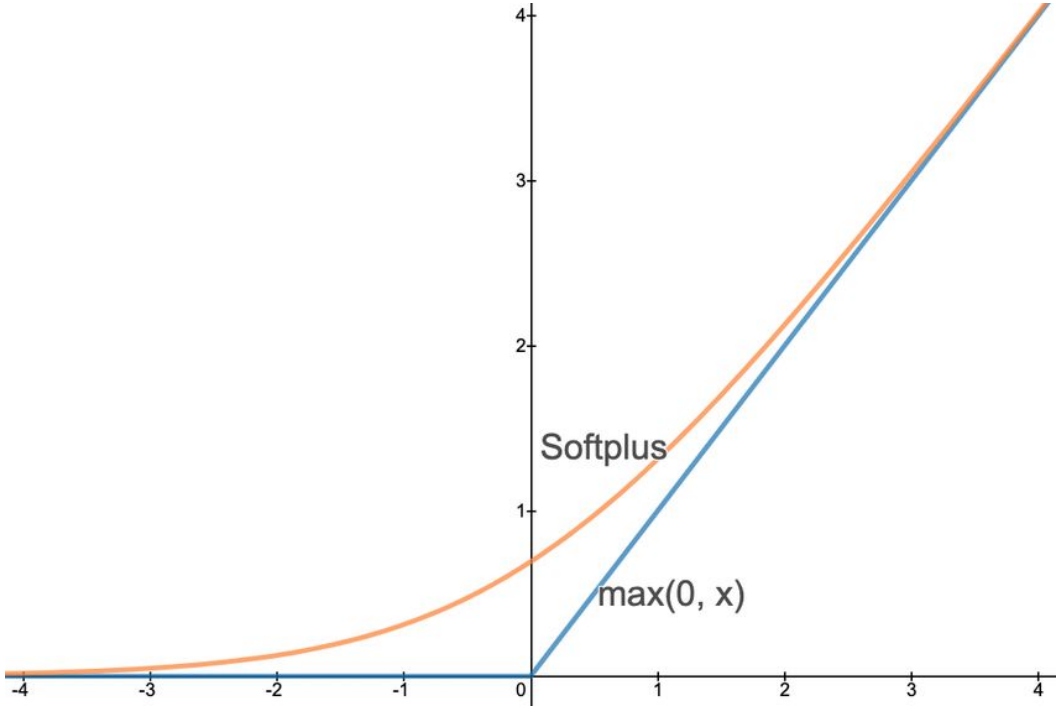
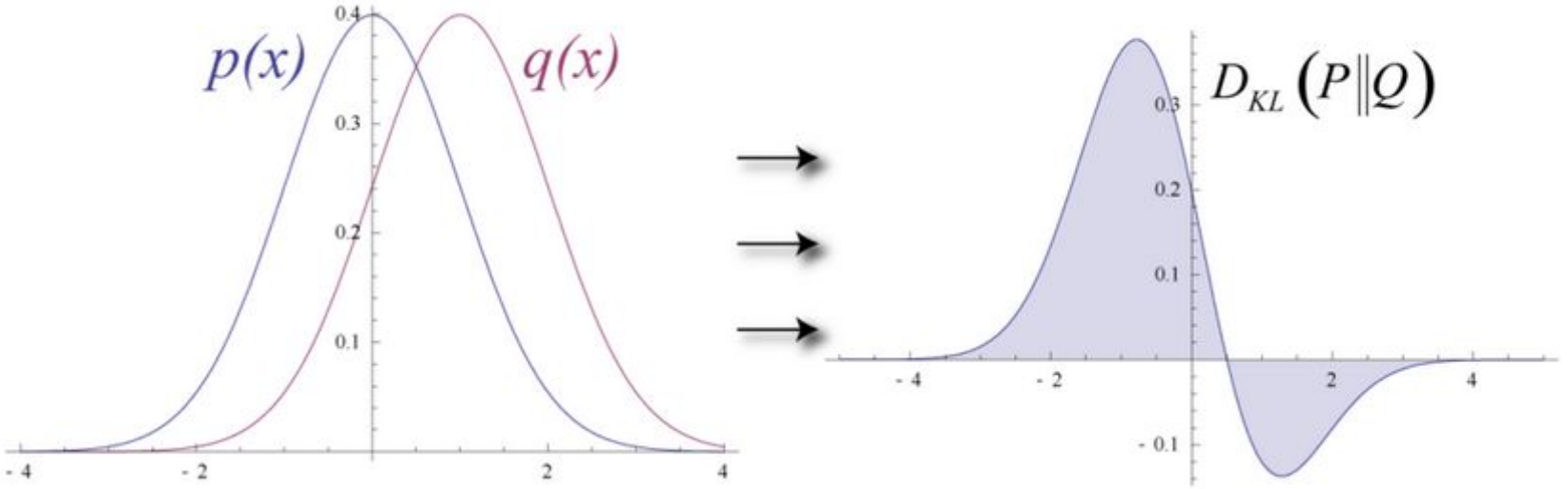


Image source: https://www.researchgate.net/figure/The-Softplus-function-ln1-exp-compared-to-max0_fig2_336602359 [accessed 30 May, 2021]

Loss Function (Kullbach-Liebler Divergence)

