機械学習が抽出する特徴とイジング模型の相転移点

船井 正太郎 （株式会社アラヤ）

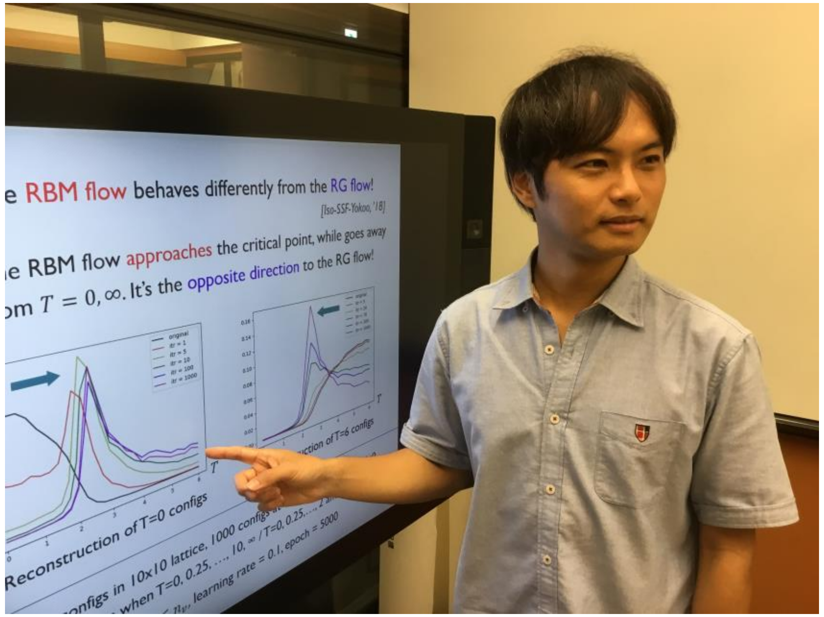
2022年10月27日 @「ディープラーニングと物理学 オンライン」

「初めまして」の方も多いと思いますので…

名前：船井 正太郎

2017年まで　柴 正太郎

論文では　Shotaro Shiba Funai



研究分野：

2015年まで　物理学の超弦理論の研究

2016年から　機械学習の研究

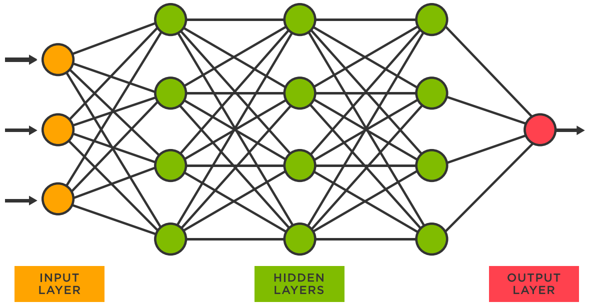
職歴：

学振研究員(PD)、益川塾とKEKの博士研究員、OISTの研究員などを経て、

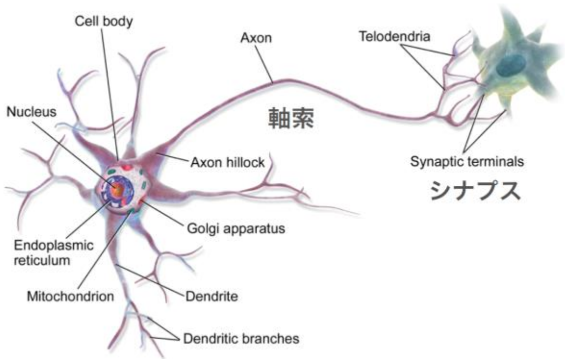
2022年10月から　株式会社アラヤのチーフリサーチャー

最も根底にある興味は…

機械学習 (AI) が数値的に捉えられる、
データの「特徴」とは結局何なのか？
可能ならば物理学の概念を応用して
理解したい！

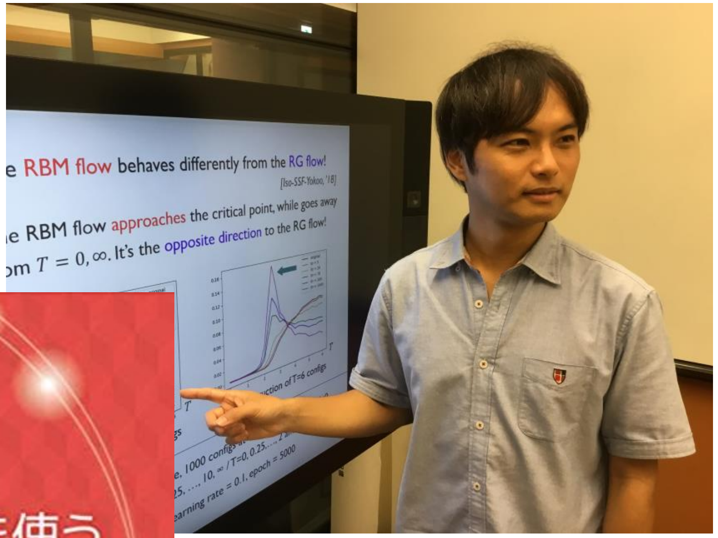


https://www.tibco.com/



橋本幸士さん編

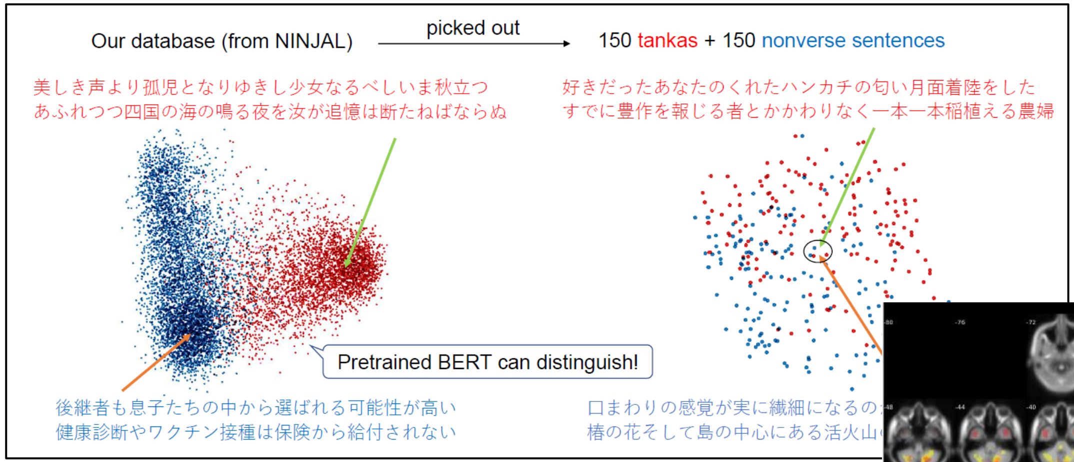
船井正太郎（第10章）

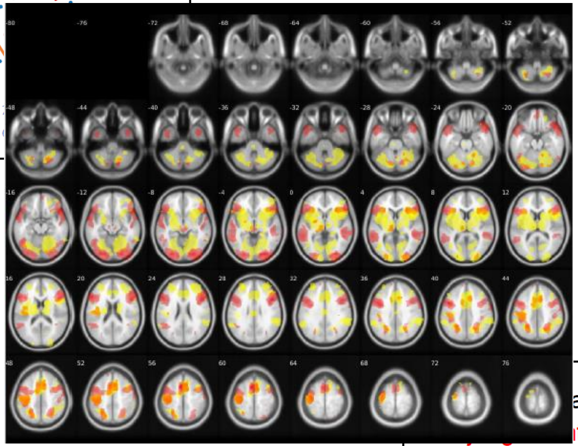


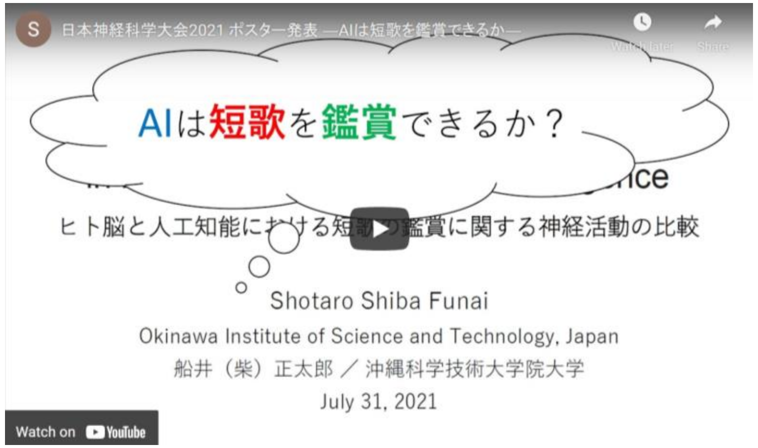
もし機械学習が本当に脳や意識の toy model と見做せるならば、
「意識の物理学」が創れるかもしれない…？

