Renormalization, Thermodynamics, and Feature Extraction of Machine Learning

Shotaro Shiba Funai (OIST)

Collaborators: S. Iso, S. Yokoo (KEK) and D. Giataganas (NCTS)

References: arXiv: 1801.07172 [hep-th], 1810.08179 [cond-mat]

Nov. 6, 2018 @ Nagoya Univ.

What is machine learning ?

- An attempt to reproduce actions of consciousness in a computer. There are two directions of research:

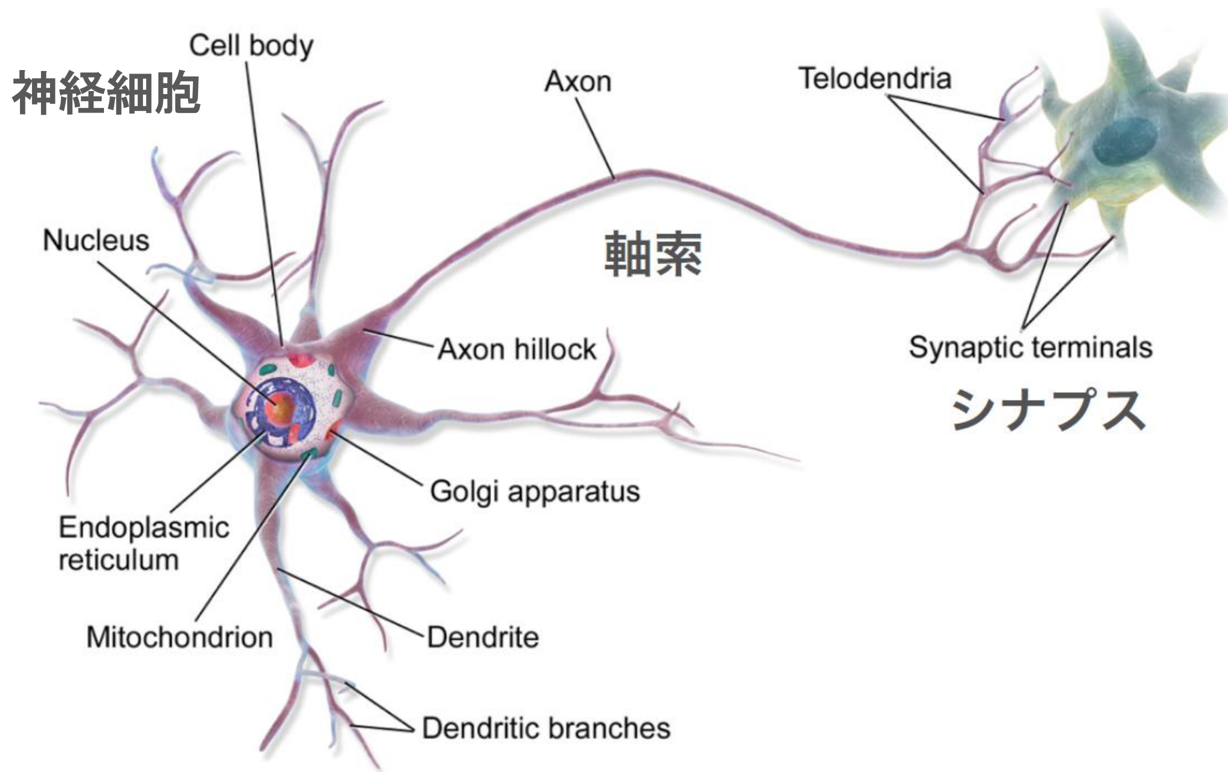
1. By teaching a computer on various rules, we design a machine which can judge things like humans.
(e.g., Expert system)
2. By emulating a structure of human brain, we design a machine which can learn and judge information.
→ Machine learning (ML)

It finds the rules!

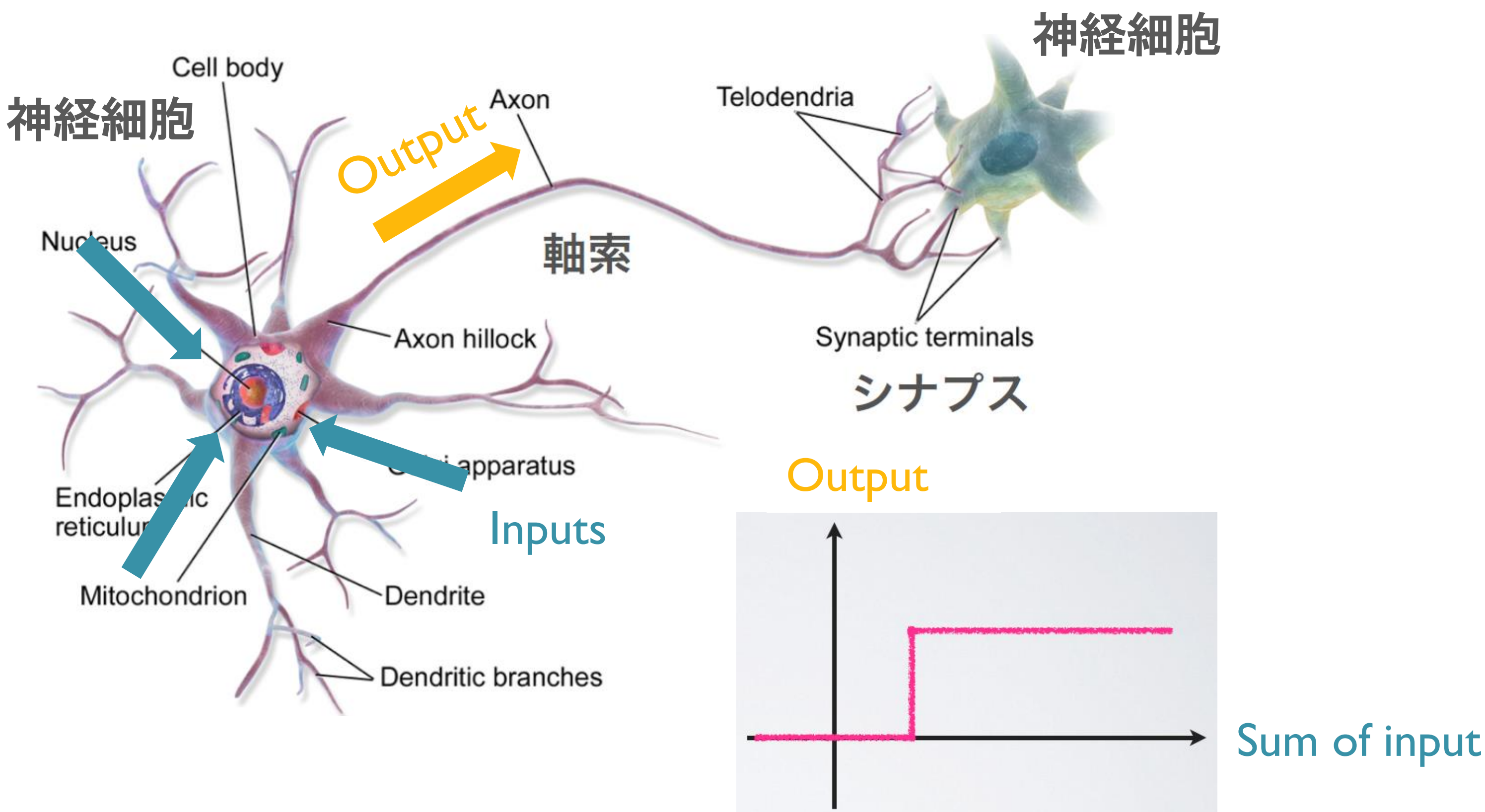
- Both researches are ongoing now, but recently the latter has been greatly developed.

Structure of human brain

- Human brain has about 100 billion neurons（神経細胞）and they are connected via axons（軸索）.

