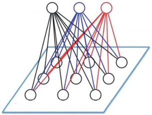
Neural Networks as Thermodynamic Physical Systems

Dimitrios Giataganas
National Sun Yat-sen Univeristy, Taiwan



Talk given for: The Cyprus Institute, CaSToRC, July 4th, 2023, Cyprus.

The usual question is what ML can do for us?

The usual question is what ML can do for us?

An equally interesting question is
what can we do for the ML?

The usual question is what ML can do for us?

An equally interesting question is
what can we do for the ML?

Underlying Link: Coarse Graining!

Outline

Outline

1. Introduction
2. Ising Model
3. Restricted Boltzmann Machines
4. RBM flow vs RG flow
5. Conclusions

This Talk

- We attempt to investigate fundamental relations between the process of learning and principles of physics or physical models.
- To do that we need to choose a theory and employ ML methods on a physical model: The Ising model and later the Potts model.
- Why Ising? It is binary, simple and has rich structure=phase transitions.
- Why Potts? It generalizes the binary states, still rich structure=phase transitions.
- We look for evidence of this relation at the "special points" of the Ising Model: The points where phase transition occurs.
  - These are the critical points of the Renormalization Group flow.
  - There the theory is scale invariant!
  - There certain thermodynamic properties take special values.

Statements for the learning process:

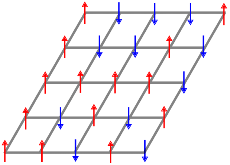
- The machine knows nothing about Hamiltonian, interactions and phase transitions!
- It is trained using (many!) state configurations we generate with Monte Carlo at a range of temperatures.

- Our ML methods spontaneously identify the critical phase, what is the reason?
- A step further: Can we compute any observables with this process?

Outline

Ising Model

- N-site square lattice with binary variables = spins s_i. *(Lenz 1920)*



- Each site, labeled by the index i, contains a spin s_i with values ± 1 which represent the two possible states.
- The Hamiltonian is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - H \sum_i s_i \,,$$

where $\langle ij \rangle$ the nearest neighbor pairs of the sites i and j,
J is the coupling of the nearest neighbors,
H is the external magnetic field.

- The magnetization M of this system is defined as the sum of all the spins $M = \sum_i s_i$.
- The partition function of the system reads

$$\mathcal{Z} = \sum_{\{s\}} \prod_i e^{K s_i s_{i+1} + h s_i}$$

where $K := J/T$ and $h := H/T$.

1-dim Ising Model

It is a spin chain and is exactly solvable. *(Ising 1924)*

- In the thermodynamic limit

$$\mathcal{Z} = \left(e^{K} \cosh h + \left(e^{-2K} + e^{2K} \sinh^2 h\right)^{1/2}\right)^{N}$$

And the magnetization per site

$$m := \frac{M}{N} = \frac{\sinh h}{\sqrt{\sinh^2 h + e^{-4K}}}\,.$$

- Trivial Phase transition!

2-dim Ising Model

It is exactly solvable without magnetic field. *(Onsager 1944)*

- It has a 2nd order phase transition at

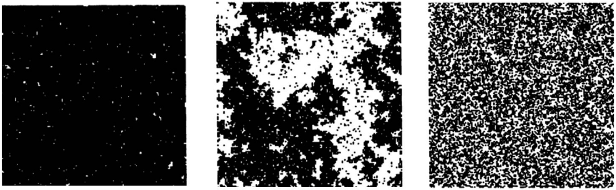
$$K_c = \frac{J}{T_c} = \frac{\log\left(1+\sqrt{2}\right)}{2} \,,$$

where the specific heat $C = (\partial E/\partial T)_H$ diverges.

- The magnetization per site below the T_c

$$m = \left(1 - \sinh^{-4} 2K\right)^{1/8} .$$

- The spin state configurations at different temperatures:



T<<Tc T~Tc T>>Tc

Renormalization Procedure: Key Idea

- Successive decimation of degrees of freedom.
- Macroscopic modes are respected while microscopic are integrated out and averaged.
- Results to effective field theory for long distance degrees of freedom with given macroscopic laws.
- “The guiding principle in formulating the new interactions (and the process) is to reproduce as accurately as possible the observed probability distribution.”

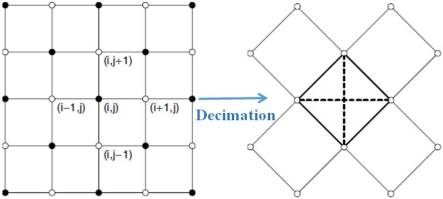
(Wilson 79)

- Intuitive similarity with the RBM methods.