研究1：機械学習は短歌を鑑賞できるのか？

（機構間連携におけるプロジェクト、2019年から）

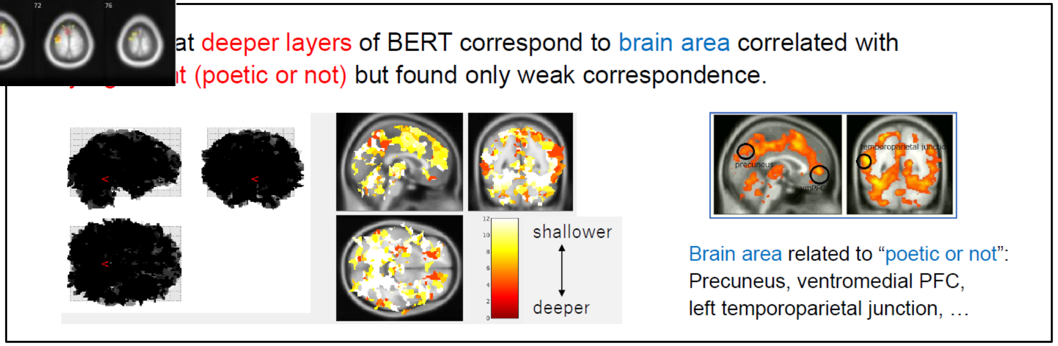






https://youtu.be/wP0RAWF6_h0

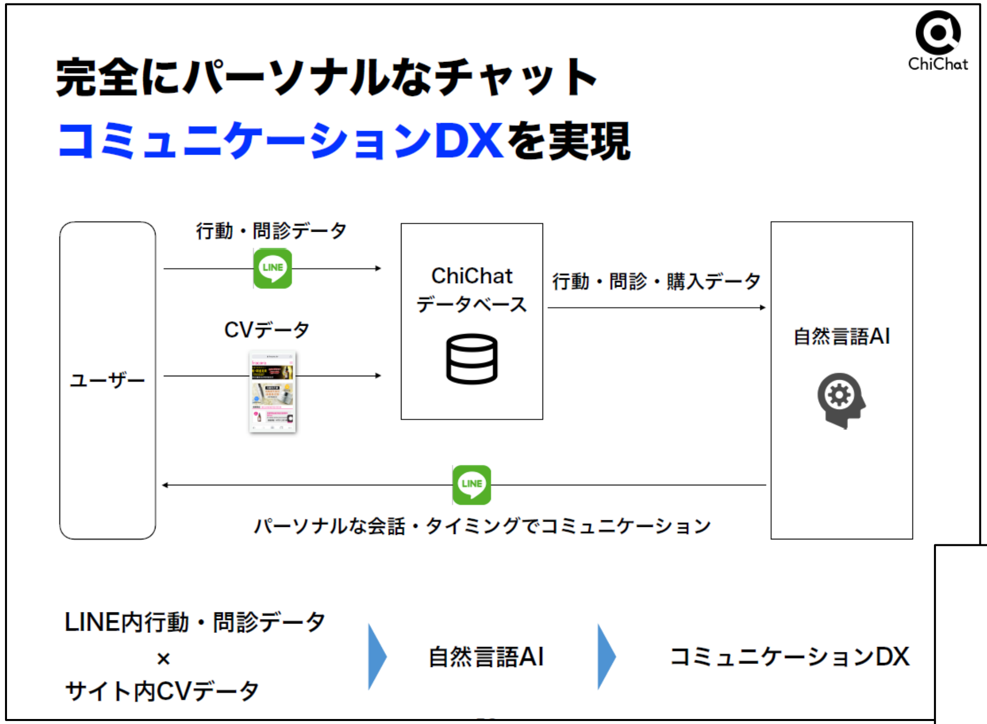
短歌の「特徴」をBERTのベクトルで表現して解析



短歌を読むときの人間の脳をfMRIで計測して、BERTベクトルとの対応（各部位・各層の相関）を解析

研究2：チャットコマースにおける文章・画像の自動生成

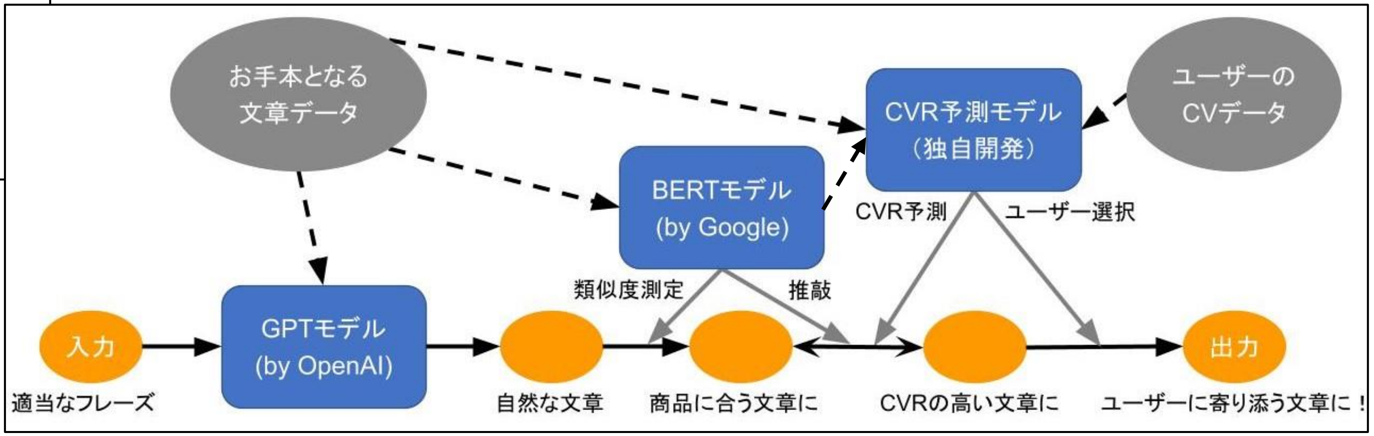
（株式会社人々との共同研究、2020年から）



CEO石川さんの目標は…

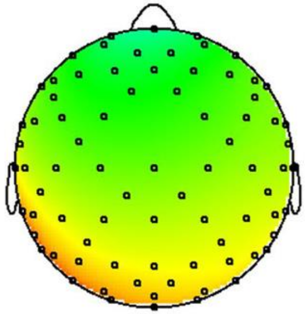
一人ひとりのユーザーが気持ちよく
商品を買えるようなメッセージを
生成できる、ユーザーに寄り添える
AIシステムを作りたい！

そうしたシステムを開発すべく、BERTやViTなどが
捉える文章や画像の「特徴」を活用しています。



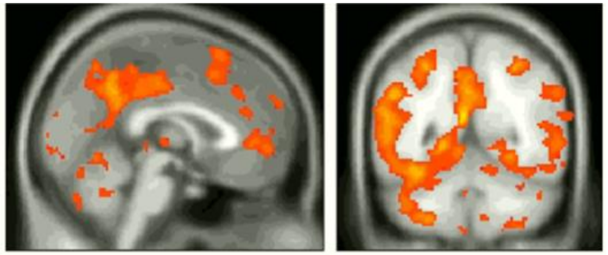
研究3：脳活動を統合的に理解する研究（脳活動データ間の変換）

（株式会社アラヤでの研究、ムーンショットプロジェクト、2022年から）



脳波測定 (EEG) のデータ：
時間分解能は良いが、空間分解能は悪い。
脳の深いところはよく見えない。

fMRI測定のデータ：
空間分解能は良いが、時間分解能は悪い。
脳の深いところまでよく見える。



機械学習で両方の「特徴」を捉えた上で、
相互に変換する技術を目指します。

研究4：学習物理学の研究

（学術変革Aでの研究、2022年から）

物理学で多用される概念の中で、主にトポロジーを使って、機械学習の新しい側面を見出していきたいと思っています。

その土台となるかもしれない研究として…

MLPhYs 学術変革領域研究(A) 学習物理学の創成
Foundation of "Machine Learning Physics"

CONTACT | Members only | En | Jp

領域概要 | 研究組織 | イベント | 成果 | アウトリーチ

計画研究 B03

機械学習への位相幾何学的アプローチ

研究代表者	福嶋 健二	東京大学 理学系研究科
	船井 正太郎（柴 正太郎）	沖縄科学技術大学院大学 物理生物学ユニット
	塩崎 謙	京都大学 基礎物理学研究所
研究分担者	三角 樹弘	近畿大学 理工学部

お誘いいただき
有難うございます！

つまり…「特徴」って何だろうか？

今日のテーマ：イジング模型におけるスピン配位の特徴抽出

- 物理学の概念（特に繰り込み）を使って「特徴」が議論できそう！
- イジング模型のハミルトニアン（1次元または2次元の正方格子を考える）

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

- 様々な温度におけるスピン配位（他のパラメータは固定: $J = 1, H = 0$; 白: $s_i = 1$, 黒: $s_i = -1$）

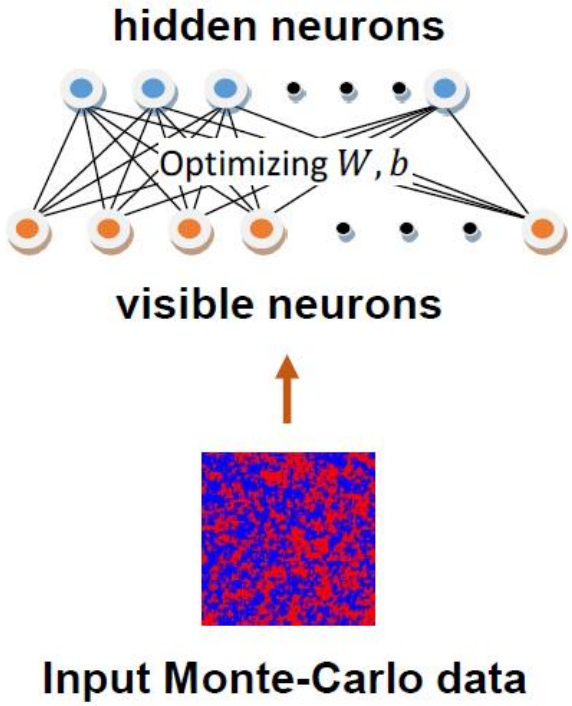
T=0.0　2.0　4.0　6.0　8.0　…

相転移点 $T = T_c = 2.27$

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

➢ 入力データ（不思議に感じるかもしれませんが…）

- スピン配位を Metropolis Monte Carlo シミュレーションで生成する。
- 入力データには、一定の間隔をもつ様々な温度（や外場 H）のスピン配位が同じ数だけ含まれるようにする。例えば、$T = 0, 0.1, 0.2, \ldots, 9.9$（100個の温度）を1000枚ずつ。
- このデータセットは物理系としては不自然なものだが、入力データに画像として様々なパターンが含まれるように選んだ。
- こうして入力データが決まれば、スピン配位の確率分布が定義できて、計算ができる。



➢ 機械学習には制限ボルツマンマシン (RBM) を使う

- 入力データと同じ確率分布でデータを出力するように学習をする、ニューラルネットワークである。
- 学習の過程で、RBM は入力データの「特徴」を抽出する。

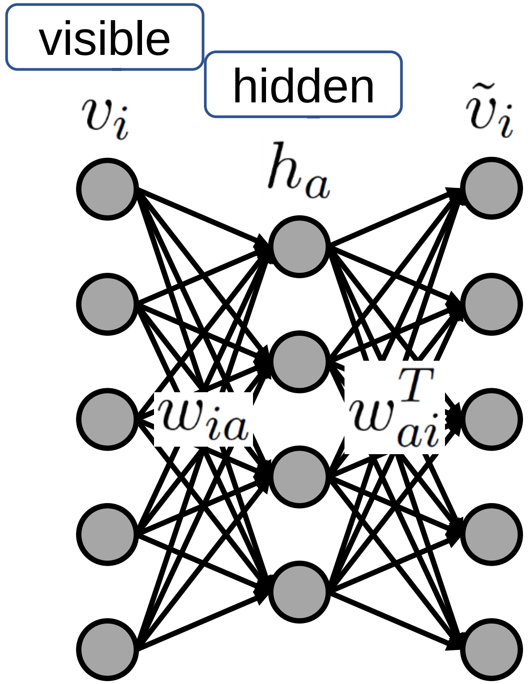
- 出力データの確率分布を、ニューラルネットワークの言葉で定義する必要があるが、RBM では visible, hidden ニューロンの値を用いて「エネルギー」関数を定義した上で