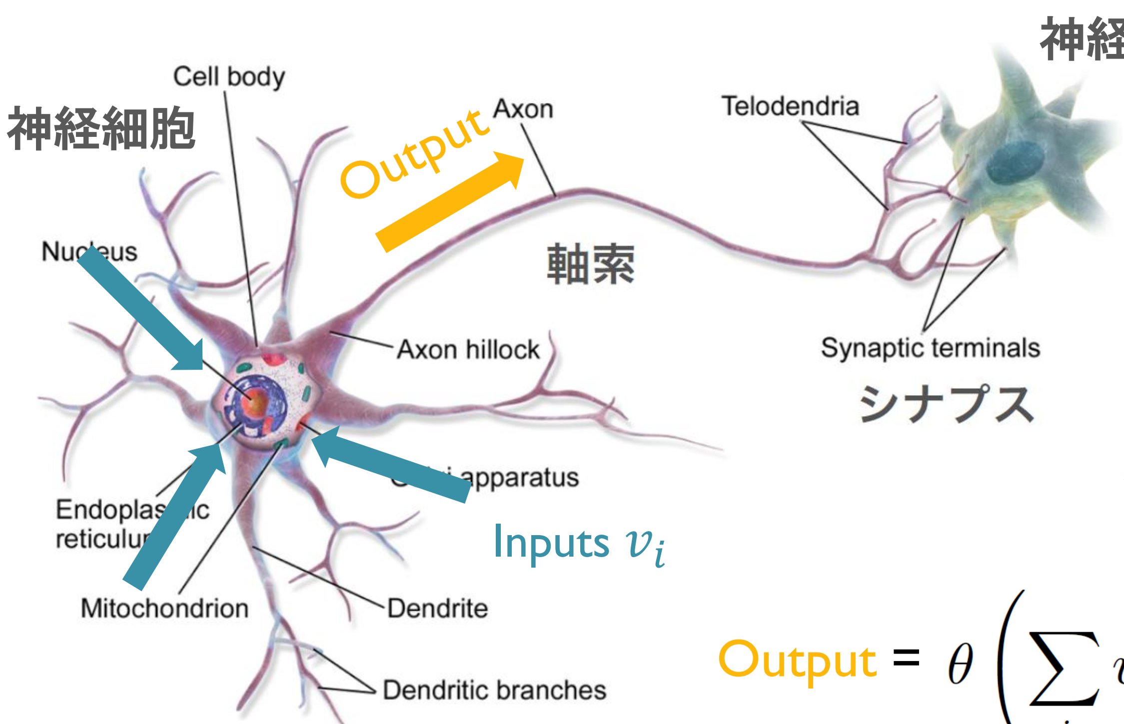


- A neuron receives electric signals sent from other neurons through axons.
- If a sum of input signal exceeds a threshold, the neuron fires and sends a signal to other connected neurons.



- Humans repeat trials and errors. After such experiences, our neurons renew a way to exchange the signals so that we can judge various things more properly.

→ This is nothing but "learning."



renewed by "learning"

weights w_i
bias (threshold) b

$$\text{Output} = \theta\left(\sum_i v_i w_i + b\right)$$

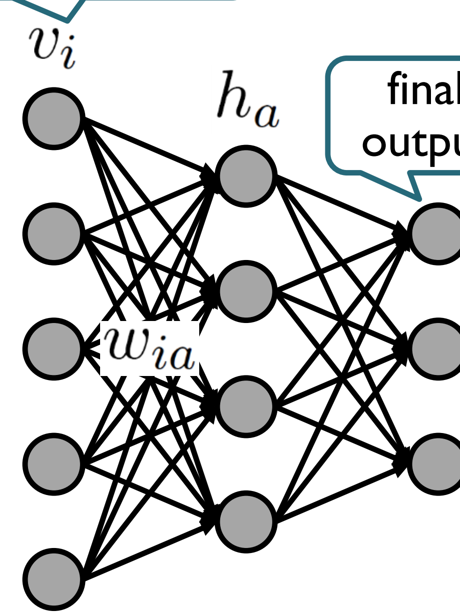
Input v_i

Algorithm of machine learning

- By emulating a structure of brain, we make an algorithm of machine learning.
- We reproduce a network of neurons exchanging signals, such that

$$h_a = f\left(\sum_i v_i w_{ia} + b_a\right)$$

nonlinear function (activation function)

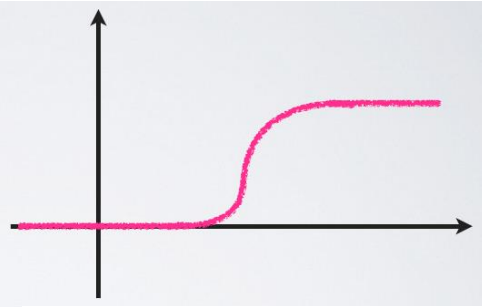


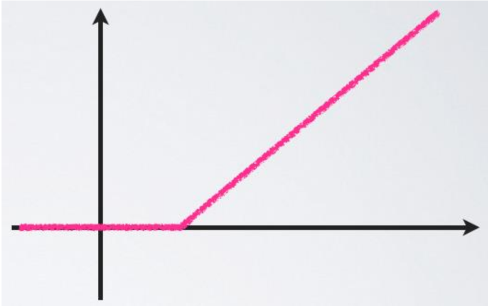
- We adjust weights w_{ia} and bias b_a so that the final output approaches desired values (answer) for us.

"training"

- As an activation function, we don't use step function but sigmoid function (left) or ReLU (right), because of analyticity.

tanh, too





$$f(x) = \frac{1}{1+e^{-x}}\,, \qquad f(x) = \begin{cases} x & \text{for } x \geq 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

- For the final output, the softmax function is often used, since we can interpret the output as probability.

$$g(x_i) = \frac{\exp x_i}{\sum_j \exp x_j}$$

- In order to adjust weights w_{ia} and bias b_a…

- We choose the loss function which evaluates difference between output at present and desired output. Square sum or relative entropy is often chosen.

$$E = \frac{1}{2}\sum_n \left(v_L^{(n)} - y^{(n)}\right)^2$$

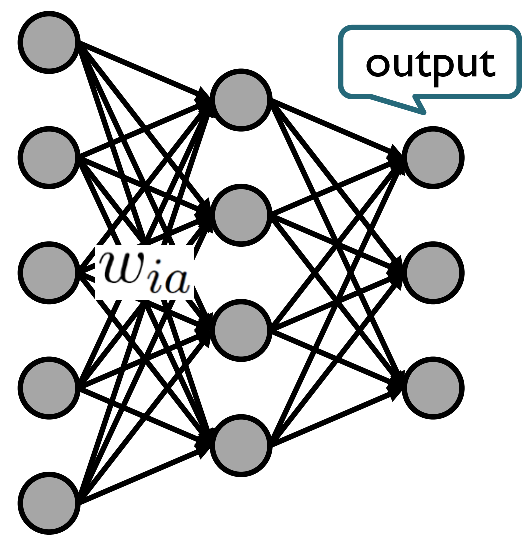
For probability distributions (later)

- Then we calculate weights and bias such that the loss function becomes the minimum.

  Analytical calculation is *impossible*, since we can't solve nonlinear eqs with many variables.

- Instead, we use *numerical* calculations to find (practically) a local minimum by iterative approximation.

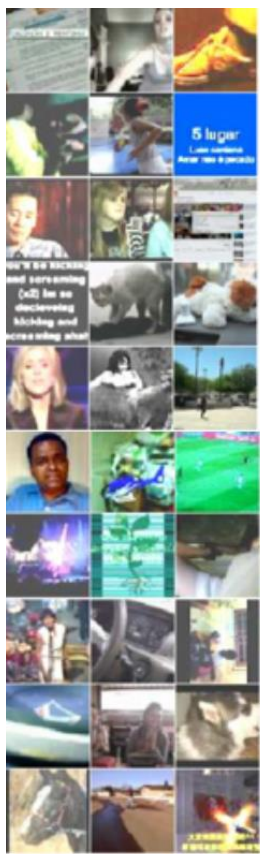
"training"



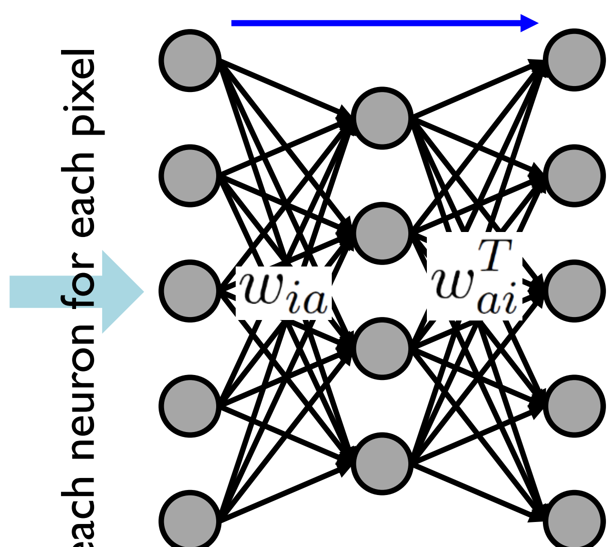
Google's cat (in 2012)

- Using such an algorithm, we can get interesting results. For example,

[Le et al., '12]



Input 10 million still images clipped from YouTube movies.