Measures the non-overlapping, or diverging, areas under the two curves

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

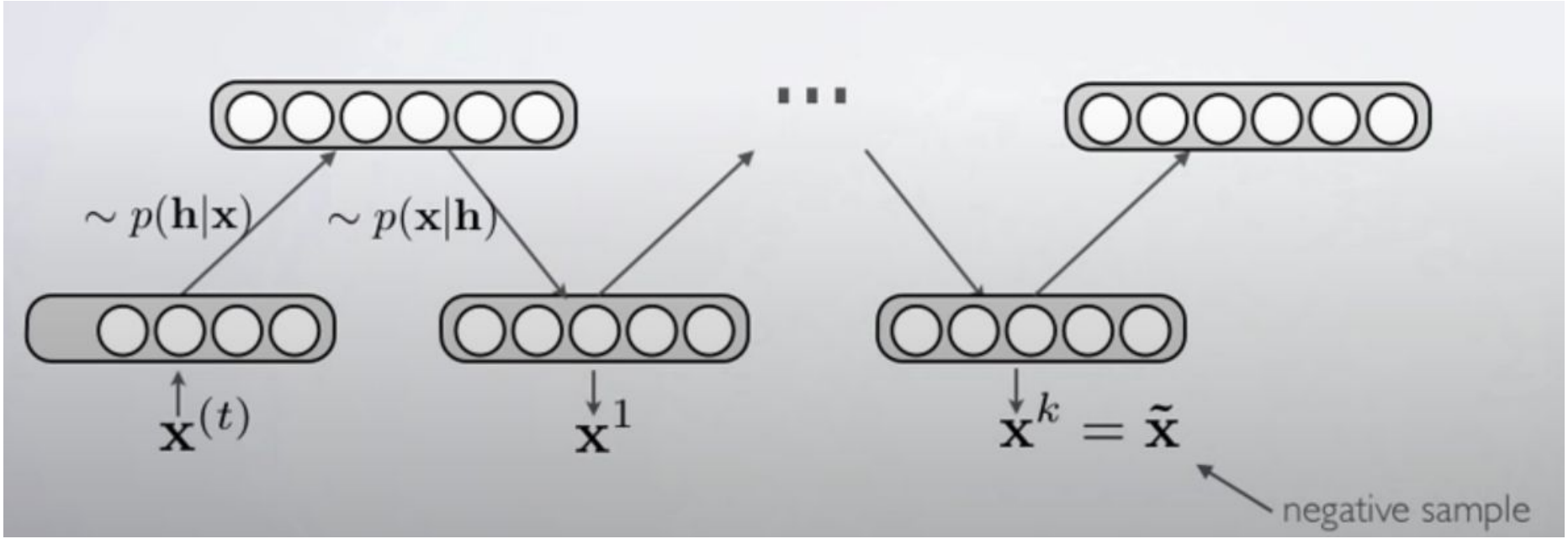


Contrastive Divergence

$$-\frac{\partial \log[p(\mathbf{x}^t)]}{\partial \theta} = \langle \partial_\theta E(\mathbf{x},\mathbf{h}) | \mathbf{x}^t \rangle_{\mathbf{h}} - \langle \partial_\theta E(\mathbf{x},\mathbf{h}) \rangle_{\mathbf{x},\mathbf{h}}$$

The first term is called the positive phase, and the second is called the negative phase. The second term is hard to compute. We use Contrastive Divergence (Hinton 2002) to evaluate the second term. We replace the average by a point estimate ($\tilde{x}$). This point estimate is obtained by gibbs sampling