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

weights w_{ia}, bias b_a, c_i

Boltzmann 分布によって確率分布を定義する。

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



- RBM を訓練する（＝weightsとbiasを最適化する）ときには、入力データと出力データの確率分布の KL divergence が極小 (local minimum) に近づくようにする。

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

- この KL divergence は2つの確率分布の「距離」のようなものである。

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

入力が v_i である確率 / 出力が v_i である確率

- 我々の研究では、入力データはスピン配位（白黒）なので $v_i = \pm 1$ に限られる。また、hidden ニューロンの値も binary ($h_a = \pm 1$) のみに限定する。

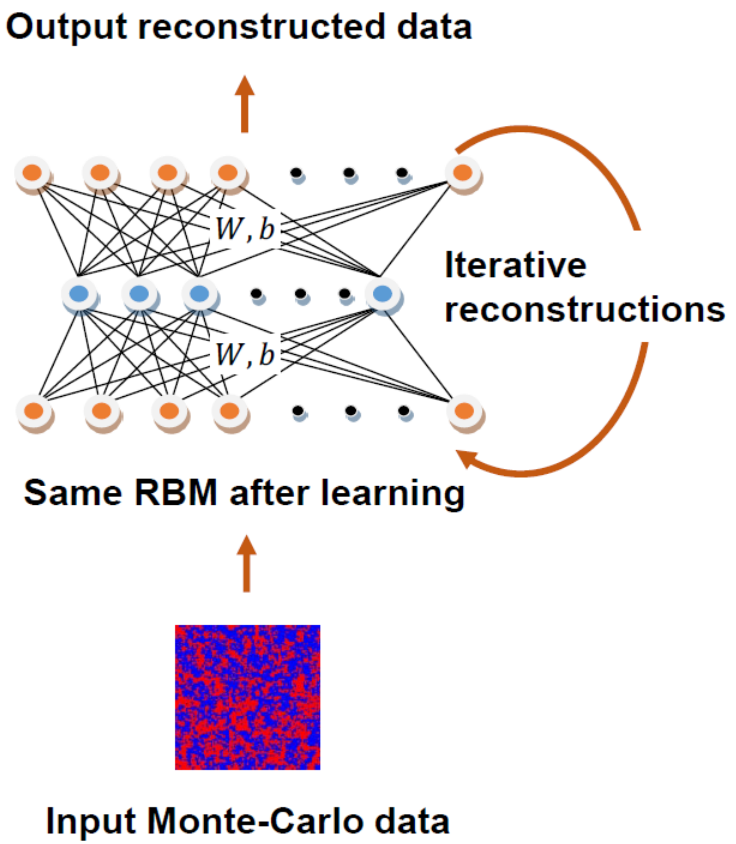
- すると、hidden ニューロンと（再構成したときの）visible ニューロンの期待値は、Boltzmann 分布を用いて計算できる。

$$\langle h_a \rangle = \tanh\left(\sum_i v_i w_{ia} + b_a\right), \qquad \langle \tilde{v}_i \rangle = \tanh\left(\sum_a h_a w^T_{ai} + c_i\right)$$

- 出力（再構成された）データも当然、$\tilde{v}_i = \pm 1$ に限定したい。そのために、配位の各スピンの期待値 $\langle \tilde{v}_i \rangle$ を確率 $(1 \pm \langle \tilde{v}_i \rangle)/2$ で、 $\tilde{v}_i = \pm 1$ に置き換える。

期待値を変えない操作

- 訓練が終わると、入力データの確率分布 $q(\{v_i\})$ と出力データの確率分布 $\tilde{p}(\{v_i\})$ は似ているけれども少し違うものになる（現実的には KL divergence は0にならないため）。
- この出力データを再び入力すると、また少し違う確率分布 $\tilde{\tilde{p}}(\{v_i\})$ の配位データが出力（再構成）される。
- この操作を繰り返し行うと、スピン配位の確率分布の flow が得られる： $q(\{v_i\}) \to \tilde{p}(\{v_i\}) \to \tilde{\tilde{p}}(\{v_i\}) \to \ldots$
- これを RBM flow と名付けた。 *[Iso-SSF-Yokoo, '18]*



➢ いろいろ素朴な疑問が湧いてくると思います

1. この RBM flow はイジング模型の RG（繰り込み群）flow と対応するか？（どちらも特徴的でない情報が欠落する flow のはず）
2. もし RBM flow が固定点を持ったら、それは「特徴」そのものを表現するか？（RBM flow に沿って特徴が強調されるはず）

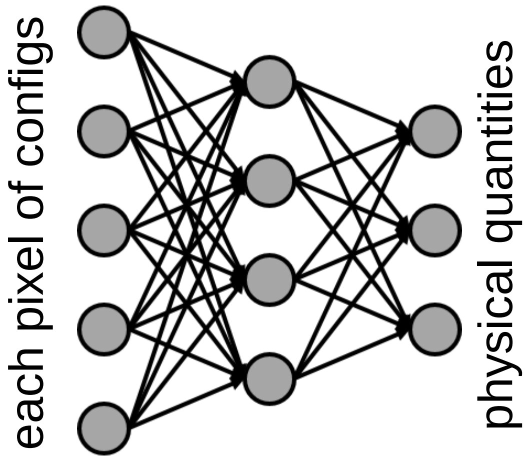
我々の結果を眺めてみよう

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

➢ RBM flow はパラメータ (T, H) の空間に固定点を持つ！

- スピン配位の固定点は存在しない。従って、RBM flow の固定点 (T, H) における新しい配位を作り続けることになる。
- 出力データ（配位）の (T, H) を見積もるために、我々は2つの方法を使った。

1. 別のニューラルネットワークを訓練して、入力データの正しい (T, H) を出力できるようにする。（教師あり学習）
   MMC シミュレーションのパラメータ
2. 外場なし $(H = 0)$ の配位のみを扱う場合は、ハミルトニアンでエネルギーを計算して、温度 T を見積もることができる。
   （1の方法と矛盾しない結果が得られる。）

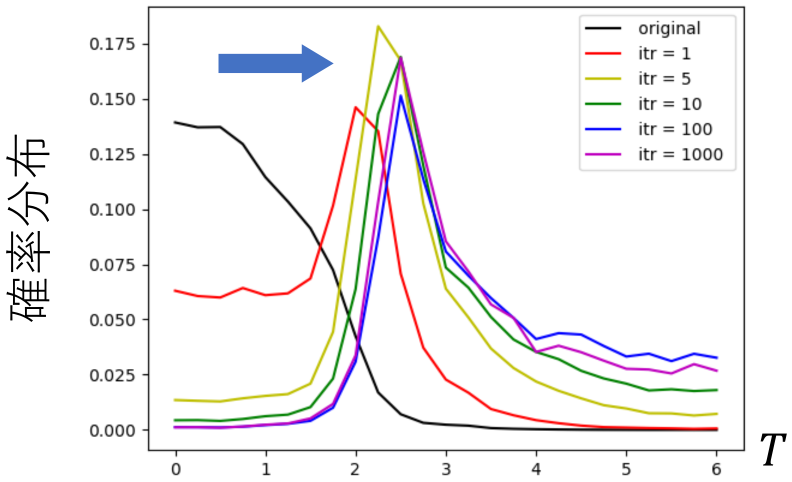


➢ RBM flow と RG flow の振る舞いは異なる！ *[Iso-SSF-Yokoo, '18]*

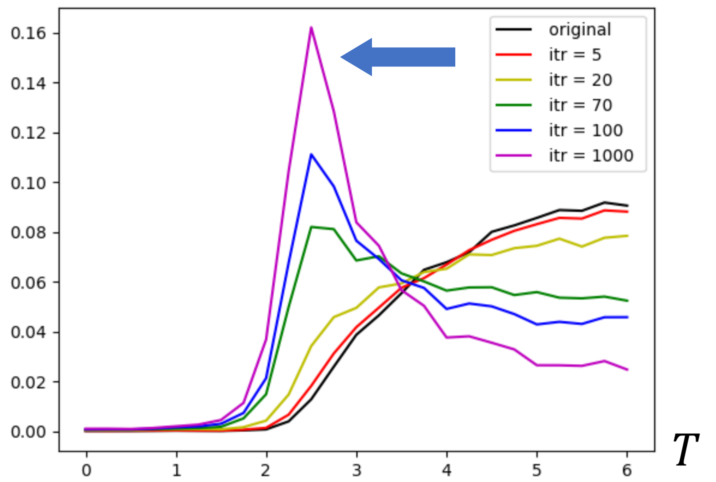
- RBM flow は相転移点 $T = T_c \sim 2.27$ に近づいていき、$T = 0, \infty$ から遠ざかっていく。
  これは RG（繰り込み群）flow とちょうど逆の方向になっている。

2d, H=0

$n_h = 81$



T=0（低温）の配位から再構成した場合



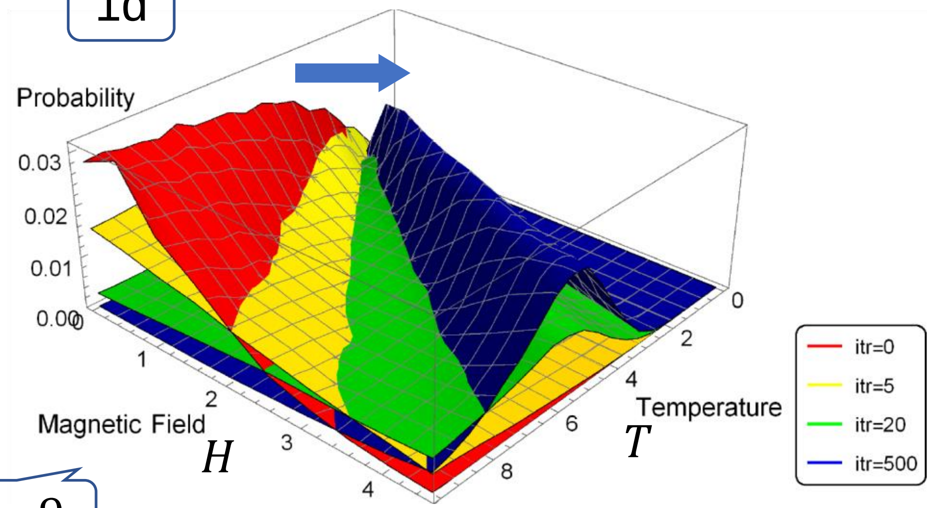
T=6（高温）の配位から再構成した場合

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

➢ 1次元と2次元のスピン配位で外場あり $(H \neq 0)$ の場合に一般化 *[SSF-Giataganas, '18]*

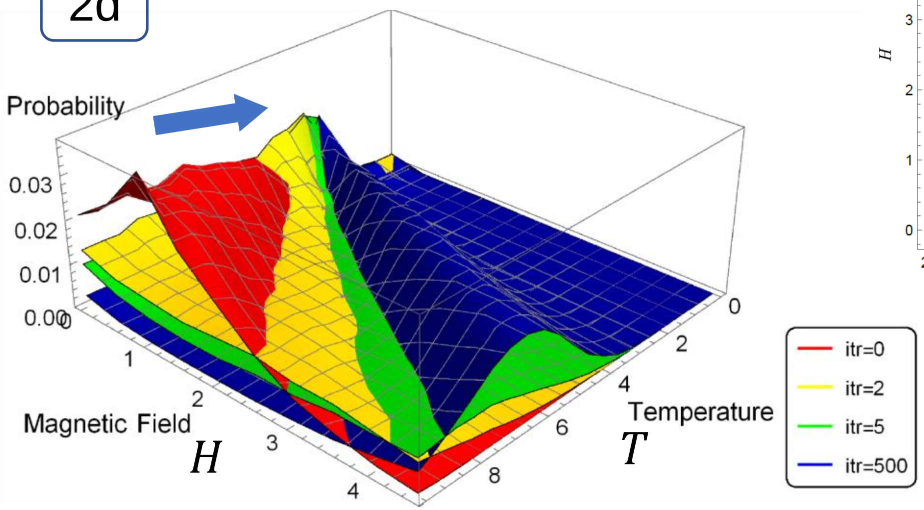
- RBM flow は、パラメータ (T, H) 空間で比熱が極大になる点に近づいていく。
  この flow とその固定点は、やはり RG flow とは異なる。