Desired output = input itself (autoencoder)

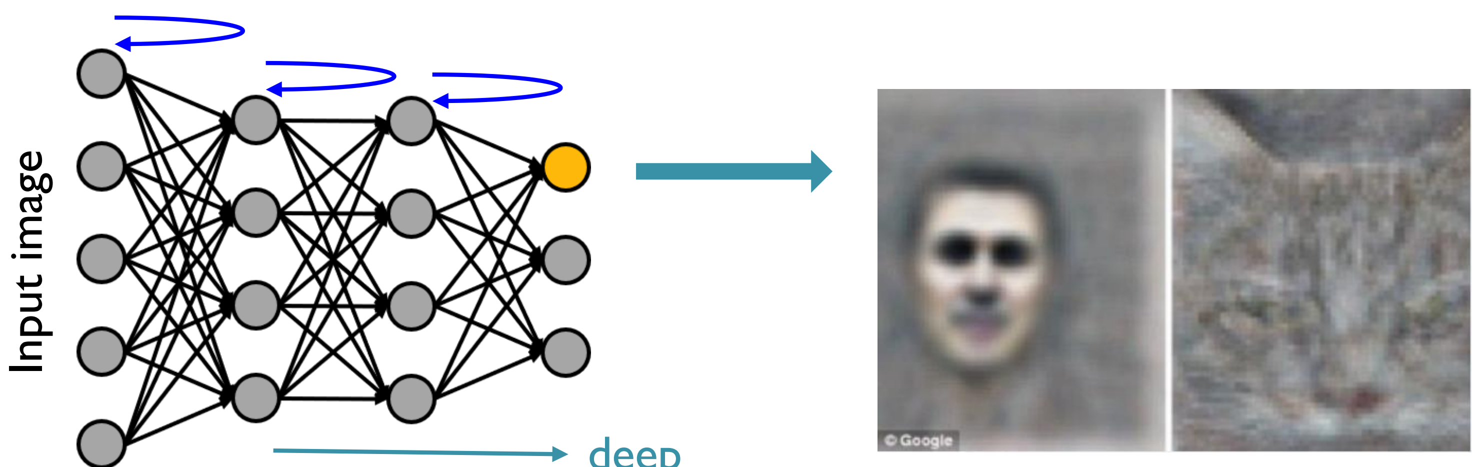
Humans don't teach anything but inputs! (unsupervised)

A network of neurons (neural network, NN)

➢ What do the neurons learn?

- If we make images which only a specific neuron react to, a human face or a cat's face appears.
- There are also neurons which react to simpler figures, such as a line, an edge or a triangle.
- In general, neurons in deep layer react to complicated things. (We deepen a NN to combine many autoencoders.)
- This may reproduce a human process of grasping "features."

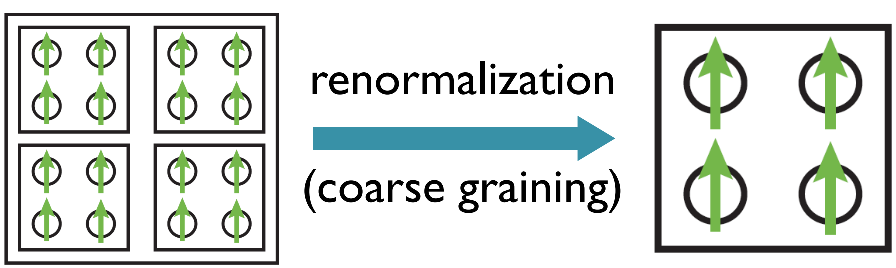
from large number of input data



- What does grasping "features" mean?
  - An image contains various information, but we extract an important part as its features and drop the other parts.
  - It is similar to the coarse graining, and then may be related to the renormalization group (RG).
  - Going along the RG flow, relevant parameters (~features) are emphasized while irrelevant parameters are dropped.

iterative RG transformations

- Let us discuss a relation of feature extraction in ML and renormalization in physics!

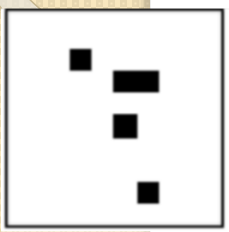


[Mehta-Schwab, '14]
[Lin-Tegmark-Rolnick, '16]
[Sato, '16]
[Aoki-Kobayashi, '16]
[Koch-Janusz, Ringel, '17]

Our experiments and results

Our experiment (1)

[Iso-SSF-Yokoo, '18]
[SSF-Giataganas, '18]

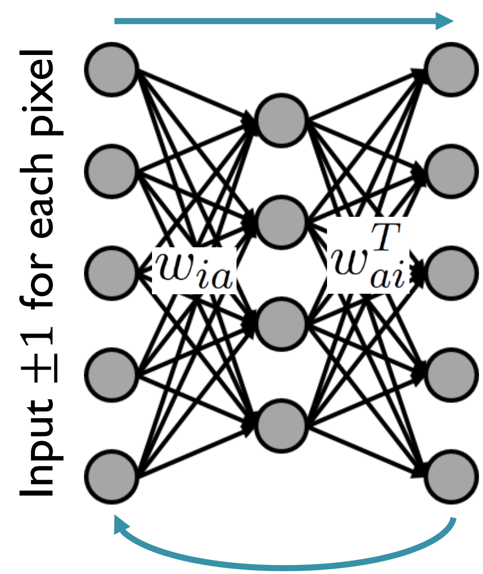


T=2, H=0

T=6, H=0

1. We generate spin configurations of Ising model (black-and-white images) using Monte Carlo simulation, since we know well about RG in Ising model.
2. We train a NN so that when we input the configs it outputs images as similar as possible to inputs. (Autoencoder, unsupervised learning)
3. After training (the weight is fixed), we input again the output configs. Doing this iteratively, we obtain the flow of configs.

Relation to RG flow?

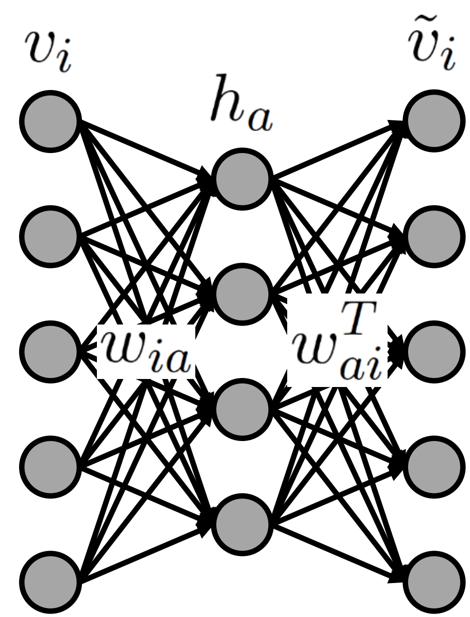


Autoencoder (unsupervised learning)

- An autoencoder, which plays important roles in "Google's cat," is believed to extract "features" of input images.

- It can be related to the coarse graining: a NN compresses images and then reconstructs them.
- We train a NN so that it outputs the (ideally) same images as inputs with the same probability.
- This type of autoencoder is called Restricted Boltzmann Machine (RBM).