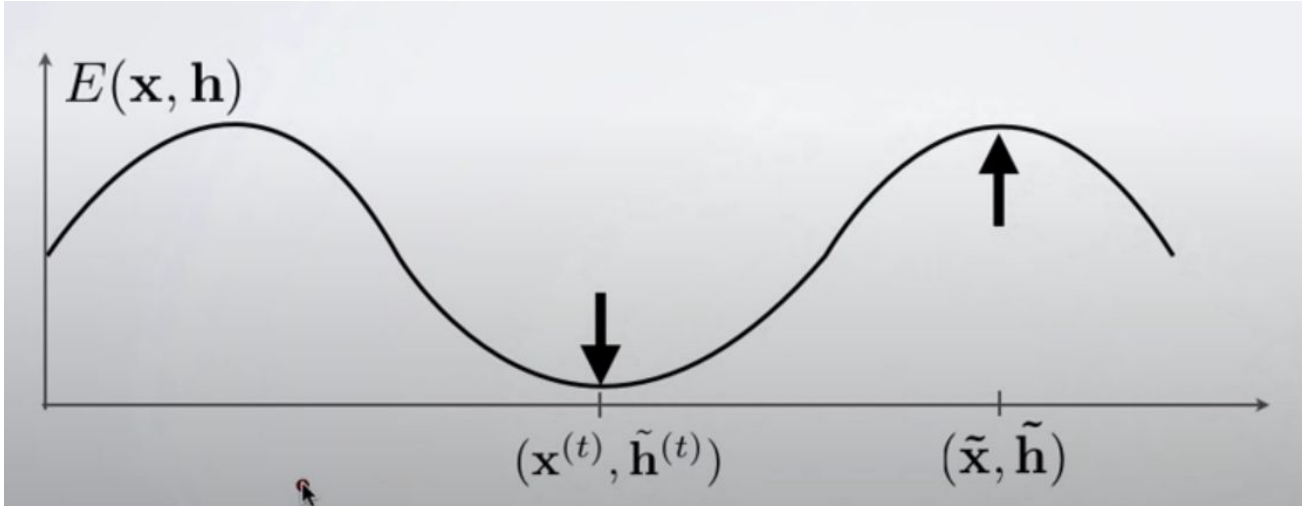
Gibbs Sampling



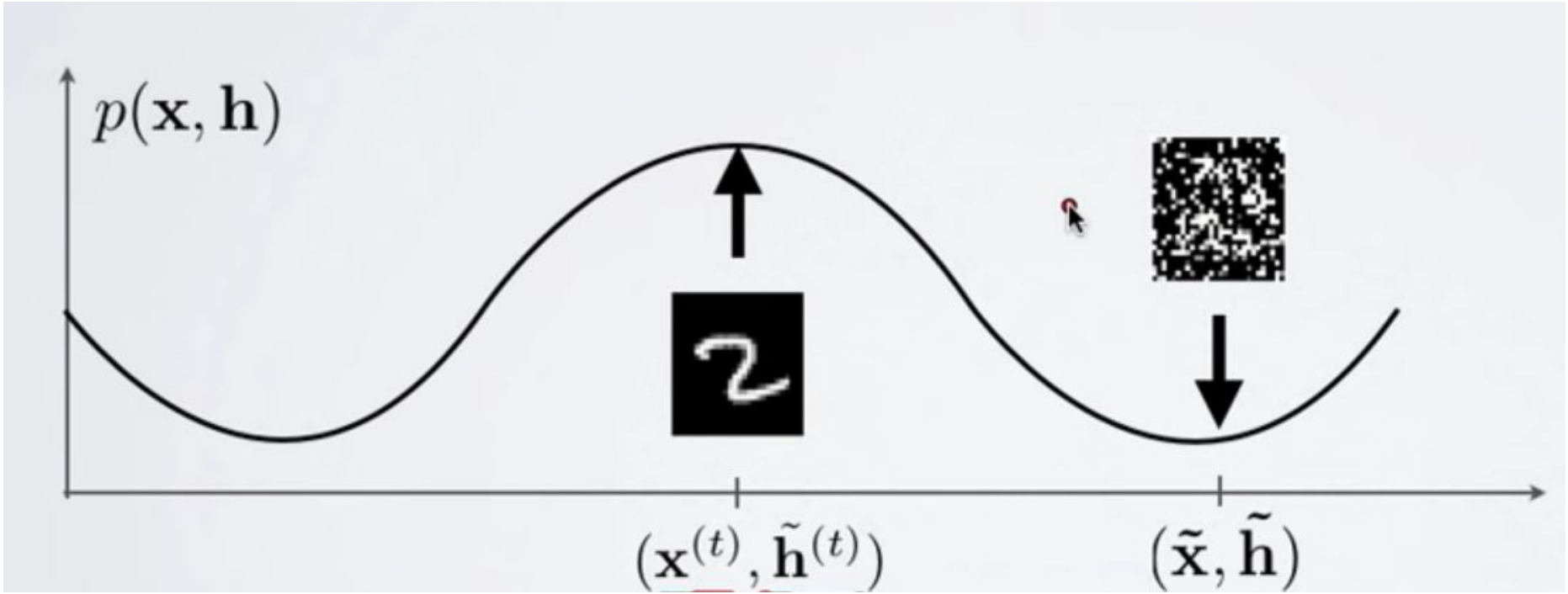
Contrastive Divergence

$$\langle \partial_\theta E(\mathbf{x},\mathbf{h}) | \mathbf{x}^t \rangle_{\mathbf{h}} \approx \partial_\theta E(\mathbf{x}^t, \mathbf{h}^t) \; where \; \mathbf{h}^t = p(\mathbf{h} = \mathbf{1} | \mathbf{x}^t)$$

$$\langle \partial_\theta E(\mathbf{x},\mathbf{h}) \rangle_{\mathbf{x},\mathbf{h}} \approx \partial_\theta E(\tilde{\mathbf{x}}, \tilde{\mathbf{h}}) \; where \; \tilde{\mathbf{h}} = p(\mathbf{h} = \mathbf{1} | \tilde{\mathbf{x}})$$



Contrastive Divergence



Parameter Update

$$\mathbf{W} += \alpha(\mathbf{h}^t \mathbf{x}^{t^T} - \tilde{\mathbf{h}}\tilde{\mathbf{x}}^T)$$

$$\mathbf{b} += \alpha(\mathbf{h}^t - \tilde{\mathbf{h}})$$

$$\mathbf{c} += \alpha(\mathbf{x}^t - \tilde{\mathbf{x}})$$

Persistent Contrastive Divergence

Initialise the Gibbs Chain using negative sample from previous iteration instead of training example from current iteration