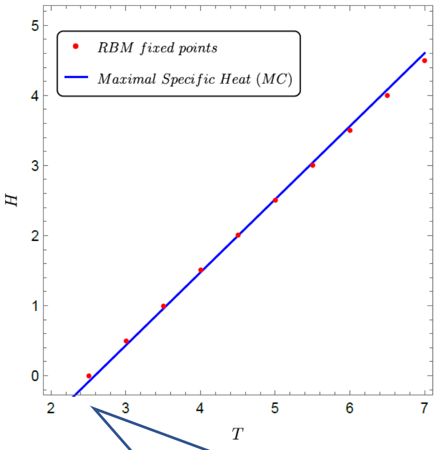
1d



$n_h = 9$

2d





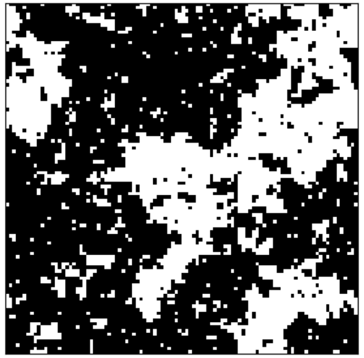
相転移点
$T = T_c, H = 0$
を含んでいる
（一般化）

- 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 hyperparameters: $n_v = 100, n_h \leq 16,$ learning rate = 0.001, epoch = 10000

面白い結果には見えるけれども…

- 理由がはっきりしない：スケール不変性と関係があるのか？
- 条件がはっきりしない：様々なパラメータ依存性を調べるべき。

なぜ、いつ、RBM flow の固定点は $T = T_c$ に現れるのだろうか？（$H = 0$ の場合）



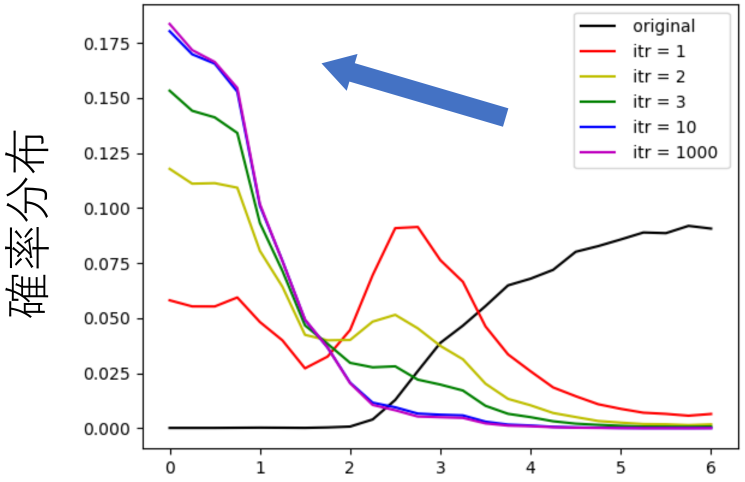
相転移点における
スピン配位は
スケール不変性をもつ

なぜ？：スケール不変性をもつ証拠

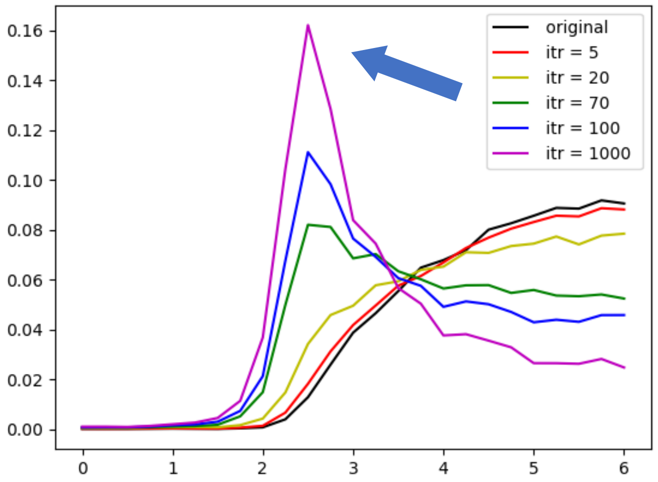
➢ 次の2つの RBM を作って、RBM flow と weight 行列を解析しよう。
そして、比較してみよう。 *[Iso-SSF-Yokoo, '18]*

- 低温の（大きなスケールをもつ）配位のみで訓練した RBM
- 幅広い温度 $T = 0, 0.25, \ldots, 6$ の配位で訓練した RBM（先ほどと同じ）

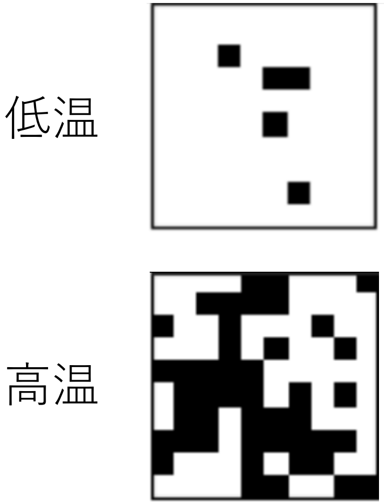
2d, H=0



T=0の配位のみを学習した RBM



T=0, …, 6 の配位を学習した RBM



➢ Weight 行列 $\sum_a w_{ia} w_{ja}$ の固有値

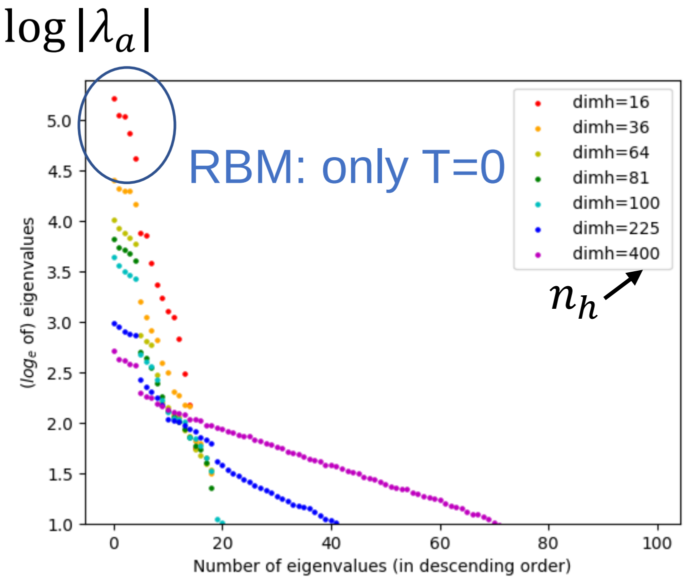
Hidden ニューロンの基底に依存しない積を取った

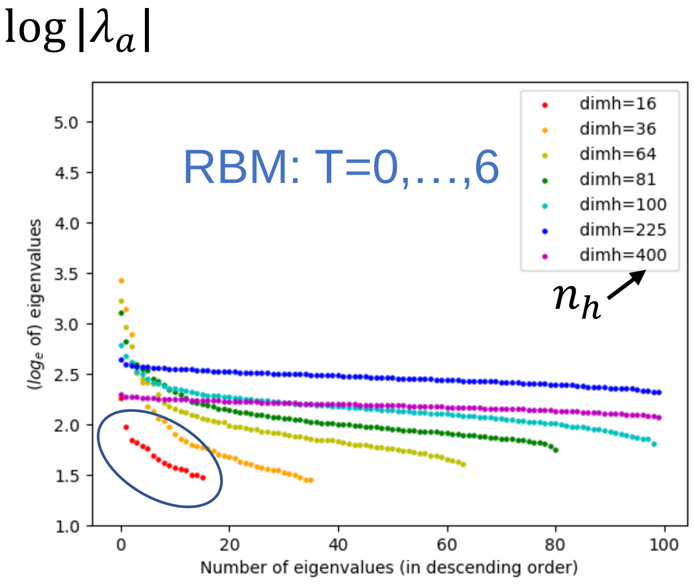
- RBM が低温の配位のみ学習した場合、少数（5個）の固有値だけが特に大きくなる。