Inputs contain configs at *various* (T, H)



- The probability to output an image is defined, using the "energy" function

$$E(\{v_i\},\{h_a\}) = \sum_{i,a} v_i w_{ia} h_a + \sum_a b_a h_a + \sum_i c_i v_i$$

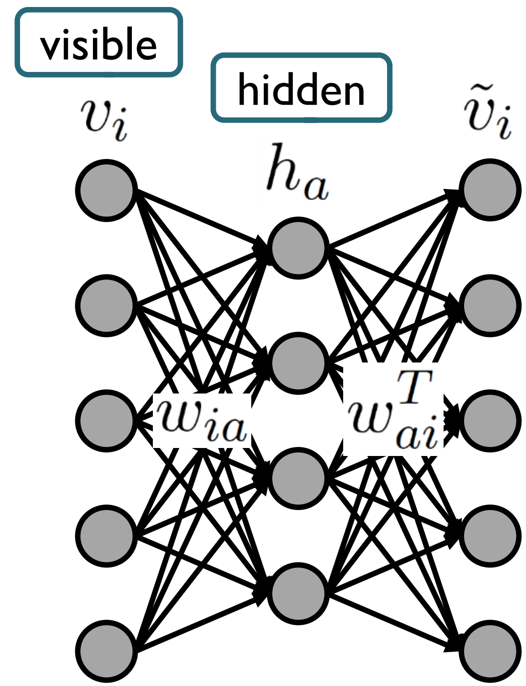
weights w_{ia}, bias b_a, c_i

Statistical physics

by Boltzmann distribution

$$p(\{h_a\}) = \sum_{\{v_i\}} \frac{e^{-E(\{v_i\},\{h_a\})}}{\mathcal{Z}}$$

$$\tilde{p}(\{\tilde{v}_i\}) = \sum_{\{h_a\}} \frac{e^{-E(\{\tilde{v}_i\},\{h_a\})}}{\mathcal{Z}}$$



- We train the RBM (weights and bias) so that the relative entropy approaches a local minimum.

Distance between probability distributions

$$\sum_{\{v_i\}} q(\{v_i\}) \log \frac{q(\{v_i\})}{\tilde{p}(\{v_i\})}$$

- The relative entropy is also called KL divergence:

$$\sum_{\{v_i\}} q(\{v_i\}) \log \frac{q(\{v_i\})}{\tilde{p}(\{v_i\})}$$

prob of an input image = v_i / prob of an output image = v_i

➢ In our experiments, the input images are the spin configs in Ising model: $v_i = \pm 1$ for a white/black pixel.

- The expectation values of outputs are those of spins:

$$\langle h_a \rangle = \tanh\left(\sum_i v_i w_{ia} + b_a\right)$$

$$\langle \tilde{v}_i \rangle = \tanh\left(\sum_a h_a w^T_{ai} + c_i\right)$$

- The final output (reconstructed) images have spins $\tilde{v}_i = \pm 1$ by replacing an EV $\langle \tilde{v}_i \rangle$ with a probability $(1 \pm \langle \tilde{v}_i \rangle)/2$.

To keep same EV

- The probability distribution of input configs $q(\{v_i\})$ and that of output configs $\tilde{p}(\{v_i\})$ are slightly different, even after the training finished.
  (It's because the KL divergence cannot be zero, practically.)
- If we input again the output configs, we obtain another prob distribution $\tilde{\tilde{p}}(\{v_i\})$ of reconstructed configs.
- Doing this iteratively, we get the flow of prob distribution of spin configs: $q(\{v_i\}) \to \tilde{p}(\{v_i\}) \to \tilde{\tilde{p}}(\{v_i\}) \to \dots$

➢ Questions:

This is a well-defined question!

1. Does the "RBM flow" correspond to the RG flow?
2. Does it have the fixed points describing the "features"?
   (The features are *emphasized* along the RBM flow.)

Our experiment (2)

[Iso-SSF-Yokoo, '18]
[SSF-Giataganas, '18]

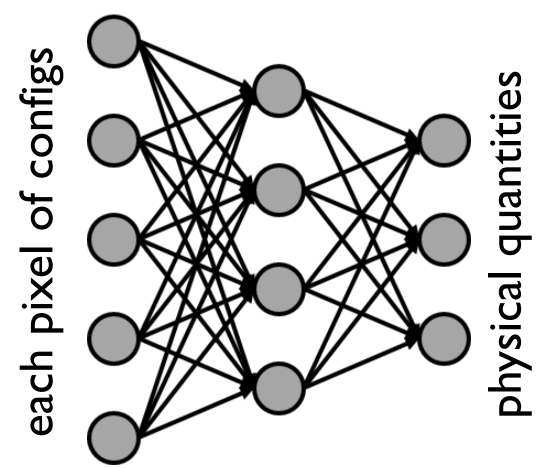
➢ We check if the RBM flow is related to the RG flow.

- Let us translate the flow of spin configs into a flow of physical quantities (temperature T and magnetic field H), since it makes our discussion easier.

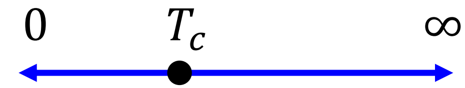
- To do this, we train *another* NN to output correct values of (T, H) of input configs. (supervised learning)

  parameters when generated by MC simulation

- Using this NN, we get a flow of the physical quantities $\phi = (T, H)$:
  $\phi(\{v_i\}) \to \tilde{\phi}(\{v_i\}) \to \tilde{\tilde{\phi}}(\{v_i\}) \to \ldots$



- For example, in 2d Ising model (at H=0),



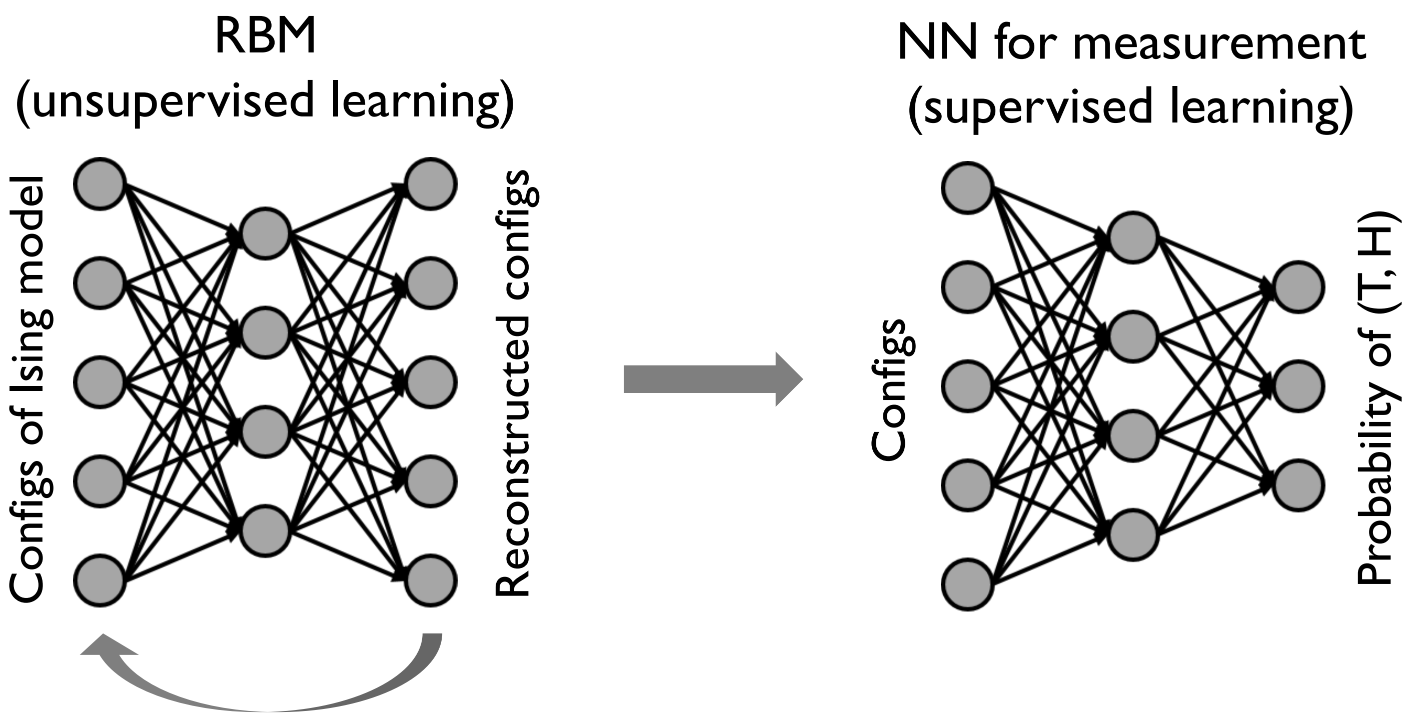
- RG flow goes away from the critical temperature $T_c = 2.27$ and approaches to $T = 0, \infty$.