Learning Thermodynamics with Boltzmann Machines

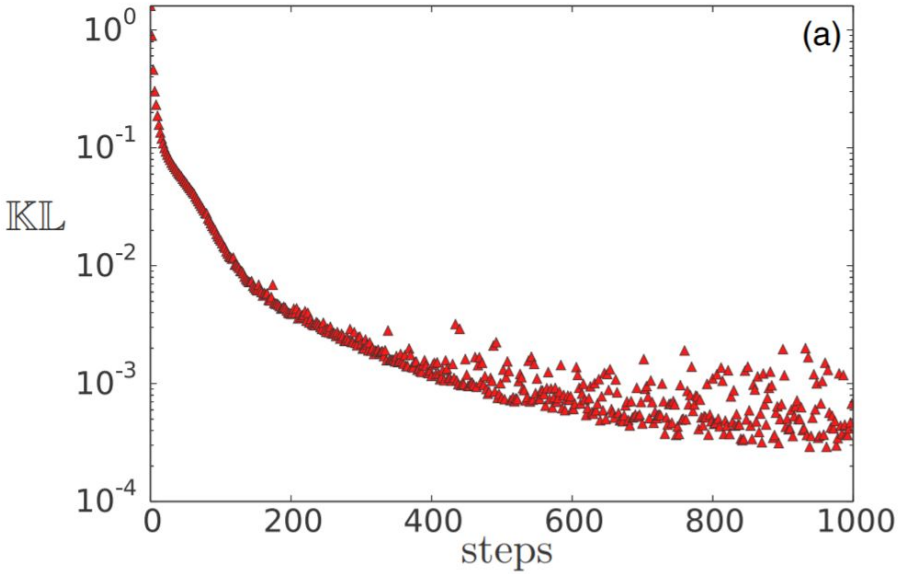
PHYSICAL REVIEW B **94**, 165134 (2016)

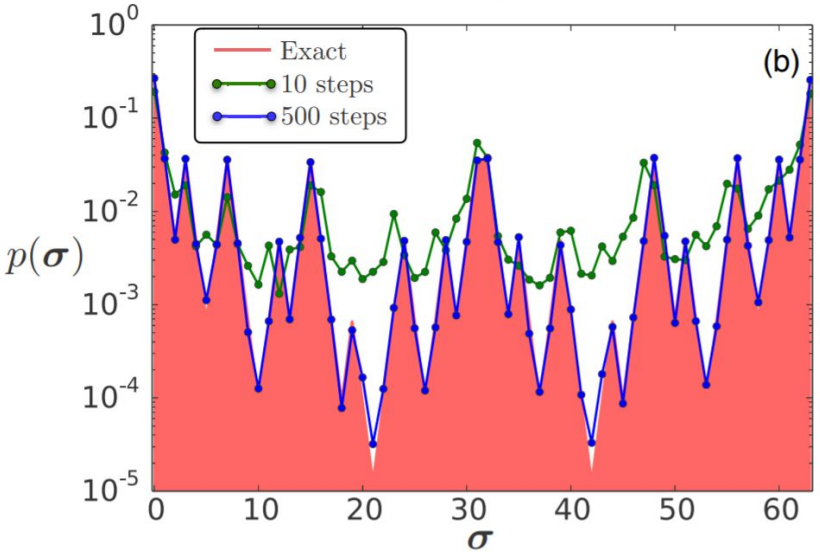
Learning thermodynamics with Boltzmann machines

Giacomo Torlai and Roger G. Melko
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5
and Department of Physics and Astronomy, University of Waterloo, Ontario, Canada N2L 3G1
(Received 17 June 2016; revised manuscript received 8 September 2016; published 17 October 2016)

A Boltzmann machine is a stochastic neural network that has been extensively used in the layers of deep architectures for modern machine learning applications. In this paper, we develop a Boltzmann machine that is capable of modeling thermodynamic observables for physical systems in thermal equilibrium. Through unsupervised learning, we train the Boltzmann machine on data sets constructed with spin configurations importance sampled from the partition function of an Ising Hamiltonian at different temperatures using Monte Carlo (MC) methods. The trained Boltzmann machine is then used to generate spin states, for which we compare thermodynamic observables to those computed by direct MC sampling. We demonstrate that the Boltzmann machine can faithfully reproduce the observables of the physical system. Further, we observe that the number of neurons required to obtain accurate results increases as the system is brought close to criticality.