$$\boxed{ww^T u_a = \lambda_a u_a}$$

- RBM が様々な温度 $T = 0, 0.25, \dots, 6$（高温を含む）の配位を学習した場合、すべての固有値が似たような値をもつ。

様々な温度（＝様々なスケール）の配位を学習するために、多くの hidden ニューロンが必要になるのだろう。

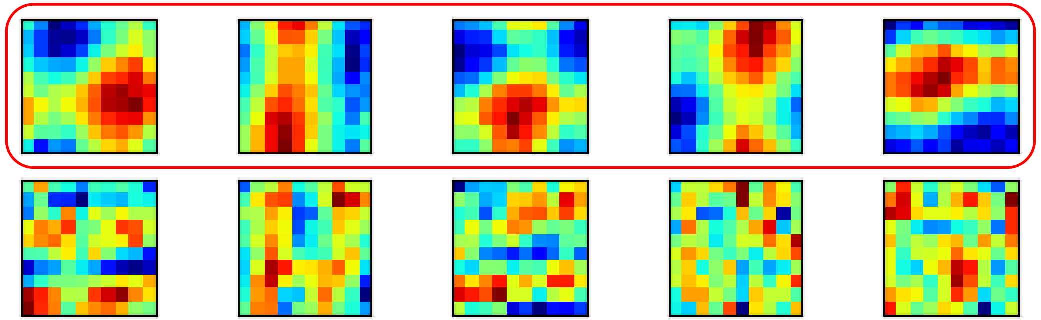




➢ Weight 行列 ww^T の固有ベクトル

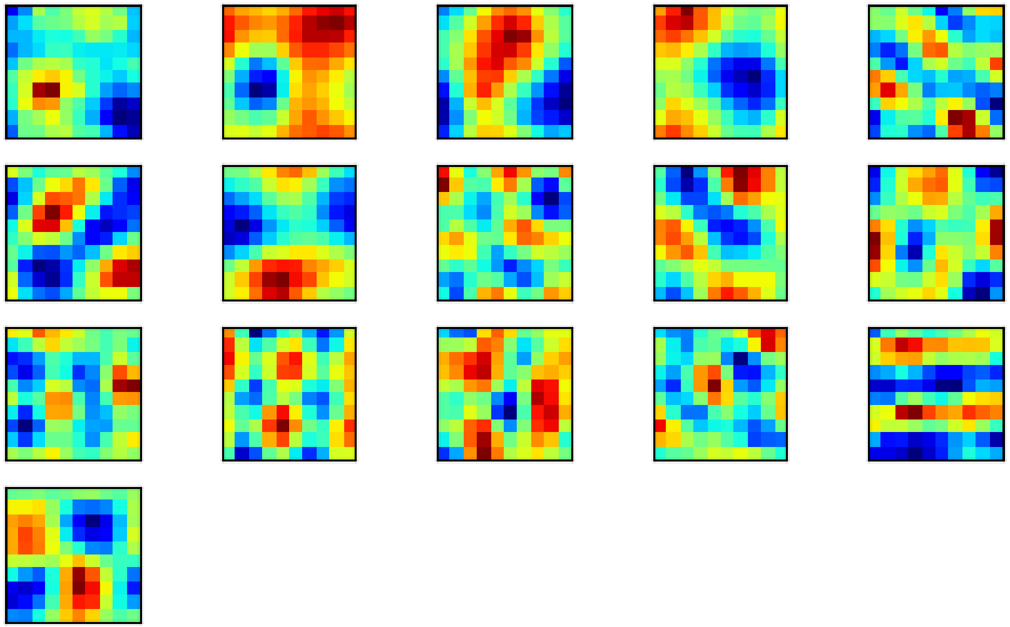
$$\boxed{ww^T u_a = \lambda_a u_a}$$

- 低温の配位のみを学習した RBM ($T = 0, \dots, 2, n_h = 16$)



大きなスケールをもつ配位が
大きな固有値をもつ。

- 様々な温度の配位を学習した RBM ($T = 0, \dots, 6, n_h = 16$)



様々なスケールをもつ配位が
似たような固有値をもつ！

⬇

それらすべての配位が出力
（再構成）データに現れる

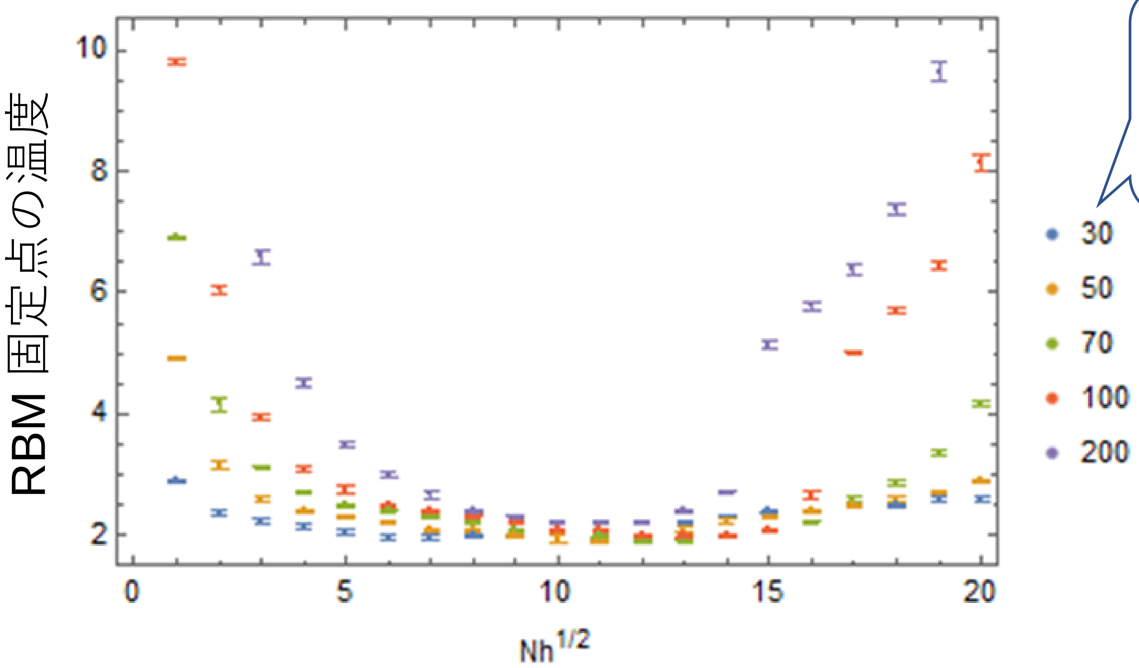
⬇

スケール不変性？

いつ？：RBM 固定点が $T \sim T_c$ に現れる条件

[SSF, '21]

Nv=20^2（visible ニューロンの数＝配位のサイズ）



N_{temp}: 学習データに含まれる温度の数
$T = 0, 0.1, \dots, 0.1 \times (N_{temp} - 1)$ の
配位が同数ずつ（外場なし）。