Phase transition occurs

coupling $J = 1$ fixed

- If RBM flow behaves similarly, it should correspond to the RG flow (as we expected).

Summary of our setup

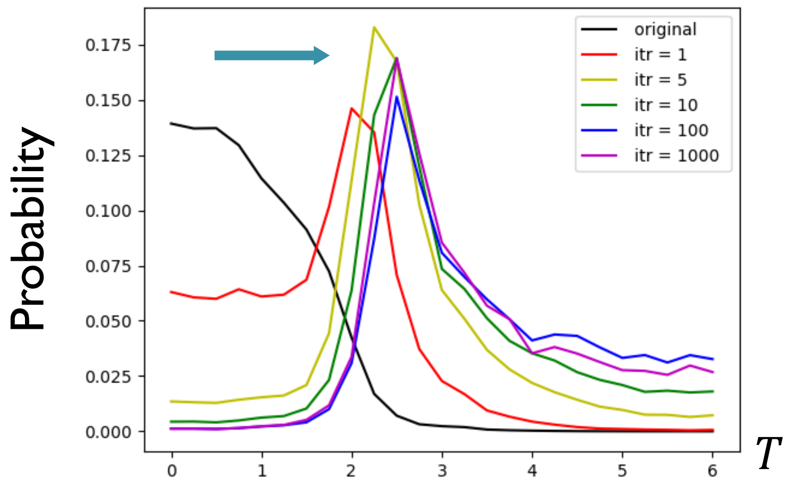


Results: obviously different!

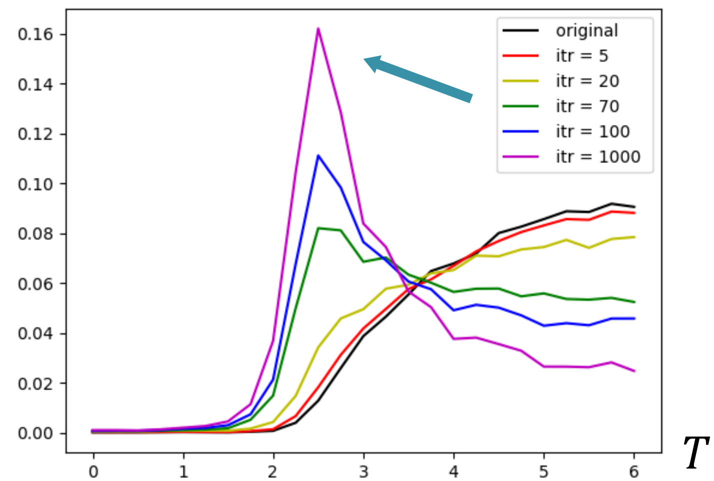
[Iso-SSF-Yokoo, '18]

2d case at H=0

- The RBM flow approaches the critical point, while goes away from $T = 0, \infty$. It's the opposite direction to the RG flow!



Reconstruction of T=0 configs



Reconstruction of T=6 configs

- Data: configs in 10x10 lattice, 1000 configs at each T=0, 0.25, …, 6, H=0. (Same results when T=0, 0.25, …, 10, ∞ / T=0, 0.25,…, 2 and 4, 4.25,…, 6.)

$n_h = 81$

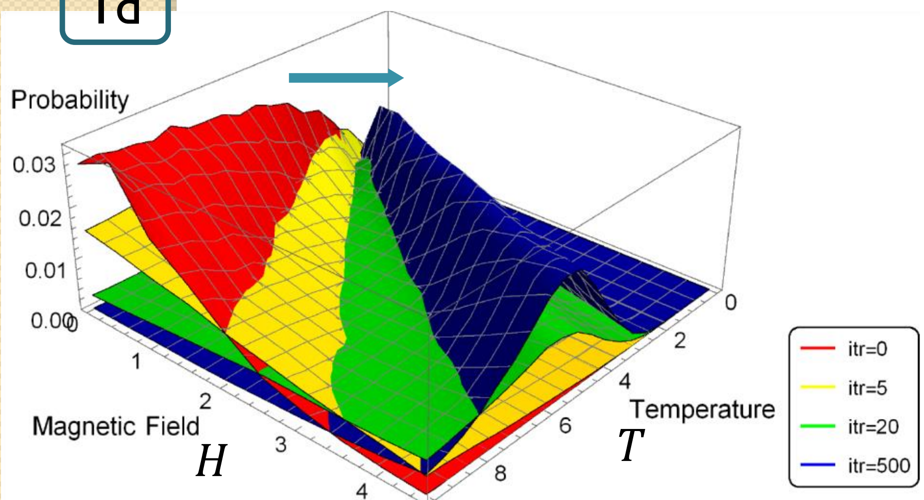
- RBM: $n_v = 100, n_h \leq n_v$, learning rate = 0.1, epoch = 5000

[SSF-Giataganas, '18]

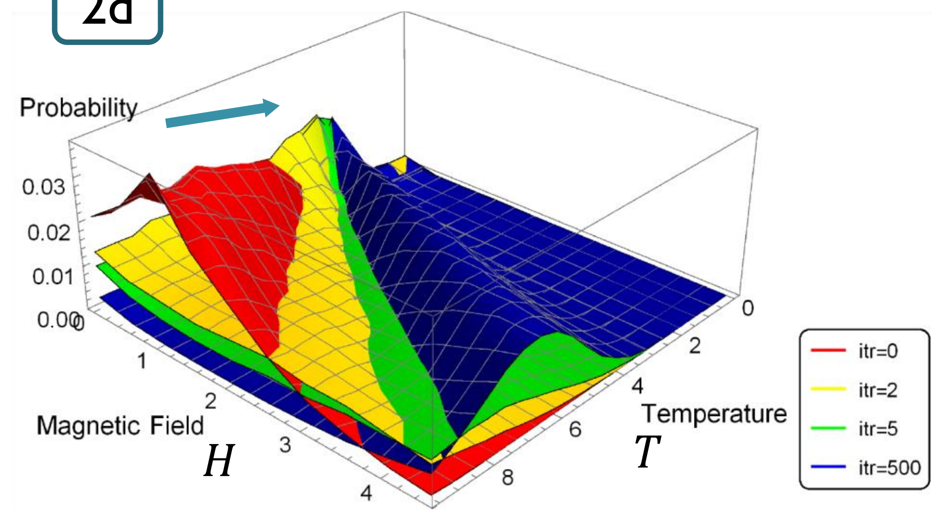
- In 1d and 2d Ising models including $H \neq 0$ region:
  - The RBM flow approaches the fixed points in (T, H) space. Wherever a start point is, the flow arrives at the *same* points.
  - But they are different from the RG flow and its fixed points.

explain it later

1d



2d



- Data: configs in 100 (1d) or 10x10 (2d) lattice, 1000 configs at each (T,H), where T=0, 0.5, …, 9.5 and H=0, 0.5, …, 4.5.
- RBM: $n_v = 100, n_h \leq 16$, learning rate = 0.001, epoch = 10000

$n_h = 9$

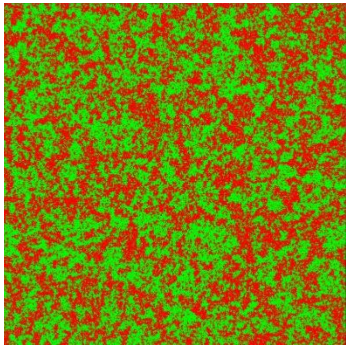
Discussions on our results

RBM flow has fixed points!