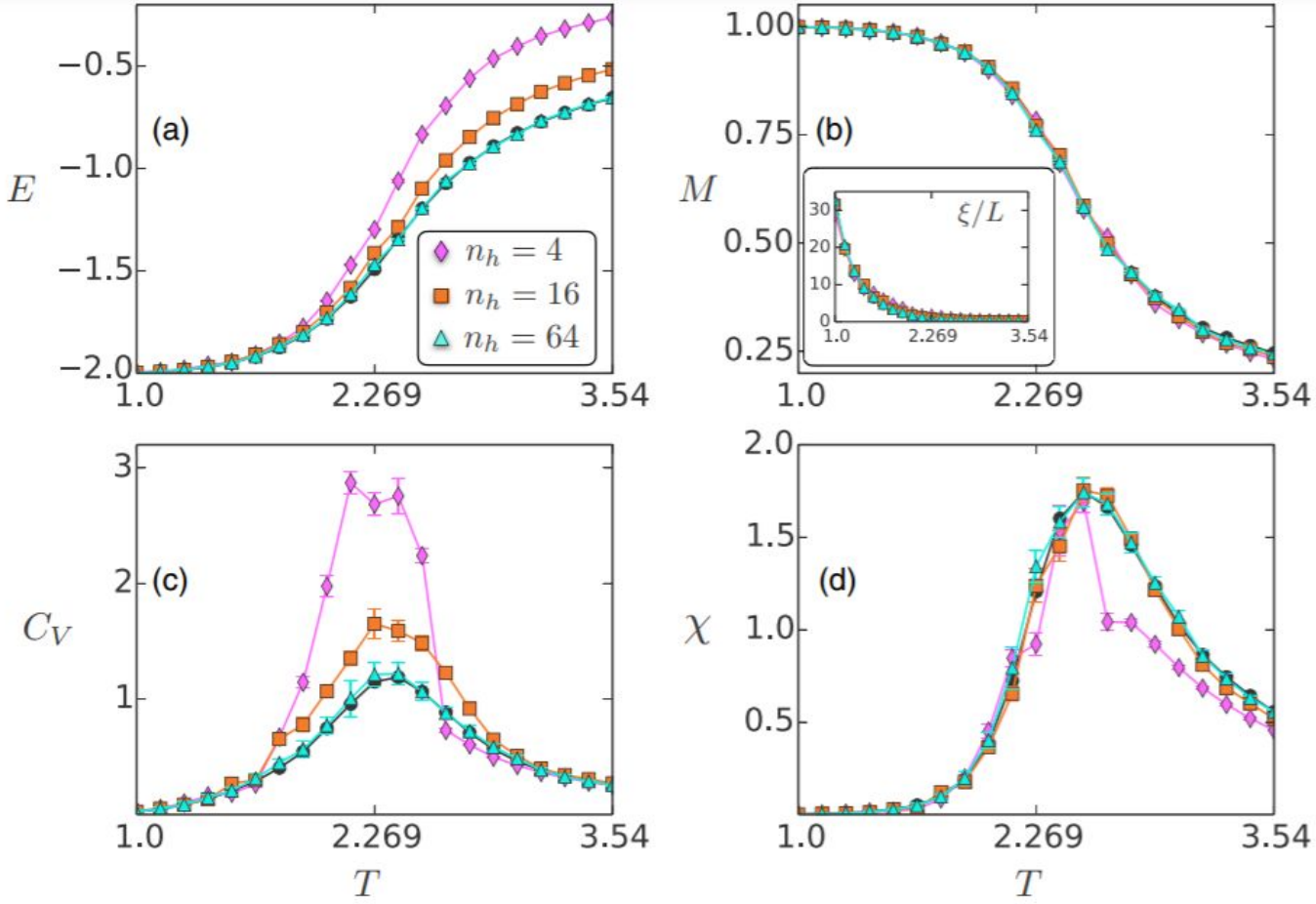
DOI: 10.1103/PhysRevB.94.165134



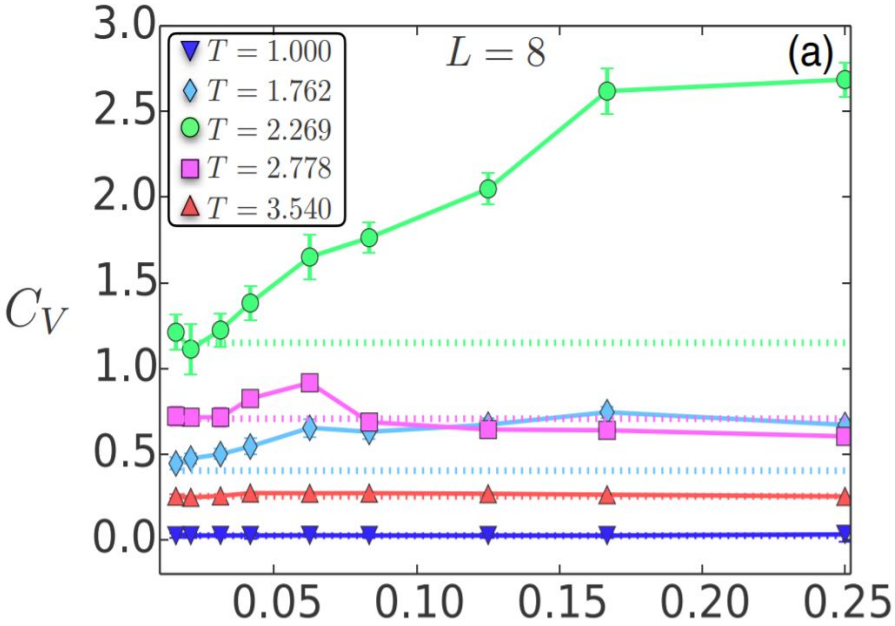


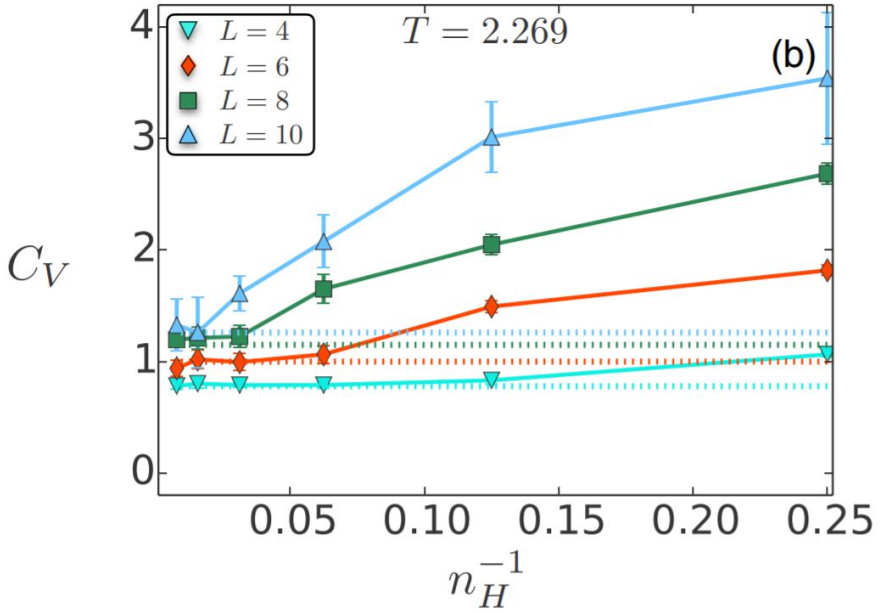
(a) KL divergence vs number of training steps
(b) Comparision of probability distribution obtained from RBM with exact results

1D Ising model with N=6



2D Ising Model on square lattice with periodic boundaries (N=64)





Scaling of the specific heat with the number of hidden nodes. In (a) scaling at different temperatures T with fixed system size. In (b) we see the scaling at criticality for different system sizes L. Dotted lines represent the exact value computed on the spin configurations of the training data set. The number of hidden nodes for faithful generation increases with system size, and for a fixed system size, it is large near the critical region

Solving Statistical Mechanics Using Variational Autoregressive Networks

Solving Statistical Mechanics Using Variational Autoregressive Networks

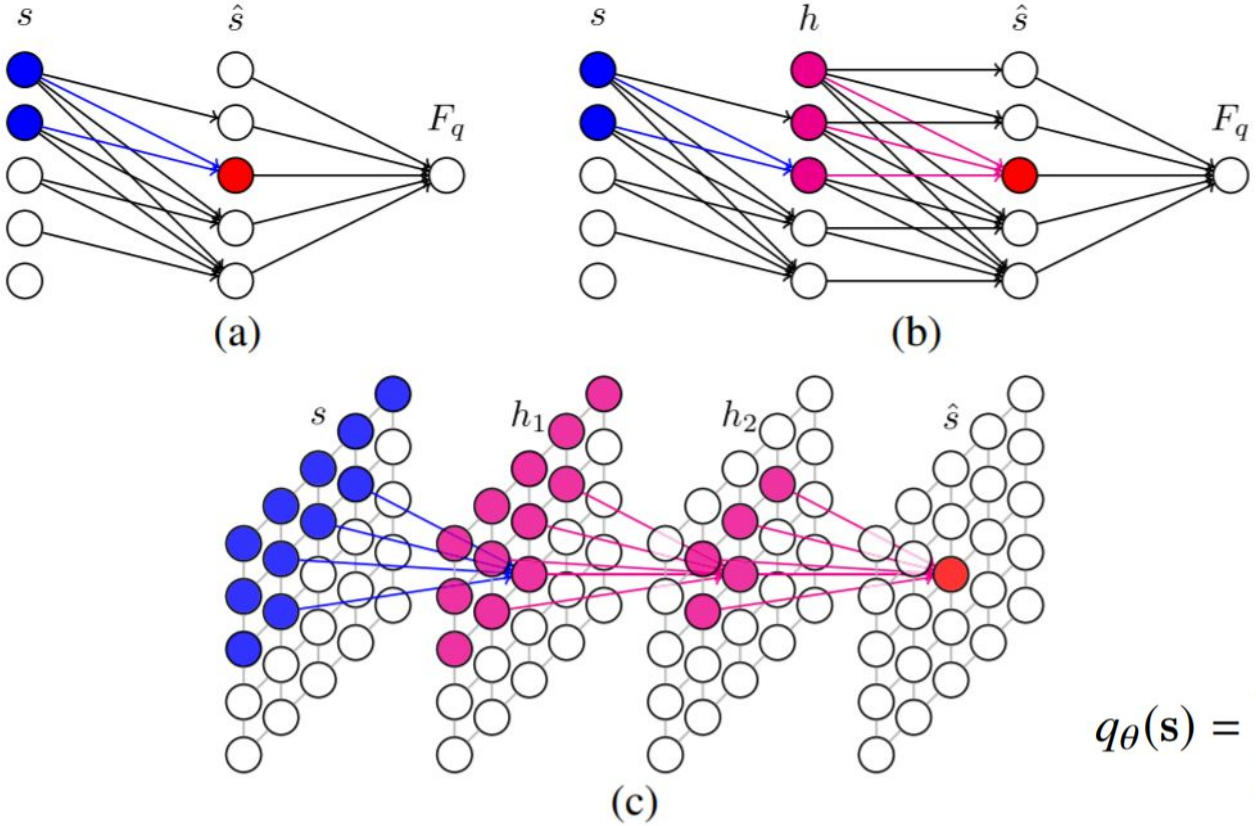
Dian Wu,[1] Lei Wang,[2,3,4,*] and Pan Zhang[5,†]
[1]*School of Physics, Peking University, Beijing 100871, China*
[2]*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[3]*CAS Center for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China*
[4]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*
[5]*Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

(Received 8 November 2018; published 28 February 2019)

We propose a general framework for solving statistical mechanics of systems with finite size. The approach extends the celebrated variational mean-field approaches using autoregressive neural networks, which support direct sampling and exact calculation of normalized probability of configurations. It computes variational free energy, estimates physical quantities such as entropy, magnetizations and correlations, and generates uncorrelated samples all at once. Training of the network employs the policy gradient approach in reinforcement learning, which unbiasedly estimates the gradient of variational parameters. We apply our approach to several classic systems, including 2D Ising models, the Hopfield model, the Sherrington-Kirkpatrick model, and the inverse Ising model, for demonstrating its advantages over existing variational mean-field methods. Our approach sheds light on solving statistical physics problems using modern deep generative neural networks.