2d, H=0

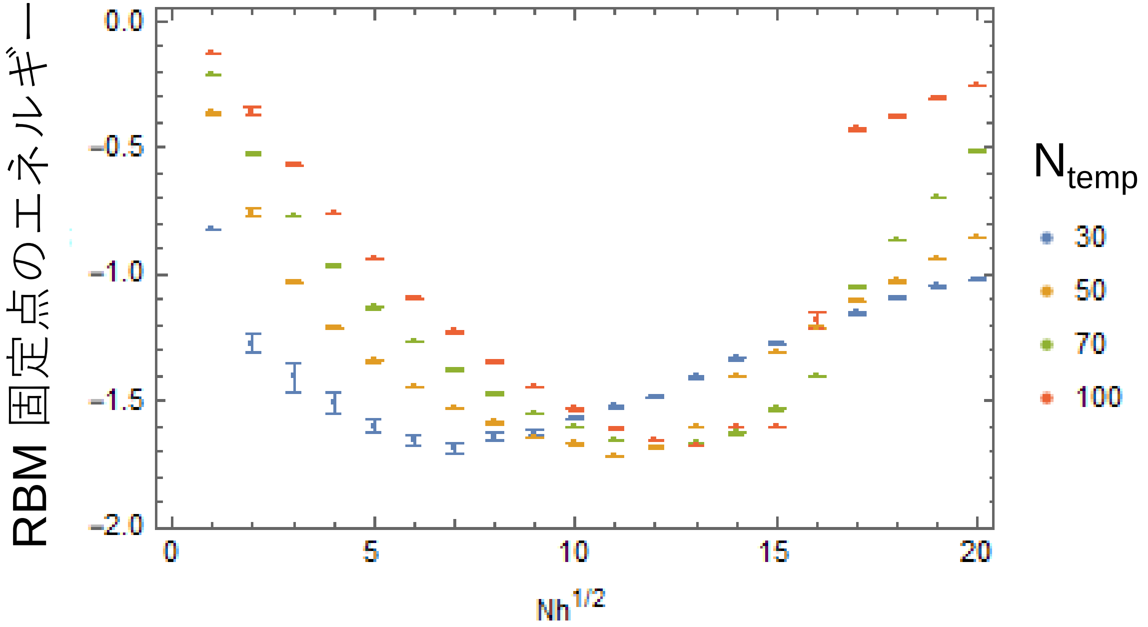
Nh: hidden ニューロンの数

Nh/Nv ~ 1/4 の辺りで
RBM 固定点の温度が最低になり、
その温度が $T \sim T_c$ になるようだ。

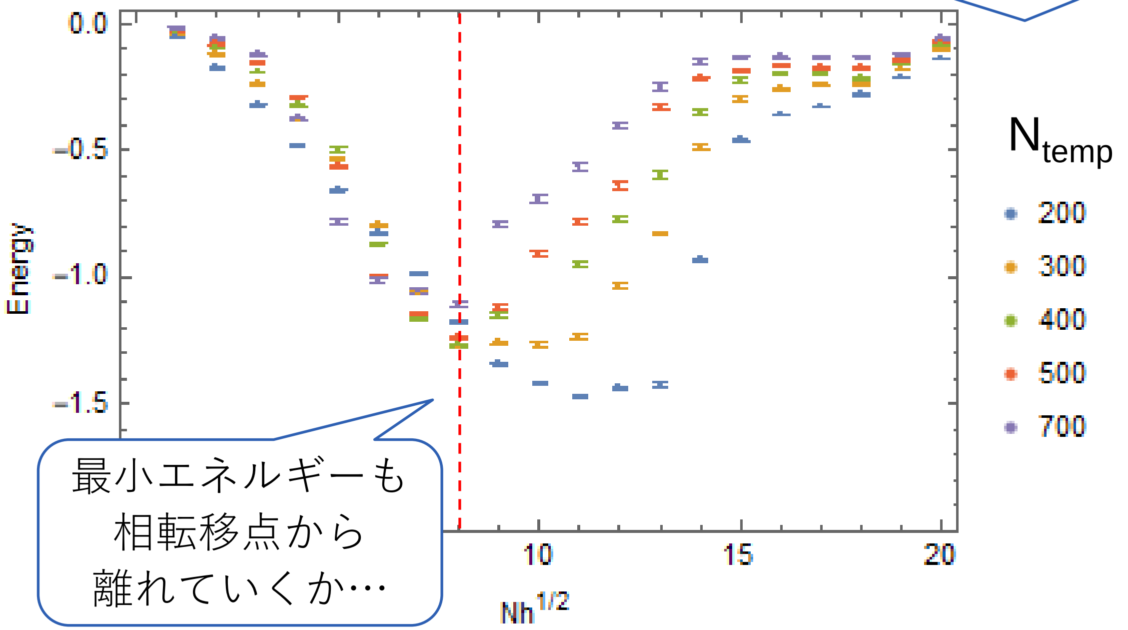
温度の代わりにエネルギーを使って、さらに詳しく結果を見よう。

（エネルギーは $T \sim T_c$ で大きく変化するため）

Nv=20^2



Nv=20^2



学習データの大部分が
超高温（ほぼランダム）だと…

最小エネルギーも
相転移点から
離れていくか…

RBM 固定点のエネルギーが最小となる Nh は、N_{temp}（温度の範囲）に依存して変化する。

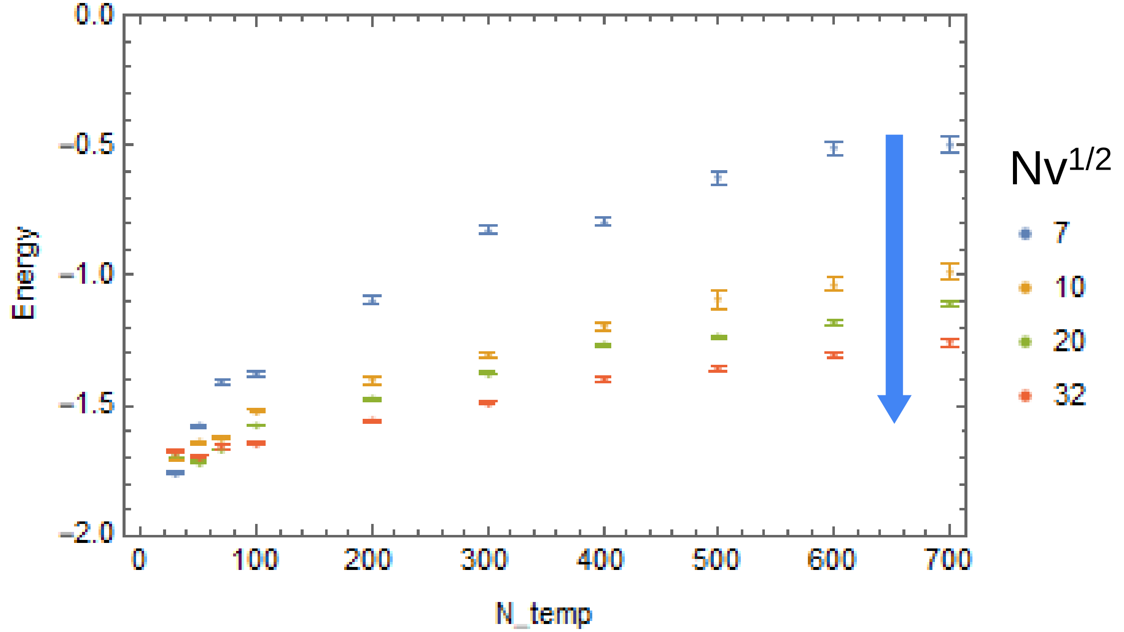
特に N_{temp} が大きくなると、エネルギーが最小となる Nh は Nh/Nv ~ 0.4^2 に収束するようだ。

ここから先は、エネルギーが “最小” となる Nh=(integer)^2 のみに注目しよう。

スピン配位のサイズが大きくなれば、N_{temp} が大きくなっても、RBM 固定点の最小エネルギーは相転移点から離れにくくなる（よりゆっくり上昇する）。

RBM 固定点の最小エネルギー

Nh fixed



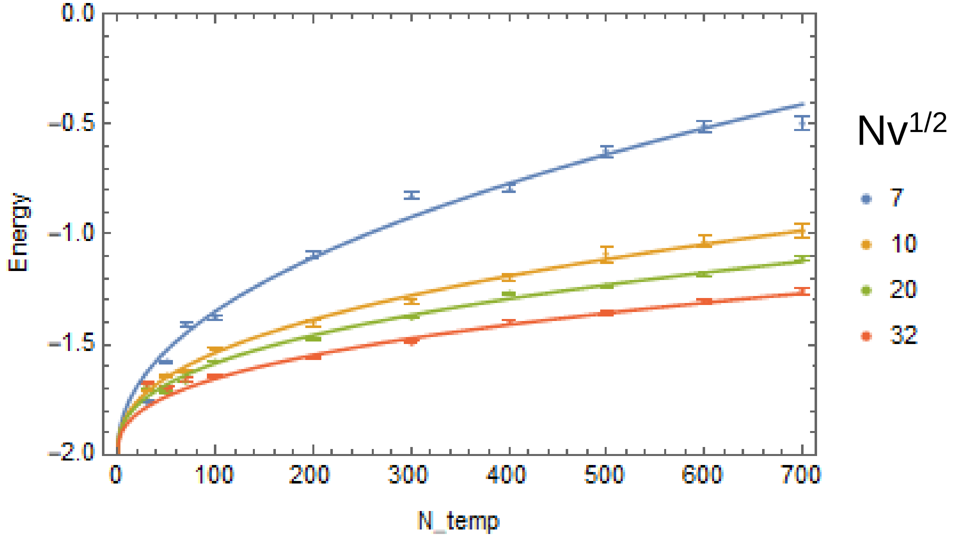
Nv: visible ニューロンの数＝配位のサイズ
N_{temp}: 学習データに含まれる配位の温度は
$T = 0, 0.1, \dots, 0.1 \times (N_{temp} - 1)$

このプロットから推測できることは…

✓ $N_{temp} \to \infty$, Nv: fixed の場合：
RBM 固定点は $E \sim 0, T \to \infty$ になる。

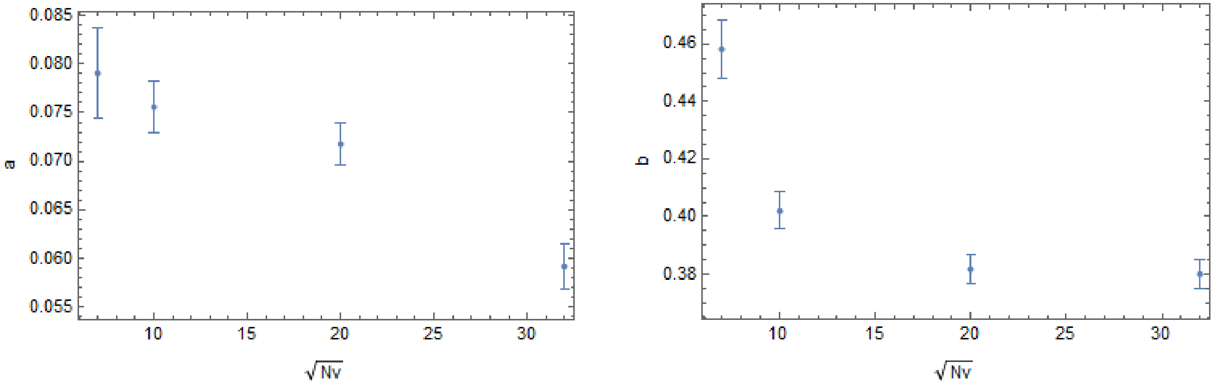
✓ N_{temp}: fixed, Nv $\to \infty$ の場合：
RBM 固定点は低温に向かうが、$T \neq 0$ の有限温度で収束するだろう。
小さい $N_{temp} < 100$ の領域で、固定点は $E \sim -1.7, T \sim T_c$ であることを考えれば、収束する温度は $T \sim T_c$ になるだろう。

参考までに、これらの RBM 固定点を $\boxed{\mathrm{Energy} = -2 + a\, N_{temp}{}^{b}}$ で fitting すると、
かなりきれいに fit できて…

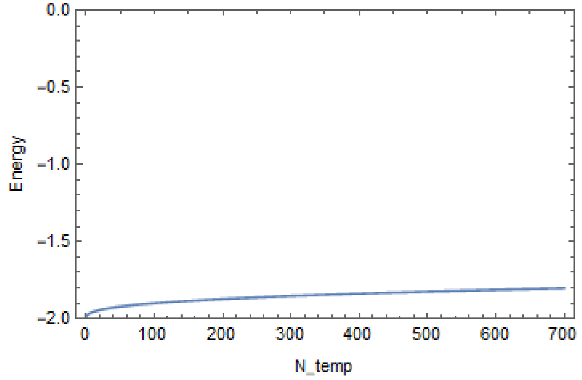


（$N_{temp} \geq 100$ の点のみ fitting に使った）

パラメータ *a* と *b* のプロット



Nv が大きくなると、パラメータ *a, b* はどちらも減少する。
従って、N_{temp} vs. energy のプロットは Nv → ∞ の極限でよりフラットになるだろう。



以上の議論に基づいて、私は次の<u>仮説</u>を提案します！

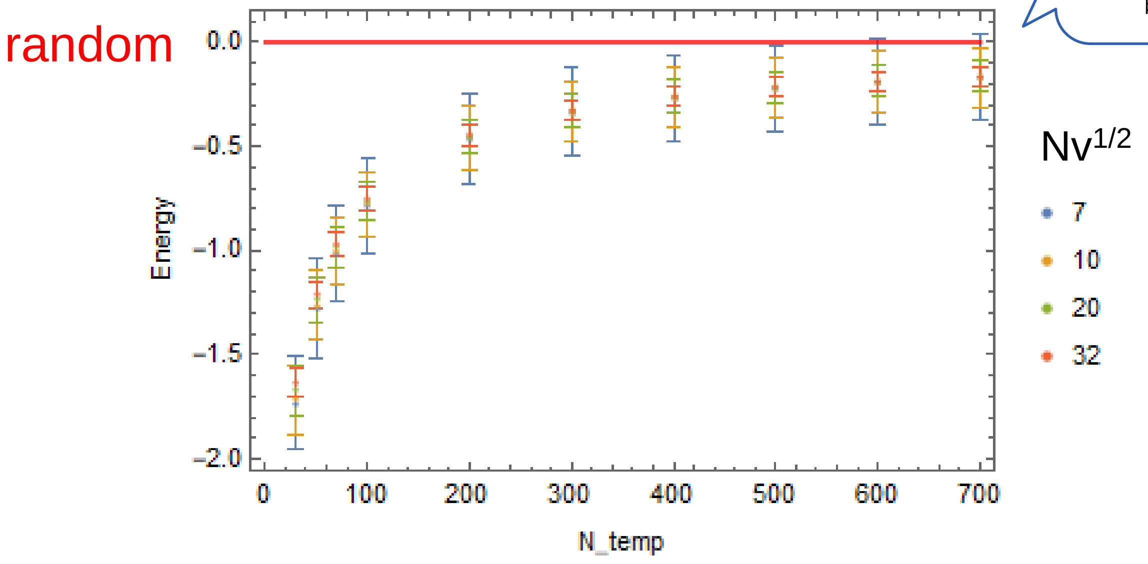
> N_{temp} が大きくても固定されていれば、Nv を充分に大きくすると、RBM 固定点は $T \sim T_c$ に現れるだろう。