- This is (perhaps the only) similarity to RG flow.
  - The fixed points are in the space of physical quantities (T, H), not that of configurations.
  - Along RBM flow the extracted features are emphasized, then its fixed points should be the "features" of learning data.
- In 2d case at H=0, fixed points exist at the same point.
  - But the flows go in the opposite directions. (stable pt in RBM flow = unstable pt in RG flow)
  - What is the "feature" extracted by the RBM? It is probably the scale invariance…?

$T = T_c$

proposal



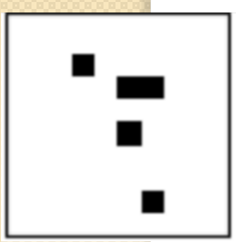
Some evidence for scale invariance

- Let us compare the two kinds of RBM by analyzing the RBM flows and their weights. *[Iso-SSF-Yokoo, '18]*

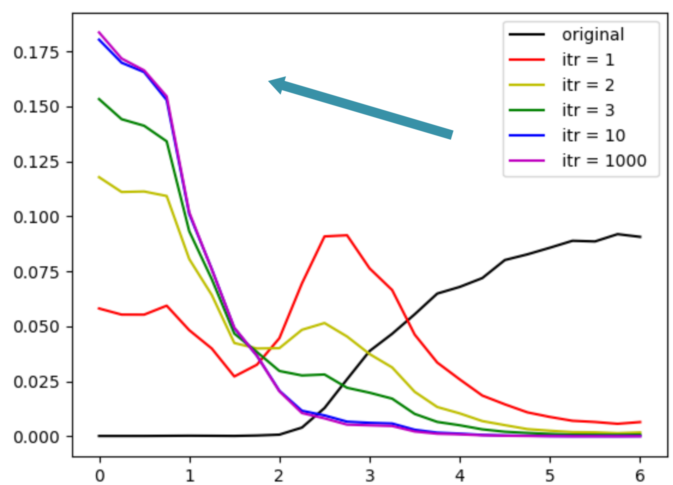
- One is the RBM trained by configs at only low temps.
- The other is the RBM learning various temps $T = 0, 0.25, \dots, 6$ (and H=0).

large scale

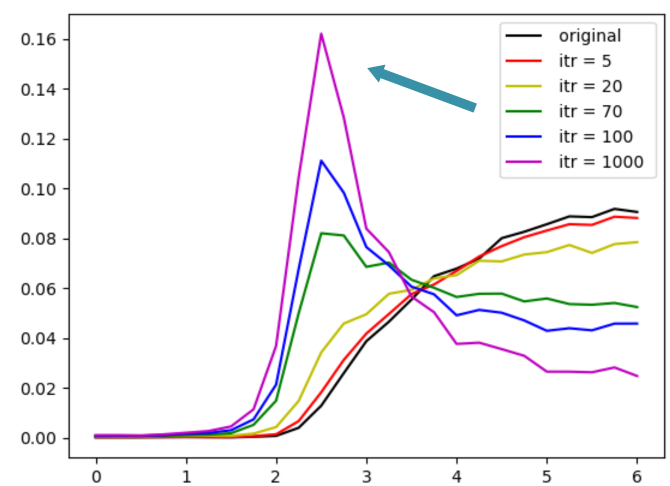
low temp



high temp



RBM learning only T=0



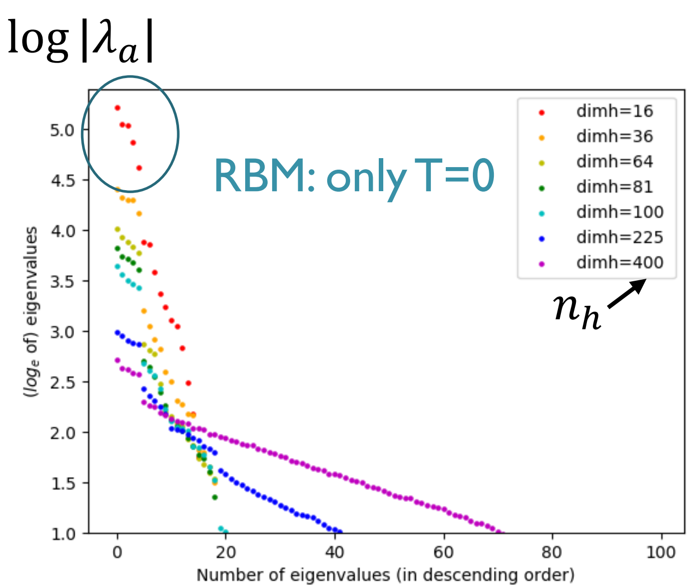
RBM learning T=0, …, 6

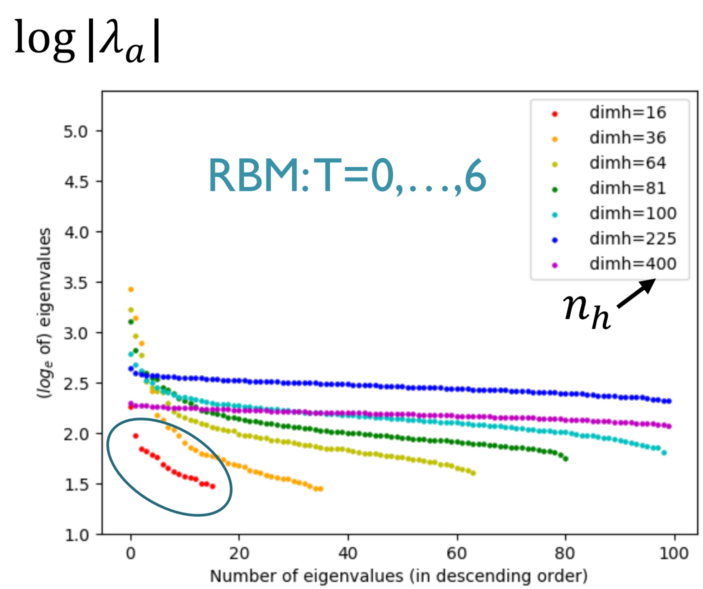
$$ww^T u_a = \lambda_a u_a$$

- Eigenvalues of weights $\sum_a w_{ia} w_{ja}$

- If the RBM learns configs at only low temps,
  only a few (~5) eigenvalues are especially large.

- If the RBM learns configs at $T = 0, 0.25, \ldots, 6$ (including high temp)
  all the eigenvalues have similar values.
  This may be because many hidden neurons are needed to learn configs at various temps (=various scales).