DOI: 10.1103/PhysRevLett.122.080602

Autoregressive Networks



$$q_\theta(\mathbf{s}) = \prod_{i=1}^{N} q_\theta(s_i \mid s_1, \ldots, s_{i-1}).$$

Autoregressive networks with different architectures

Variational Autoregressive Networks

Consider a Boltzmann Distribution of the kind $p(\mathbf{s}) = exp(-\beta E(\mathbf{s}))/Z$
The variational approach adopts an ansatz for the joint distribution $q_\theta(\mathbf{s})$ parametrized by variational parameters θ, and adjusts them so that $q_\theta(\mathbf{s})$ is as close as possible to the Boltzmann distribution $p(\mathbf{s})$.
The closeness between the two can be measured by the KL Divergence.

$$D_{\mathrm{KL}}(q_\theta \| p) = \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \ln\left(\frac{q_\theta(\mathbf{s})}{p(\mathbf{s})}\right) = \beta(F_q - F), \quad (2)$$

where

$$F_q = \frac{1}{\beta}\sum_{\mathbf{s}} q_\theta(\mathbf{s})[\beta E(\mathbf{s}) + \ln q_\theta(\mathbf{s})] \quad (3)$$

is the variational free energy corresponding to distribution $q_\theta(\mathbf{s})$.

Variational Autoregressive Network

$$\hat{s}_i = \sigma\left(\sum_{j<i} W_{ij} s_j\right)$$

$$\ln q_\theta(\mathbf{s}) = \sum_{i=1}^{N} \ln\left[\hat{s}_i \delta_{s_i,+1} + (1-\hat{s}_i)\delta_{s_i,-1}\right]$$

$$F_q = \frac{1}{\beta} \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \left[\beta E(\mathbf{s}) + \ln q_\theta(\mathbf{s})\right]$$

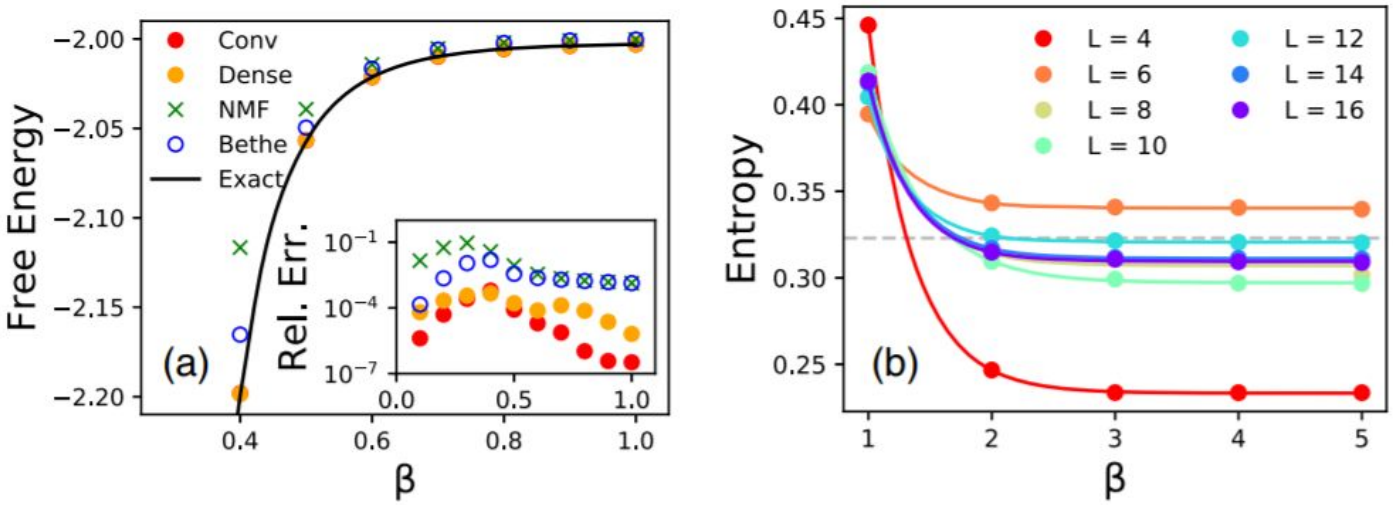
Gradient Estimator

$$\beta \nabla_\theta F_q = \nabla_\theta \sum_{\mathbf{s}} [q_\theta(\mathbf{s}) \cdot (\beta E(\mathbf{s}) + \ln q_\theta(\mathbf{s}))]$$

$$= \sum_{\mathbf{s}} [\nabla_\theta q_\theta(\mathbf{s}) \cdot (\beta E(\mathbf{s}) + \ln q_\theta(\mathbf{s})) + q_\theta(\mathbf{s}) \nabla_\theta \ln q_\theta(\mathbf{s})]$$

$$= \mathbb{E}_{\mathbf{s} \sim q_\theta(\mathbf{s})} \left[\nabla_\theta \ln q_\theta(\mathbf{s}) \cdot \underbrace{(\beta E(\mathbf{s}) + \ln q_\theta(\mathbf{s}))}_{R(\mathbf{s})} \right]$$

Ising Model using VAN



(a) Free energy per site and its relative error of ferromagnetic Ising model on 16×16 square lattice with periodic boundary condition. (b) Entropy per site of antiferromagnetic Ising model on triangular lattices of various sizes L with periodic boundary condition.

