Machine learning determination of dynamical parameters: The Ising model case

Guido Cossu, Luigi Del Debbio, Tommaso Giani, Ava Khamseh, Michael Wilson

December 20, 2018

System of spins on a lattice

- system of spins on a lattice

$$\mathbf{s} = \{s_1, .., s_n\}$$

System of spins on a lattice

- system of spins on a lattice

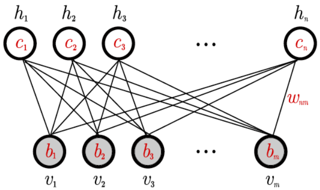
$$\mathbf{s} = \{s_1, .., s_n\}$$

- unknown underlying Boltzman probability distribution

$$p(\mathbf{s}) = \frac{1}{Z} e^{-H(\mathbf{s})}$$

$$H(\mathbf{s}) = \sum_i A_i^{(1)} s_i + \sum_{i,j} A_{ij}^{(2)} s_i s_j + \sum_{i,j,k} A_{ijk}^{(3)} s_i s_j s_k + ..$$

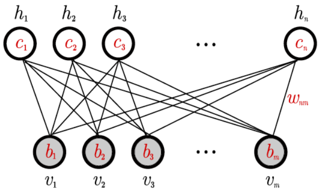
RBM



$$E_\theta(\mathbf{v},\mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$

$$(\mathbf{v},\mathbf{h}) \in \{0,1\}^{n+m}$$

RBM



$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$

$$(\mathbf{v}, \mathbf{h}) \in \{0, 1\}^{n+m}$$

$$\boxed{p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$

RBM

$$p_{\text{RBM}}(\mathbf{v}|\theta) = \sum_{\mathbf{h}} p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-\mathcal{E}(\mathbf{v}|\theta)}$$

RBM

$$p_{\text{RBM}}(\mathbf{v}|\theta) = \sum_{\mathbf{h}} p_{\text{RBM}}(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-\mathcal{E}(\mathbf{v}|\theta)}$$

$\mathbf{v}$: visible states $\rightarrow$ physical states

$\mathbf{h}$: hidden states $\rightarrow$ additional degrees of freedom

RBM

$$p_{\text{RBM}}(\mathbf{v}|\theta) = \sum_{\mathbf{h}} p_{\text{RBM}}(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-\mathcal{E}(\mathbf{v}|\theta)}$$

$\mathbf{v}$: visible states $\rightarrow$ physical states

$\mathbf{h}$: hidden states $\rightarrow$ additional degrees of freedom

$$p_{\text{RBM}}(h_i = 1|\mathbf{v},\theta) = \mathrm{sig}\left(\sum_j w_{ij} v_j + c_i\right)$$

$$p_{\text{RBM}}(v_j = 1|\mathbf{h},\theta) = \mathrm{sig}\left(\sum_i w_{ij} h_i + b_j\right)$$

RBM

$$p_{\text{RBM}}(\mathbf{v}|\theta) = \sum_{\mathbf{h}} p_{\text{RBM}}(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-\mathcal{E}(\mathbf{v}|\theta)}$$

$\mathbf{v}$: visible states $\rightarrow$ physical states
$\mathbf{h}$: hidden states $\rightarrow$ additional degrees of freedom

$$p_{\text{RBM}}(h_i = 1|\mathbf{v},\theta) = \mathrm{sig}\left(\sum_j w_{ij} v_j + c_i\right)$$

$$p_{\text{RBM}}(v_j = 1|\mathbf{h},\theta) = \mathrm{sig}\left(\sum_i w_{ij} h_i + b_j\right)$$

$$\mathrm{sig}(x) = \frac{1}{1+e^{-x}}$$

Training of an RBM

Identify each $\mathbf{v}$ with one physical state $\mathbf{s}$ and construct the log-likelihood

Training of an RBM

Identify each $\mathbf{v}$ with one physical state $\mathbf{s}$ and construct the log-likelihood

$$S = \{\mathbf{s}_1, \mathbf{s}_2, .., \mathbf{s}_n\} \to \{\mathbf{v}_1, \mathbf{v}_2, .., \mathbf{v}_n\}$$

$$\mathcal{L}(\theta|S) = \prod_i p_{\mathrm{RBM}}(\mathbf{v}_i|\theta)$$

Training of an RBM

Identify each $\mathbf{v}$ with one physical state $\mathbf{s}$ and construct the log-likelihood

$$S = \{\mathbf{s}_1, \mathbf{s}_2, .., \mathbf{s}_n\} \rightarrow \{\mathbf{v}_1, \mathbf{v}_2, .., \mathbf{v}_n\}$$

$$\mathcal{L}\left(\theta|S\right) = \prod_i p_{\text{RBM}}(\mathbf{v}_i|\theta)$$

$$\boxed{\log \mathcal{L}\left(\theta|S\right) = \sum_i \log p_{\text{RBM}}(\mathbf{v}_i|\theta)}$$