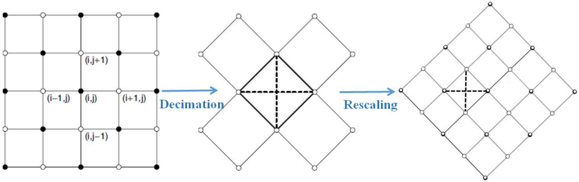
Lets be more precise.

Renormalization Procedure

- RG is a semi-group of transformations R.
- $\mathcal{H}' = R[\mathcal{H}]$, R is a non-linear transformation of the coupling parameters.
  It does a coarse graining/decimation removing the degrees of freedom $N' = N/b$, while keeping the partition function invariant $\mathcal{Z}_{N'}[H'] \sim \mathcal{Z}_N[H]$.



- It is combined with a rescaling of lengths: $r' = r/b$, to restore spatial density and renormalization of physical variables to restore the relative fluctuations.



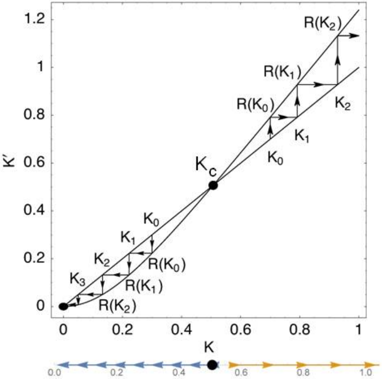
- A fixed point of the RG transformation defining a fixed point Hamiltonian H_0, is a point in the coupling parameter space where $R[H_0] = H_0$.

RG in 2-dim Ising model

- Approximately the RG flow can be encoded in the coupling equation

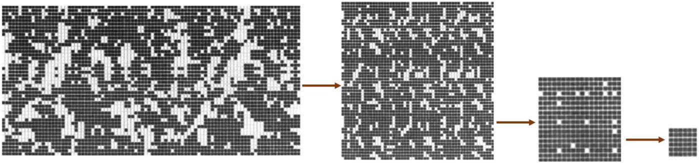
$$K' = \frac{3}{8} \log\cosh 4K \, .$$

- The fixed points are at $K = 0, \infty$ and $K_c \simeq 0.507$.
- K_c is an unstable critical point:

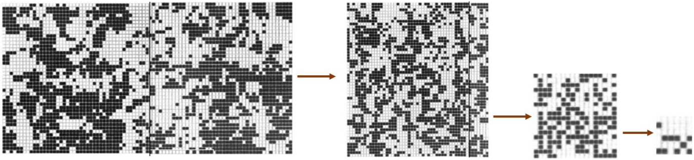


Block-Spin decimation on the 2-dim Ising Lattice

- Performing it just below T_c we quickly obtain the ordered state.



- Performing it at T_c we remain to the scale invariant state.

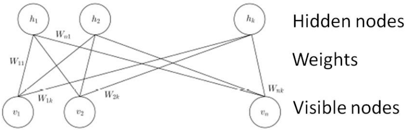


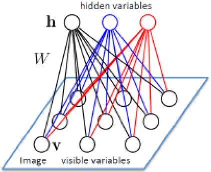
- Performing it just above T_c we obtain the the random state.

Outline

Restricted Boltzmann Machines (RBM)

- RBM: Energy based, undirected graphical models, which can be interpreted as stochastic neural networks.





- RBM: No connection between nodes of the same layer.
- Two layers: one visible to represent data (e.g. one visible unit for each pixel) and one hidden (e.g. model dependencies of the pixel of images).
- The hidden layer is where the network stores its internal abstract representation of the training data.
- W is the connection strength between visible and hidden neurons. $v_i(h_j)$ is the relevant state of the visible (hidden) unit.