Nv を大きくすると RBM 固定点のエネルギーが小さくなる理由は…

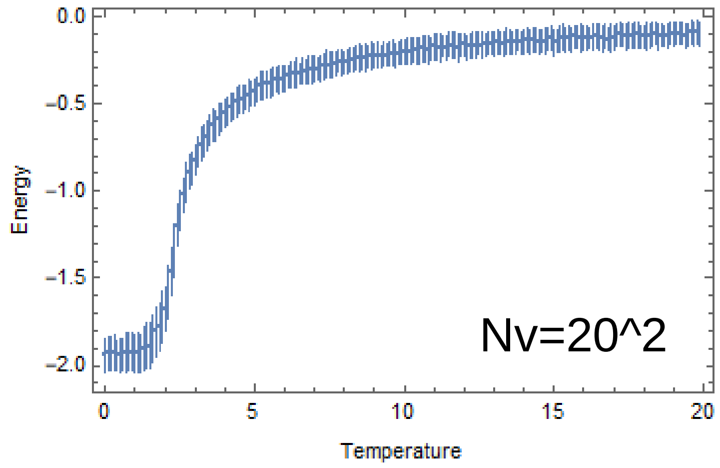
- 学習データに含まれる（完全に）ランダムなスピン配位が少なくなるから。
- ランダムな配位が少なくなれば、RBM はより多くのパターンを学習できる。
- 逆に、もしランダムな配位しかなければ、RBM は何も学習できない。
  （loss function の値を小さくする術がない…）

エラーバー (1σ) を見ると、
Nv が大きくなれば、ランダムな
配位は減ることがわかる。

学習データの平均エネルギー



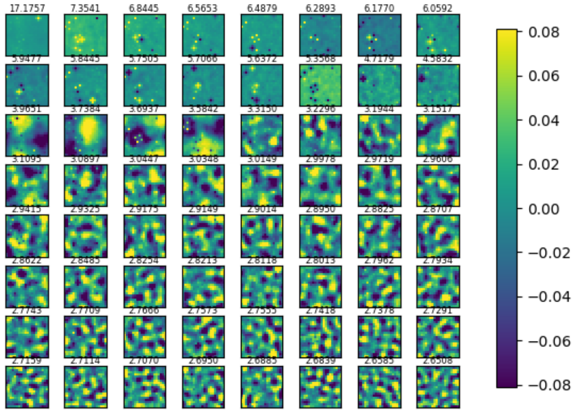
cf. 各温度の配位がもつエネルギー



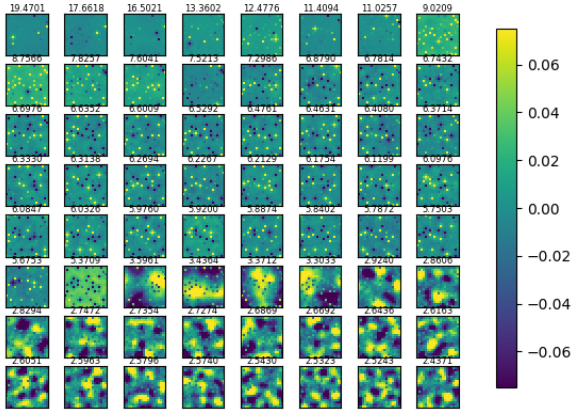
RBM が学習したパターンは weight 行列 ww^T の固有ベクトルの中に見つけられる。

（以下はすべて Nv=20^2, Nh=8^2 の結果）

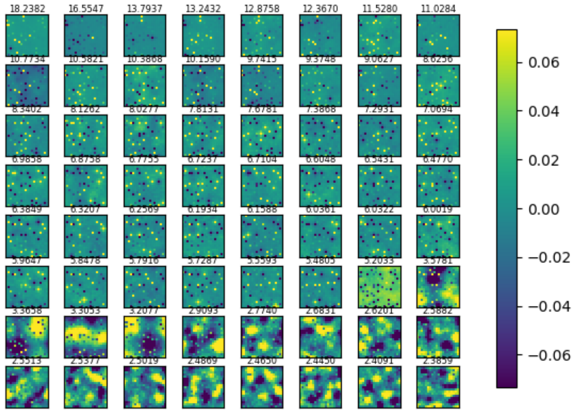
N_temp=200



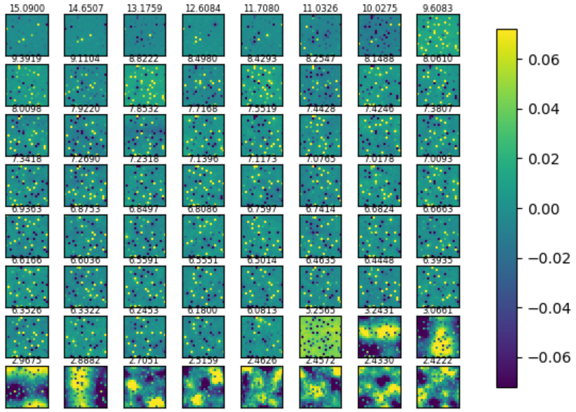
N_temp=300



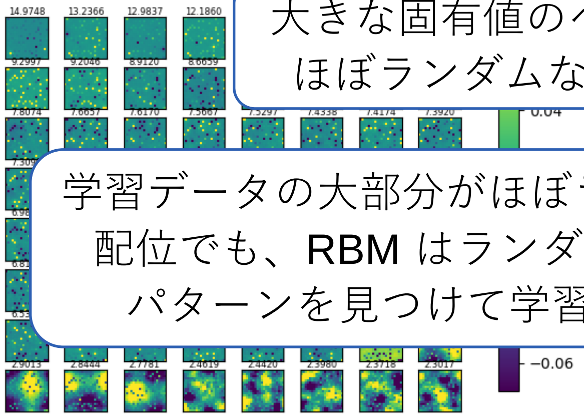
N_temp=400



N_temp=500



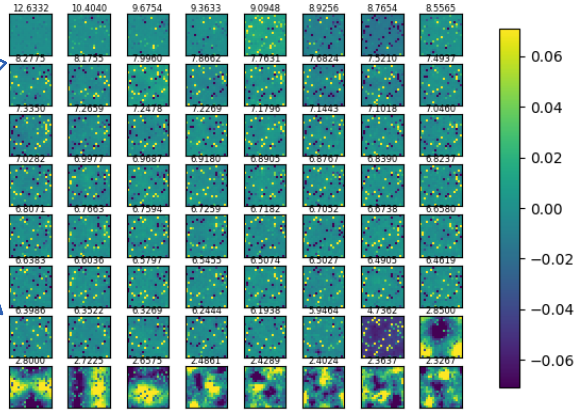
N_temp=600



大きな固有値のベクトルは
ほぼランダムなパターン

学習データの大部分がほぼランダムな
配位でも、RBM はランダムでない
パターンを見つけて学習する！

N_temp=700



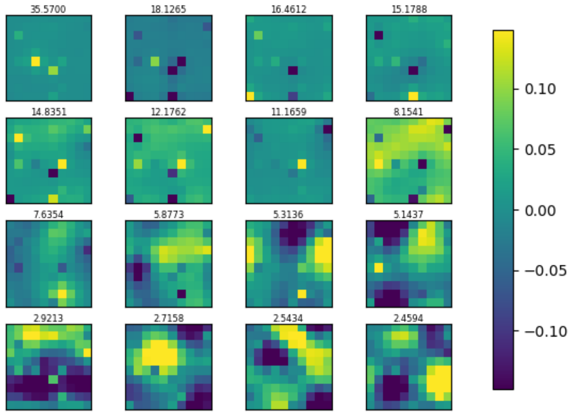
14%

8^2 個中 9 個の固有ベクトルがランダムでないパターン

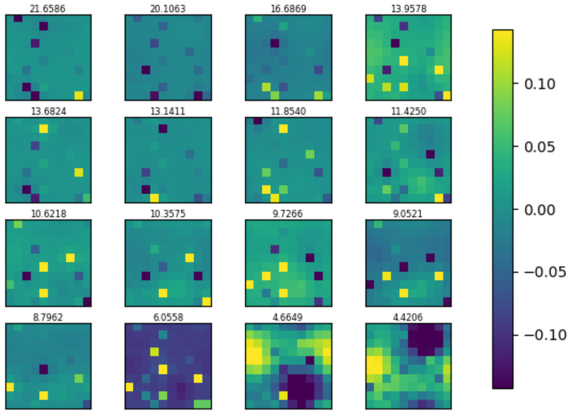
RBM が学習したパターンは weight 行列 ww^T の固有ベクトルの中に見つけられる。