Training of an RBM

Identify each $\mathbf{v}$ with one physical state $\mathbf{s}$ and construct the log-likelihood

$$S = \{\mathbf{s}_1, \mathbf{s}_2, .., \mathbf{s}_n\} \rightarrow \{\mathbf{v}_1, \mathbf{v}_2, .., \mathbf{v}_n\}$$

$$\mathcal{L}\left(\theta | S\right) = \prod_i p_{\text{RBM}}(\mathbf{v}_i | \theta)$$

$$\boxed{\log \mathcal{L}\left(\theta | S\right) = \sum_i \log p_{\text{RBM}}(\mathbf{v}_i | \theta)}$$

$$\theta_{n+1} = \theta_n + \lambda \partial_\theta \left(\log \mathcal{L}\right)$$

Derivatives of the log-likelihood

$$\frac{1}{|S|}\sum_{\mathbf{v}\in S}\left(\frac{\partial \log \mathcal{L}(\theta|\mathbf{v})}{\partial w_{ij}}\right) = \ldots$$

$$= \frac{1}{|S|}\sum_{\mathbf{v}\in S} p(h_i = 1|\mathbf{v};\theta)v_j - \left\langle p(h_i = 1|\mathbf{v}';\theta)v_j'\right\rangle_{p_{\mathrm{RBM}}(\mathbf{v}')}$$

Derivatives of the log-likelihood

$$\frac{1}{|S|}\sum_{\mathbf{v}\in S}\left(\frac{\partial \log \mathcal{L}(\theta|\mathbf{v})}{\partial w_{ij}}\right) = \dots$$

$$= \frac{1}{|S|}\sum_{\mathbf{v}\in S} p(h_i = 1|\mathbf{v};\theta)v_j - \left\langle p(h_i = 1|\mathbf{v}';\theta)v_j'\right\rangle_{p_{\mathrm{RBM}}(\mathbf{v}')}$$

$$= \underbrace{\boxed{\mathbb{E}_{\mathrm{data}}\left[p(h_i = 1|\mathbf{v};\theta)v_j\right]}}_{\text{ok..}} - \underbrace{\boxed{\mathbb{E}_{\mathrm{model}}\left[p(h_i = 1|\mathbf{v}';\theta)v_j'\right]}}_{\text{???}}$$

Gibbs sampling

Build a MC which converges to the target distribution $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$.

Gibbs sampling

Build a MC which converges to the target distribution $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$.

- Use $p_{\text{RBM}}(h_i = 1|\mathbf{v}, \theta)$ and $p_{\text{RBM}}(v_j = 1|\mathbf{h}, \theta)$ to build

$$\mathbf{v}_0 \rightarrow \mathbf{h}_0 \rightarrow \mathbf{v}_1 \rightarrow \mathbf{h}_1 \rightarrow .... \rightarrow \mathbf{v}_k \rightarrow \mathbf{h}_k$$

Gibbs sampling

Build a MC which converges to the target distribution $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$.

- Use $p_{\text{RBM}}(h_i = 1|\mathbf{v}, \theta)$ and $p_{\text{RBM}}(v_j = 1|\mathbf{h}, \theta)$ to build

$$\mathbf{v}_0 \rightarrow \mathbf{h}_0 \rightarrow \mathbf{v}_1 \rightarrow \mathbf{h}_1 \rightarrow .... \rightarrow \mathbf{v}_k \rightarrow \mathbf{h}_k$$

- At some point $(\mathbf{v}, \mathbf{h})$ will be sampled according to $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$

Gibbs sampling

Build a MC which converges to the target distribution $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$.

- Use $p_{\text{RBM}}(h_i = 1|\mathbf{v}, \theta)$ and $p_{\text{RBM}}(v_j = 1|\mathbf{h}, \theta)$ to build

$$\mathbf{v}_0 \rightarrow \mathbf{h}_0 \rightarrow \mathbf{v}_1 \rightarrow \mathbf{h}_1 \rightarrow .... \rightarrow \mathbf{v}_k \rightarrow \mathbf{h}_k$$

- At some point $(\mathbf{v}, \mathbf{h})$ will be sampled according to $p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta)$
- usually $k = 1$ is enough

Estimation of the log-likelihood

How to estimate the loglikelihood to monitor the training?

$$\log \mathcal{L}\left(\theta|S\right) = \frac{1}{|S|}\sum_{i\in S} \log p_{\text{RBM}}(\mathbf{v}_i|\theta) = -\langle \mathcal{E}(\mathbf{v}|\theta)\rangle_S - \log Z_{\text{RBM}}$$

Estimation of the log-likelihood

How to estimate the loglikelihood to monitor the training?

$$\log \mathcal{L}(\theta|S) = \frac{1}{|S|}\sum_{i\in S} \log p_{\text{RBM}}(\mathbf{v}_i|\theta) = -\langle \mathcal{E}(\mathbf{v}|\theta)\rangle_S - \log Z_{\text{RBM}}$$

Choose a proposal distribution $p_0(\mathbf{v}) = p_0^*(\mathbf{v})/Z_0$ whose partition function is known exactly.