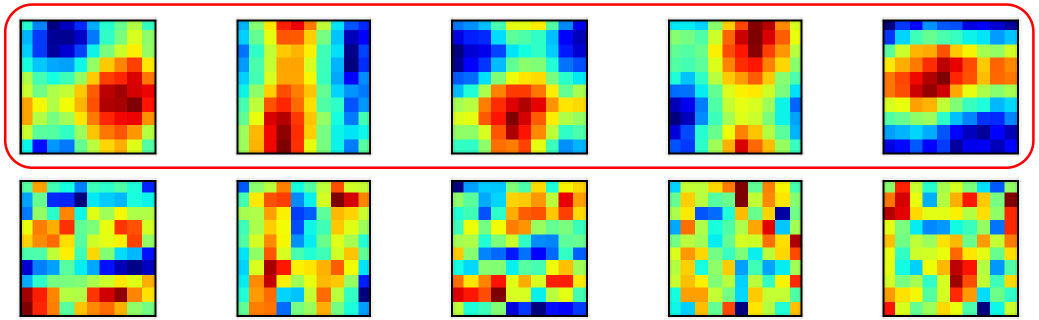




Eigenvectors of ww^T

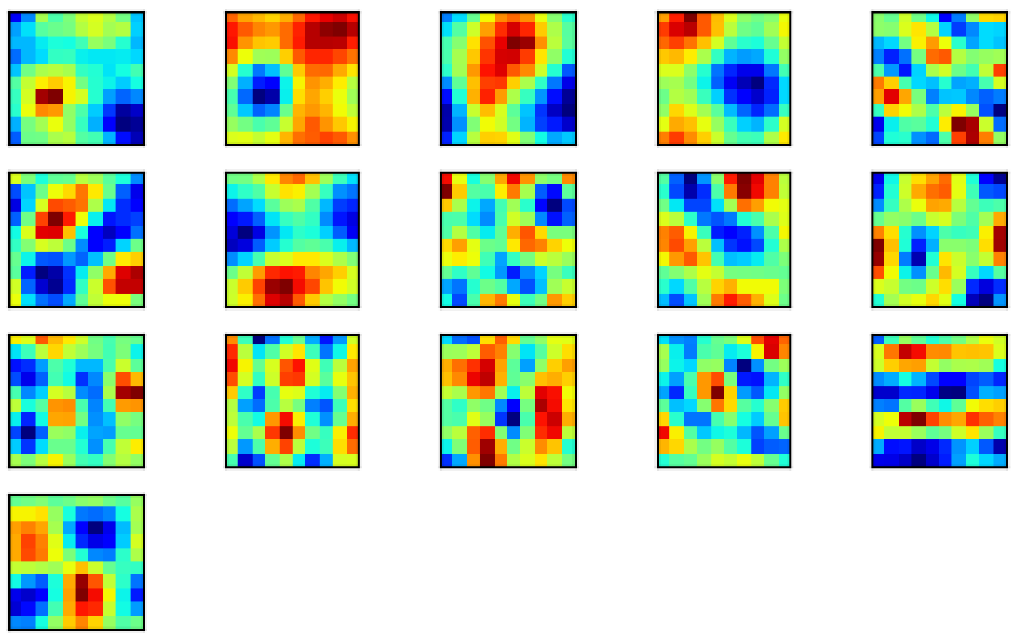
$$ww^T u_a = \lambda_a u_a$$

- RBM learning only low temps $(T = 0, \ldots, 2, n_h = 16)$



Configs with large scale have large eigenvalues.

- RBM learning various temps $(T = 0, \ldots, 6, n_h = 16)$



Configs with various scales have similar eigenvalues!

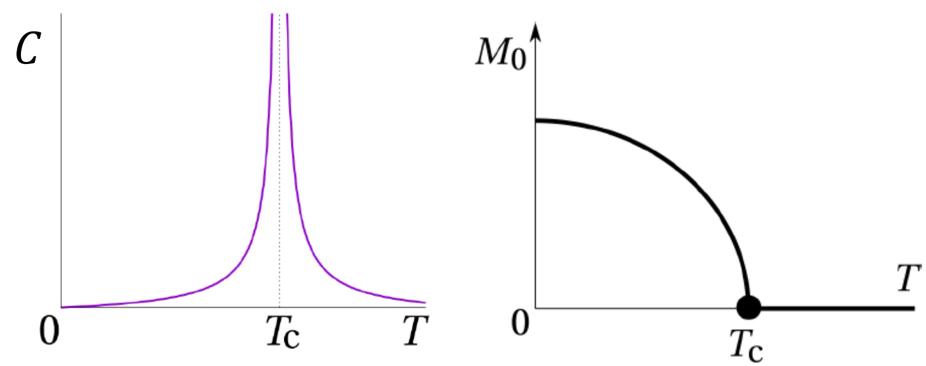
All of them appears in reconstructed images.

Scale invariance…?

General property of fixed points?

- Scale invariance is not a unique choice for "feature".
- Configs around $T = T_c$ also show the critical behavior of thermodynamic quantities (specific heat, magnetization, …).

2nd order phase transition (at H=0)



- Let us study the relation of RBM flow and thermodynamics. This gives us further understanding on its behavior.
- Then RBM learning $H \neq 0$ configs show a clear difference from RG flow and a close relation to thermodynamics.

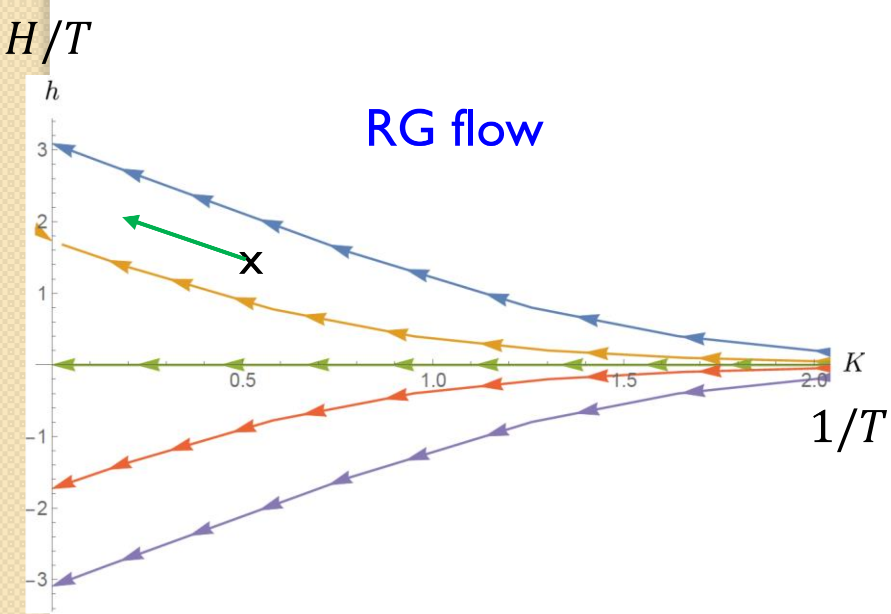
- In 1d Ising model, we can obtain the exact solutions of RG flow and thermodynamic quantities (even in $H \neq 0$).

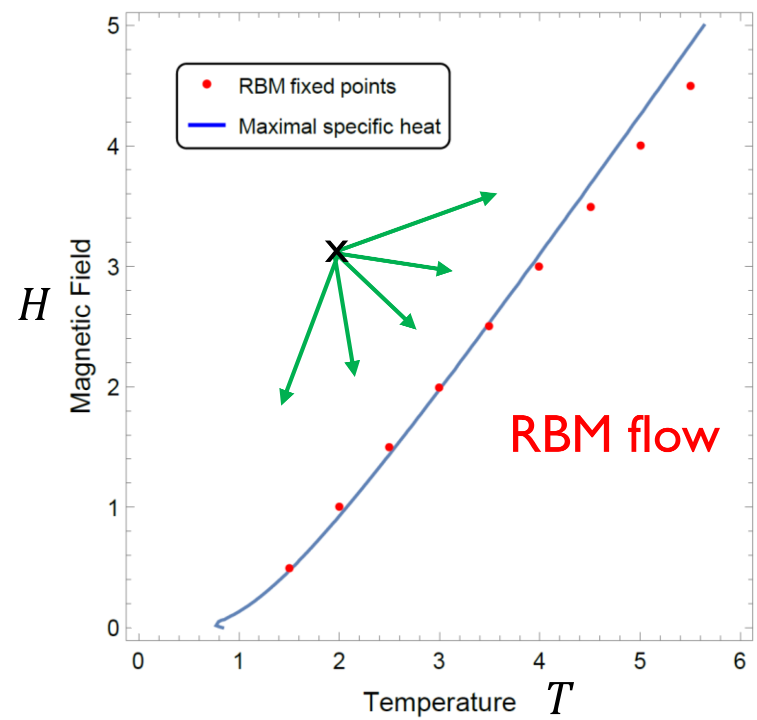
- RG flow goes in a unique direction. The fixed points are at $(T, H) = (0,0), (0, \infty), (\infty, *)$.

[SSF-Giataganas, '18]

- RBM flow approaches to aligned points, so the direction is *not* unique. These fixed points can be fit very well by (local) maximal specific heat.