- Energy function on states

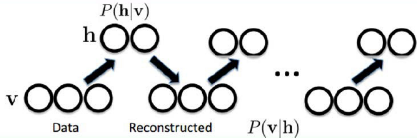
$$E(v,h) = -b_i v_i - c_\alpha h_\alpha - h_\alpha W_{\alpha i} v_i \ ,$$

  b,c biases for the visible and hidden neurons; W is the matrix of weights. $i = 1, \ldots N,\ \alpha = 1, \ldots M$.
- Weights and biases of a model $(W, b, c) :=$ model parameters θ.
- Joint Probability Boltzmann-Gibbs distribution: Probability to observe a state (v, h) via the energy of the model E.

$$p(v,h) = \frac{e^{-E(v,h)}}{\mathcal{Z}} \ , \qquad \mathcal{Z} = \sum_{v,h} e^{-E(v,h)}$$

  $\mathcal{Z}$ is the partition function and acts as a normalization.
- The marginal probability of a visible vector v assigned by the network

$$p(v) = \frac{1}{\mathcal{Z}} \sum_h p(v,h) = \frac{1}{\mathcal{Z}} \prod_j e^{b_j v_j} \prod_\alpha \left(1 + e^{c_\alpha + W_{\alpha i} v_i}\right)$$

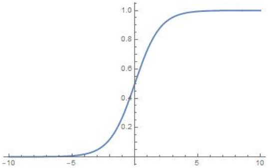


- Hidden variables are independent given the state of the visible variables and vice versa. The conditional probabilities:

$$p(h|v) = \prod_{\alpha} p(h_\alpha|v) \ , \qquad p(v|h) = \prod_{j} p(v_j|h)$$

- The conditional probabilities of a single variable expressed in sigmoid functions:

$$p(h_\alpha = 1|v) = \sigma(c_\alpha + W_{\alpha i} v_i) \ , \qquad p(v_j = 1|h) = \sigma(b_j + h_\alpha W_{\alpha j}) \ ,$$



Training an RBM

- Training an RBM: Adjusting the RBM parameters θ, such that the model probability distribution $p(v) = \frac{1}{\mathcal{Z}} \sum_h p(v, h)$ represents the given probability distribution $q(v)$ as faithfully as possible.

- Target: Minimize the distance between the distribution q of the sample data and the reconstructed distribution p.

Defining The Distance

- A candidate is the Cross Entropy

$$H(q,p) = -\sum_{v \in V} q(v) \log p(v) \ .$$

- The convenient measure is the Kullback-Liebler divergence (= information lost, relative entropy) between distributions $q(v_i)$ and $p(v_i)$:

$$KL(q,p) = H(q,p) - H(q,q) = \sum_{v \in V} \left(q(v) \log q(v) - q(v) \log p(v) \right)$$

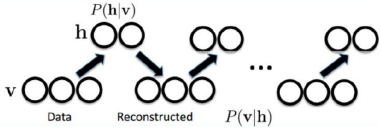
- Gibbs inequality: $KL(q,p) \geq 0$.
- **Training of RBM = minimize the KL measure.**
- $KL(q,p) \to min \quad \Rightarrow \quad \prod_j p(v_j) \to max$

Training the RBM

- Aim: Maximize $p(v)$:

$$\frac{\partial \log p(v)}{\partial \theta} = -\sum_h p(h|v)\partial_\theta E + \sum_v p(v) \sum_h p(v|h)\partial_\theta E \ .$$

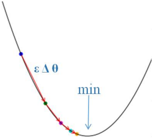
- Exponential complexity problem!
- Certain Approximations on averaging of variables.



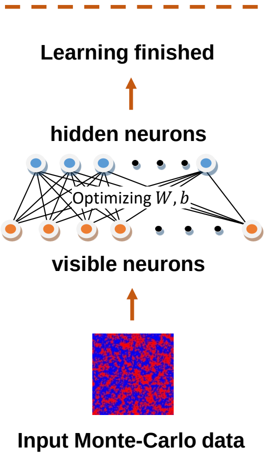
- Then gradient descent to update on the parameters

$$\theta \to \theta - \epsilon \Delta\theta \ ,$$

with ϵ the learning rate.



Learning process

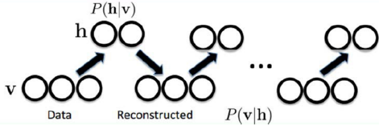


Example of Data: 10×10 (2d) lattice, 1000 configs at each (T,H), where T=0, 0.5, ..., 9.5 and H=0, 0.5, ..., 4.5.