（以下はすべて Nv=10^2, Nh=4^2 の結果）

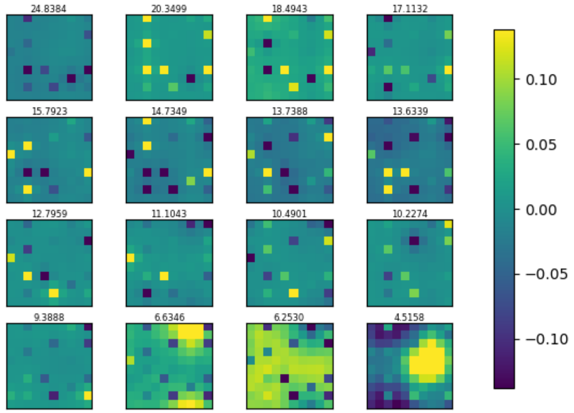
N_temp=200



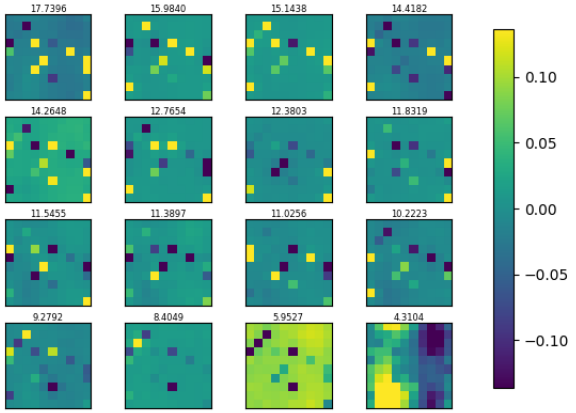
N_temp=300



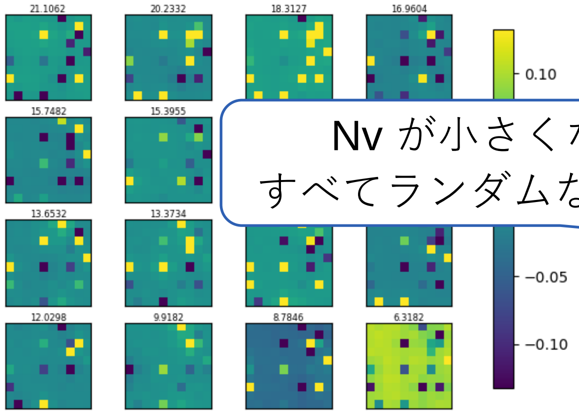
N_temp=400



N_temp=500

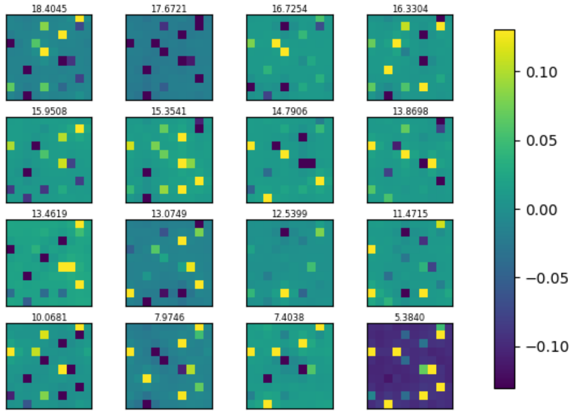


N_temp=600



Nv が小さくなると
すべてランダムなパターン

N_temp=700



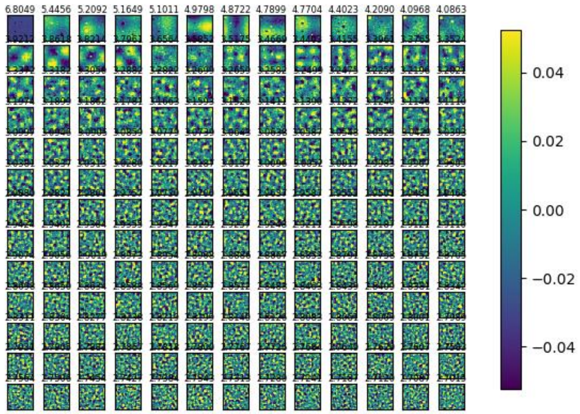
0%

0個の固有ベクトルがランダムでないパターン

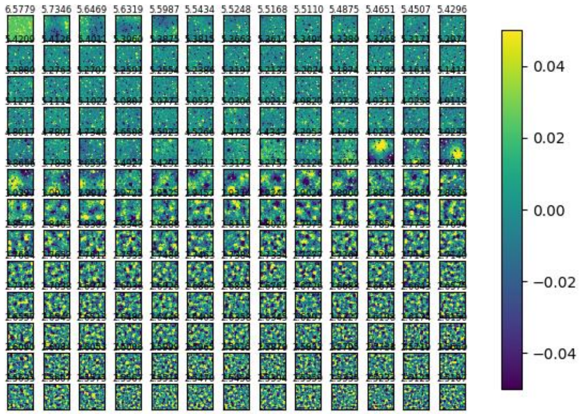
RBM が学習したパターンは weight 行列 ww^T の固有ベクトルの中に見つけられる。

（以下はすべて Nv=32^2, Nh=13^2 の結果）

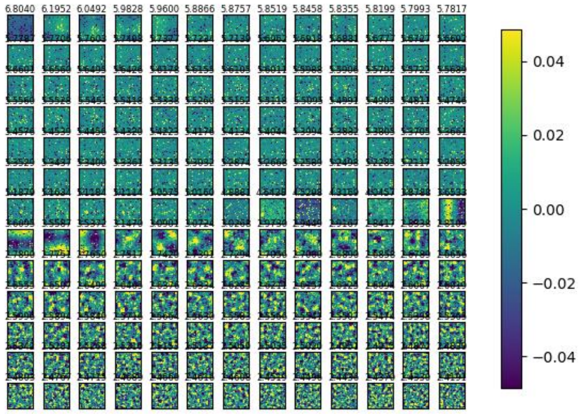
N_temp=200



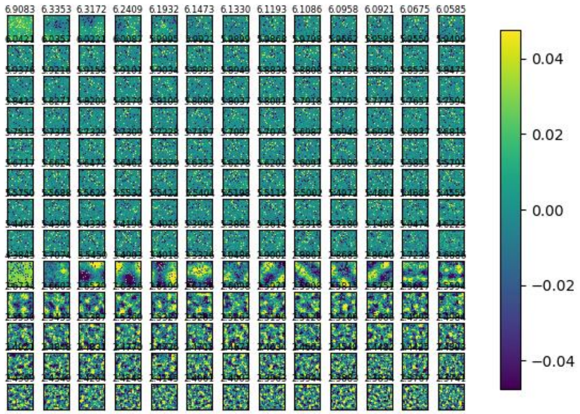
N_temp=300



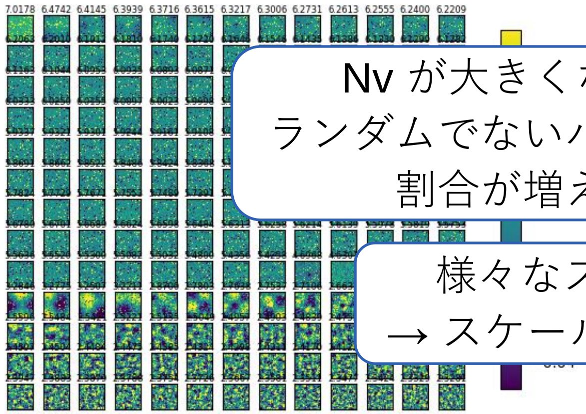
N_temp=400



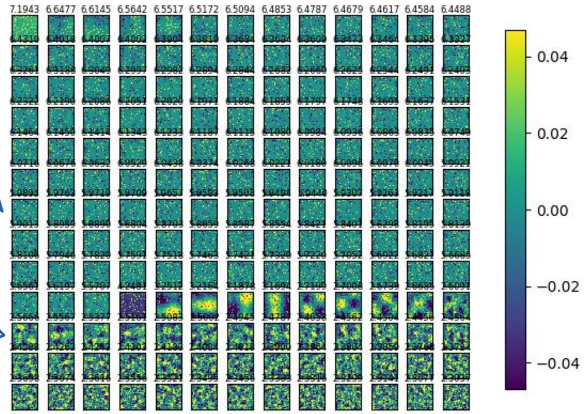
N_temp=500



N_temp=600



N_temp=700



Nv が大きくなると
ランダムでないパターンの
割合が増える

様々なスケール
→ スケール不変性に

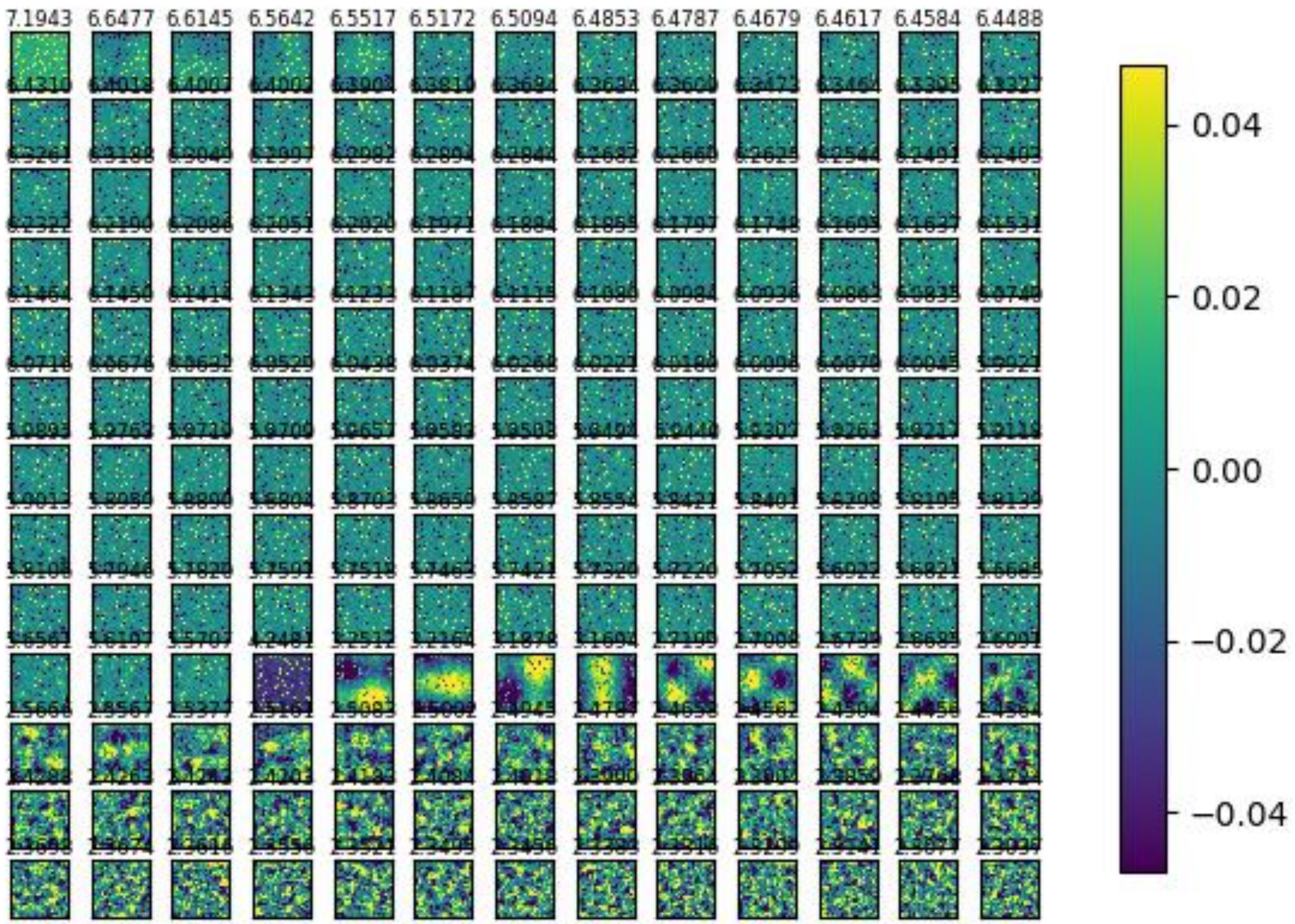
28%

13^2 個中 48 個の固有ベクトルがランダムでないパターン

Nv=32^2,

Nh=13^2,

N_temp=700



結論として

- イジング模型のスピン配位がもつ特徴を抽出するために、RBM の機械学習を行った。
- 配位の再構成によって得られる RBM flow は固定点（＝特徴）をもつ。この点は RG flow と似ているが、flow の振る舞いは明らかに異なる。
- このとき RBM が掴んだ特徴は（相転移点において配位がもつ）スケール不変性であろうと、我々は提案した。
- さらに、RBM flow の固定点が相転移点に現れるために満たすべき条件（N_{temp} は固定した上で、配位のサイズを充分大きくする）も提案した。