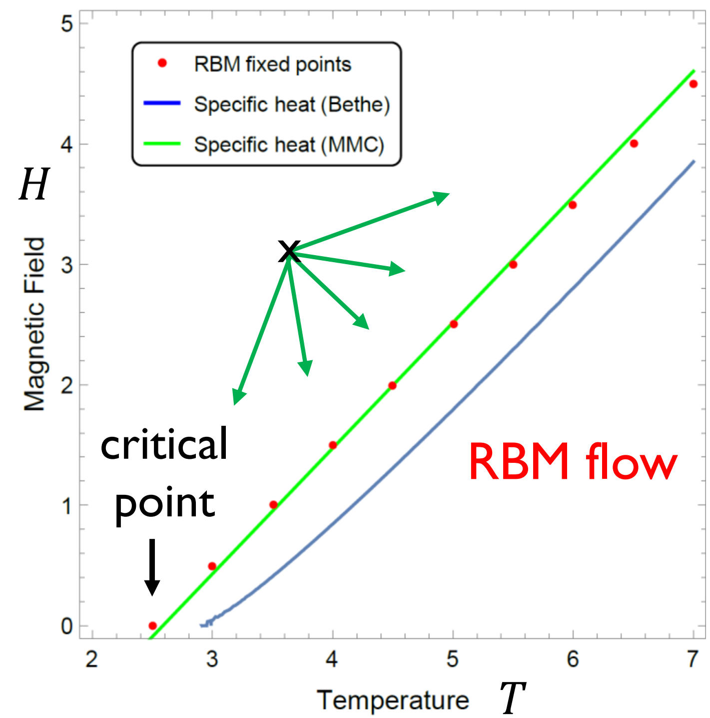
completely different!





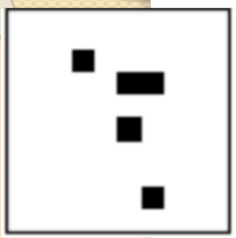
- In 2d model we cannot get the exact solutions in $H \neq 0$. Then we use the Bethe approximation for analytic calculation and numerical calculation using Monte Carlo simulation.

- RG flow has the critical fixed point only at H=0.
- RBM flow behaves similarly to 1d case: *[SSF-Giataganas, '18]* It approaches the aligned points (=fixed points).
- The fixed points are coincident *again* with local maximum of specific heat $(\partial C/\partial T)_H = 0$.
- It includes the critical point, so the maximal specific heat is its suitable generalization.
- This should be the "feature".



more general proposal

RBM can learn thermodynamics?



- It seems strange because the specific heat $C = \partial E / \partial T$ cannot be directly measured in the input configurations.

- If the RBM can, it must use wisely the "energy" function

$$E(\{v_i\},\{h_a\}) = \sum_{i,a} v_i w_{ia} h_a + \sum_a b_a h_a + \sum_i c_i v_i$$

$$p(\{h_a\}) = \sum_{\{v_i\}} \frac{e^{-E(\{v_i\},\{h_a\})}}{\mathcal{Z}}$$

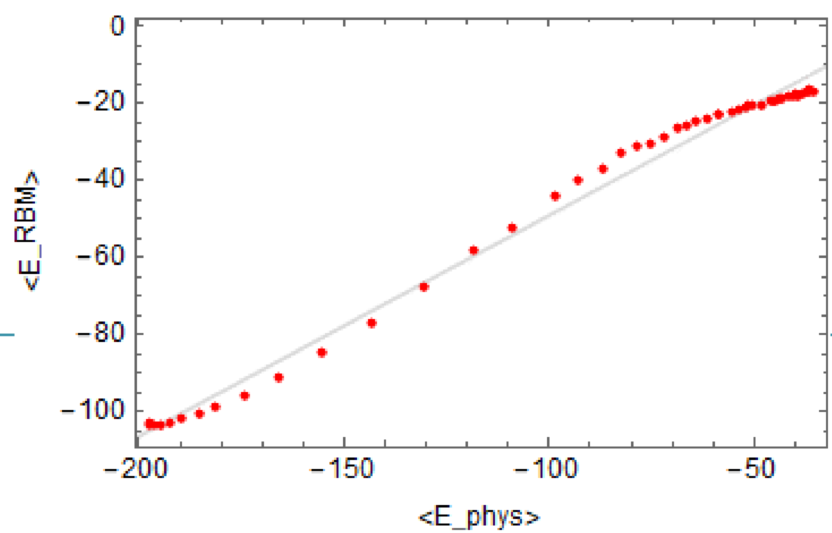
function of weights and biases in Boltzmann distribution

- This "RBM energy" seems correlated with physical energy, but *not* coincident.

Hamiltonian

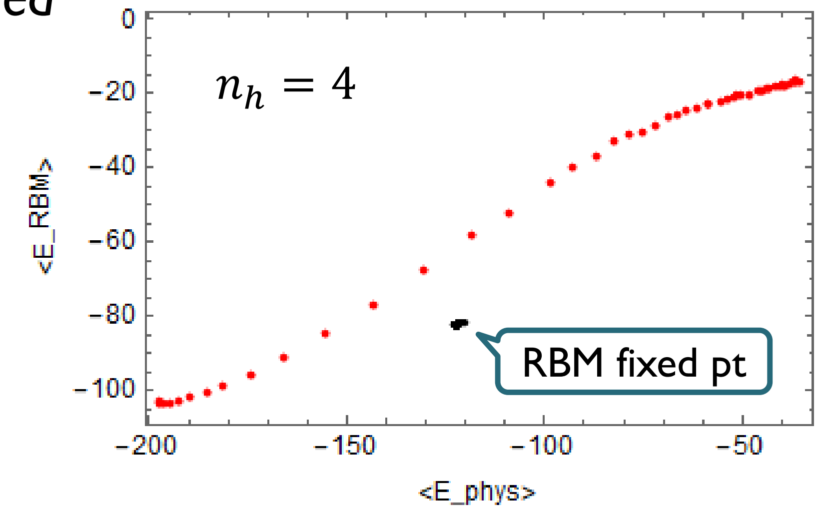
$p \sim e^{-E/T_c}$

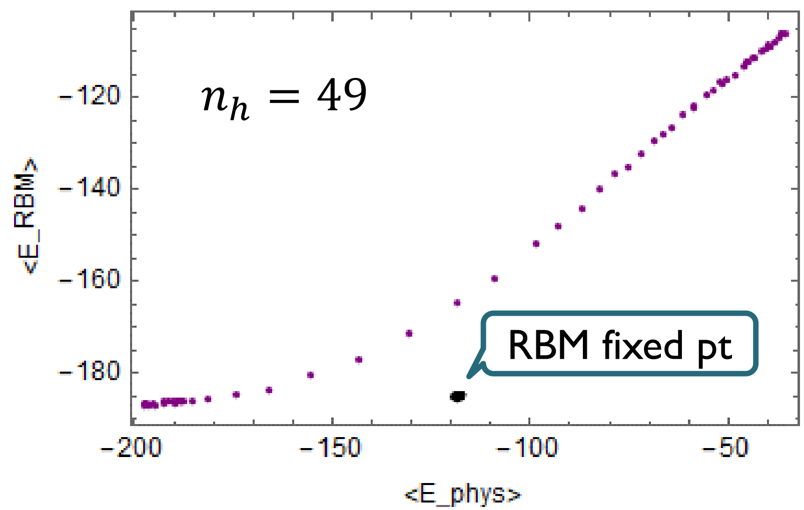
- Data: 2d configs at $T = T_c, H = 0$.
- RBM: $n_v = 100, n_h = 4$



- The relation of RBM energy and physical energy should be clarified to understand the feature extraction further.

- This relation is easily *changed* by details of training. (e.g., n_h, learning epoch, …)
- Nevertheless, RBM flows have the *same* fixed points in (T, H) space and with *similar* physical energy, corresponding to "feature".
- So far I cannot grasp what is essential here. This is an important future work.





Summary & future directions

- We perform a machine learning of the RBM to extract the features of spin configurations in Ising model.
- We find that the flow of reconstructed images by RBM has the fixed points (="features") just as the RG flow, but their behaviors are obviously different.
- We propose that the features the RBM grasps should be scale invariance and maximal specific heat.
- How the RBM learns thermodynamics must be clarified.
- Relation of "RBM energy" and physical energy is an important subject of future works.
- "RBM flow" may be related to the way of human recognition…