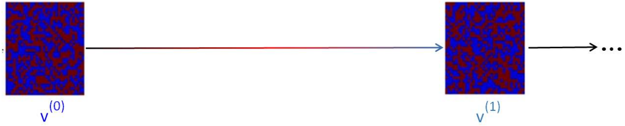
RBM: $N_v = 100$, $N_h < N_v$, learning rate: $\epsilon = 0.001$, epoch $= 10000$ (renewal procedure). **Total** $\sim 10^9$ **steps to train the RBM.**

The RBM flow of Reconstructions

- With learning we have fixed the RBM parameters (b, c, W). Once the training finished we generate the RBM flows.
- Using the initial faithful spin distribution function $v_j^{(0)} = s_j$.
- The generation the RBM flows of n steps is derived as:



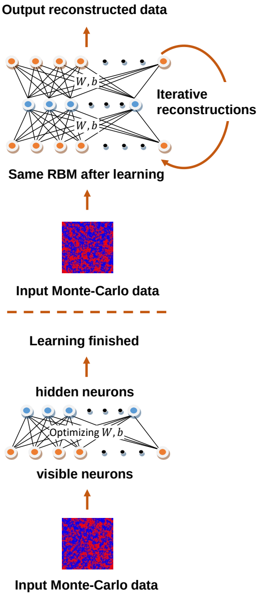
$$v_j^{(0)}(= s_j) \rightarrow h_\alpha^{(1)}(= \sigma(W_{\alpha j} s_j + c_\alpha)) \rightarrow v_j^{(1)}\left(= \sigma(W_{\alpha j} h_\alpha^{(1)} + b_j)\right) \rightarrow$$
$$h_\alpha^{(2)} \rightarrow v_j^{(2)} \rightarrow \ldots \rightarrow h_\alpha^{(n)} \rightarrow v_j^{(n)}$$



How can we interpret the outcome $v^{(n)}$?

(Iso, Funai-Shiba 2018; Funai-Shiba, D.G. 2018; D.G.,etal 2021)

Learning and Reconstruction

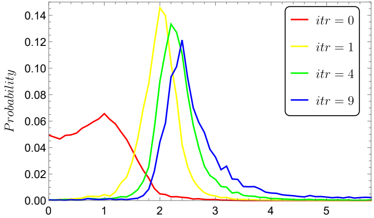


Outline

Zero Magnetic Field RBM Flow vs RG Flow

The RBM flow

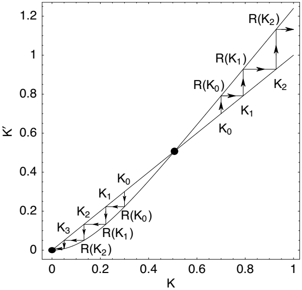
$q_0(v_j) \to r_1(h_i) \to q_1(v_j) \to \ldots \to$
$q_9(v_j) = \{v_j^0\} \to \{h_i^1\} \to \{v_j^1\} \to$
$\ldots \to \{v_j^9\}$



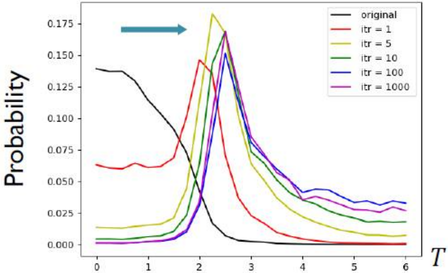
Reconstruction of $T = 1$ microstate, which flows to $T_c = 2.27$ critical point.

The RG flow

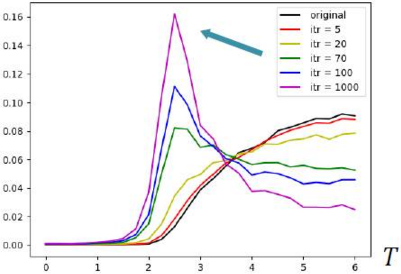
$K_0 \to K_1 \to K_2 \to \ldots \to K_n$



The RBM flow, starting at the "extremal" points:



Reconstruction of T=0 configs

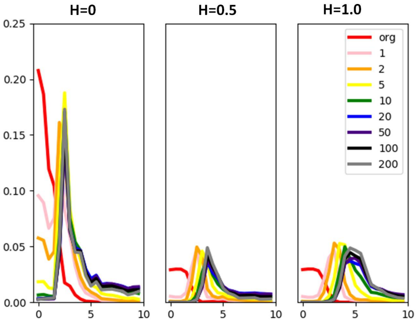


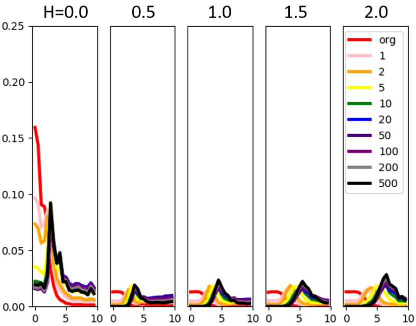
Reconstruction of T=6 configs

RBM flow spontaneously to the critical fixed point of the spin system!