Sherrington Kirkpatrick Model

Given a configuration of N Ising spins,

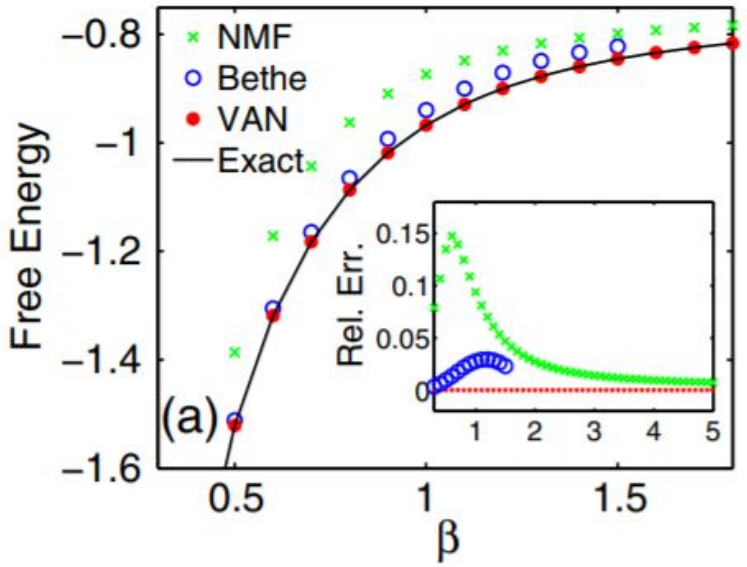
$$\sigma = (\sigma_1, \ldots, \sigma_N) \in \Sigma_N = \{-1, +1\}^N,$$

the Hamiltonian of the model is given by

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j,$$

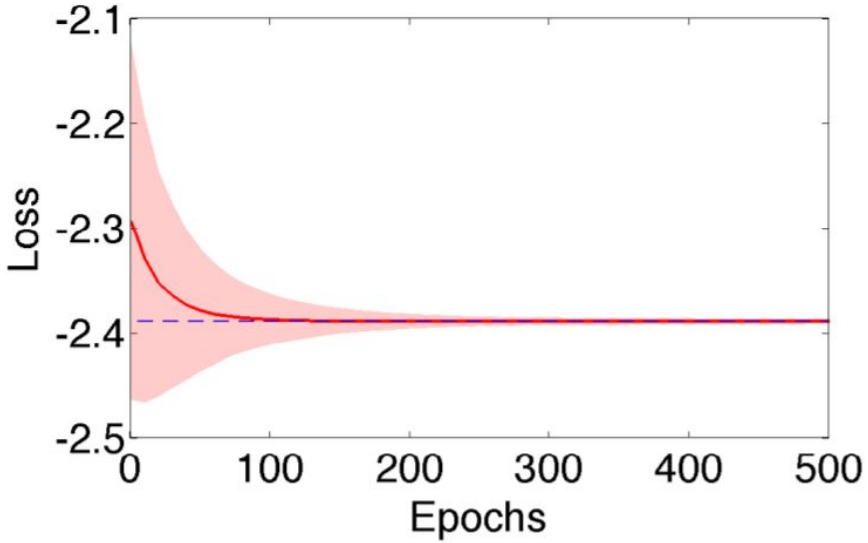
where (g_{ij}) are i.i.d. standard Gaussian random variables

Sherrington Kirkpatrick Model using VAN



Free energy of SK model with $N = 20$ spins. The inset shows relative errors to exact values in a larger β regime. Bethe converges only when $\beta \leq 1.5$.

Sherrington Kirkpatrick Model



Evolution of mean and variance of the loss function during the training of VAN on an SK model with $N = 20$ spins, $\beta = 0.3$. The light red area denotes variance, the red line denotes mean, and the blue dashed line denotes the exact free energy value.

Summary

We looked at 2 different ways of generating probability distributions:

- Restricted Boltzmann Machines are effective for identifying features in a given probability distribution, and regenerating it. This makes them useful in statistical mechanics to generate probability distributions
- In situations where the monte carlo data set is difficult to obtain, but Hamiltonian is known, variational autoregressive networks can be used to generate probability distributions

References

- Torlai, Giacomo & Melko, Roger. (2016). Learning Thermodynamics with Boltzmann Machines. Physical Review B. 94. 10.1103/PhysRevB.94.165134.
- Wu, Dian & Wang, Lei & Zhang, Pan. (2018). Solving Statistical Mechanics using Variational Autoregressive Networks.
- A Beginner's Guide to Restricted Boltzmann Machines (RBMs)
- Neura Networks course at Université de Sherbrooke by Hugo Larochelle ([Training Neural Networks] by Hugo Larochelle)