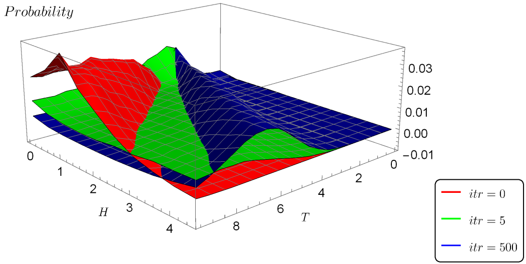
RBM Flows for Various Magnetic Field

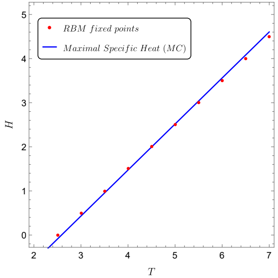
- When $H \neq 0$, no phase transition(critical fixed points), only trivial fixed points!





- There exist an RBM flow fixed point that does not match the RG fixed point.
- Puzzling Behavior!



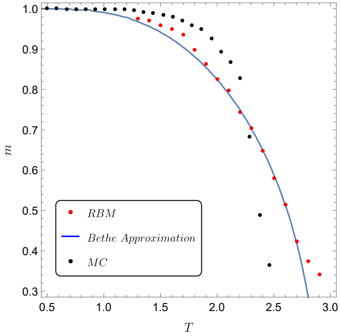


- The flow forms a pattern after a small number of iterations already already has a clear peak.
- RBM fixed points and maximal points of specific heat in 2d Ising model for reconstructed flows with fixed magnetic fields.

Ising thermodynamics relation to RBM flow instead of RG?

Critical Exponents RBM Flow?

From the recostructed configurations we can obtain the critical exponents.



Around Critical Temperature observables exhibit power law behavior.

Critical Exponents RBM Flow

- The magnetization around the critical point can be expanded to give

$$m \sim 1.222 \frac{|T - T_c|^{1/8}}{T_c}$$

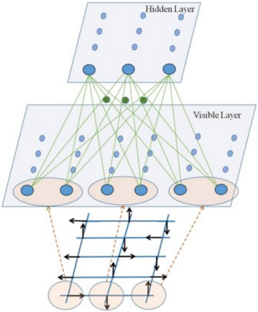
where the critical parameter is $\beta = 1/8$.

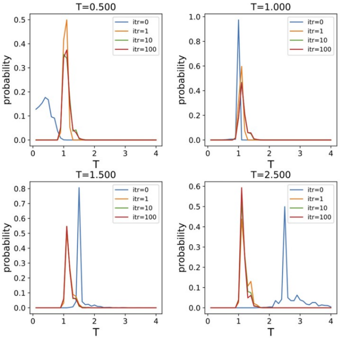
- The recostructed configurations at around T_c give with large errors

$$m \sim 0.931 \frac{|T - T_c|^{0.127}}{T_c}$$

Other Spin Models: Potts

- A generalization of the Ising model where instead of 2: spin-up, spin-down we have q states.
- The training set up is even more arbitrary, e.g. for $q = 4$:





- The RBM flow generates similar microstates converging to the critical point T_c.

Outline

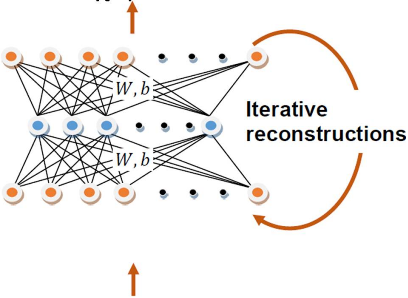
Conclusions

- No prior knowledge about the criticality of the system and its Hamiltonian for the RBM! It is trained to learn patterns of the spin configurations.

✓ The RBM flow of reconstruction approaches spontaneously the spin configurations of the RG fixed points for spin models.
The convergence depends on # of NN parameters.

- RBMs with standard Monte Carlo methods can be used as a powerful tool to study physical models and to reconstruct the thermodynamic quantities accurately.
- RBM is fundamentally related to RG and / or thermodynamics of physical systems! (many possible explanations)
(Funai-Shiba, D.G.: Phys.Rev.Res. 1810.08179 ; de Mello Koch, de Mello Koch, Cheng: 1906.05212; D.G., etal: 2102.05219; Hou, You: 2306.11054;...)

Ευχαριστώ



Thank you