Estimation of the log-likelihood

How to estimate the loglikelihood to monitor the training?

$$\log \mathcal{L}(\theta|S) = \frac{1}{|S|}\sum_{i\in S} \log p_{\text{RBM}}(\mathbf{v}_i|\theta) = -\langle \mathcal{E}(\mathbf{v}|\theta)\rangle_S - \log Z_{\text{RBM}}$$

Choose a proposal distribution $p_0(\mathbf{v}) = p_0^*(\mathbf{v})/Z_0$ whose partition function is known exactly.

Compute

$$Z_{\text{RBM}} = \int d\mathbf{v}\, p^*_{\text{RBM}}(\mathbf{v}) = Z_0 \int d\mathbf{v}\, p_0(\mathbf{v}) \frac{p^*_{\text{RBM}}(\mathbf{v})}{p_0^*(\mathbf{v})} \simeq \frac{Z_0}{m}\sum_i^m \frac{p^*_{\text{RBM}}\left(\mathbf{v}^{(i)}\right)}{p_0^*\left(\mathbf{v}^{(i)}\right)}$$

Estimation of the log-likelihood

How to estimate the loglikelihood to monitor the training?

$$\log \mathcal{L}(\theta|S) = \frac{1}{|S|} \sum_{i \in S} \log p_{\text{RBM}}(\mathbf{v}_i|\theta) = -\langle \mathcal{E}(\mathbf{v}|\theta) \rangle_S - \log Z_{\text{RBM}}$$

Choose a proposal distribution $p_0(\mathbf{v}) = p_0^*(\mathbf{v})/Z_0$ whose partition function is known exactly.

Compute

$$Z_{\text{RBM}} = \int d\mathbf{v}\, p^*_{\text{RBM}}(\mathbf{v}) = Z_0 \int d\mathbf{v}\, p_0(\mathbf{v}) \frac{p^*_{\text{RBM}}(\mathbf{v})}{p^*_0(\mathbf{v})} \simeq \frac{Z_0}{m} \sum_i^m \frac{p^*_{\text{RBM}}\left(\mathbf{v}^{(i)}\right)}{p^*_0\left(\mathbf{v}^{(i)}\right)}$$

Get

$$Z_{\text{RBM}}/Z_0 = \frac{1}{m} \sum_i^m \frac{p^*_{\text{RBM}}\left(\mathbf{v}^{(i)}\right)}{p^*_0\left(\mathbf{v}^{(i)}\right)} \qquad \text{with } \mathbf{v}^{(i)} \propto p_0(\mathbf{v})$$

Annealing importance sampling

- $p^*_{\text{RBM}} \propto e^{-\beta\mathcal{E}}$ $\qquad \beta = 1$
  $p^*_0 \propto e^{-\beta_0\mathcal{E}}$ $\qquad \beta_0 = 0 \to Z_0 = 2^N$

Annealing importance sampling

- $p^*_{\text{RBM}} \propto e^{-\beta\mathcal{E}}$ $\qquad \beta = 1$
  $p^*_0 \propto e^{-\beta_0\mathcal{E}}$ $\qquad \beta_0 = 0 \rightarrow Z_0 = 2^N$
- Introduce intermediate closer distribution $p_{\beta_0}, p_{\beta_1}, \cdots, p_{\beta_n}$ such that $0 = \beta_0 < \beta_1 < \cdots < \beta_{n-1} < \beta = 1$.

Annealing importance sampling

- $p^*_{\text{RBM}} \propto e^{-\beta\mathcal{E}}$ $\qquad \beta = 1$
  $p^*_0 \propto e^{-\beta_0\mathcal{E}}$ $\qquad \beta_0 = 0 \to Z_0 = 2^N$
- Introduce intermediate closer distribution $p_{\beta_0}, p_{\beta_1}, \cdots, p_{\beta_n}$ such that $0 = \beta_0 < \beta_1 < \cdots < \beta_{n-1} < \beta = 1$.
- Estimate Z_{RBM}/Z_0 using

$$\frac{Z_{\text{RBM}}}{Z_0} = \frac{Z_{\beta_1}}{Z_0}\frac{Z_{\beta_2}}{Z_{\beta_1}}\cdots\frac{Z_{\beta_{n-2}}}{Z_{\beta_{n-1}}}\frac{Z_1}{Z_{\beta_{n-1}}} = \prod_{j=0}^{n-1}\frac{Z_{\beta_{j+1}}}{Z_{\beta_j}} \, ,$$

Ising model

Generate a sample $D = \{s^1, s^2, \ldots\}$, $N_D \sim 10^5$, according to

$$p_D(s) = \frac{1}{Z(J,h)}\, e^{-H_{J,h}(s)}$$

$$H_{J,h} = -\sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i \quad , \quad Z(J,h) = \sum_s e^{-H_{J,h}(s)}$$

Ising model

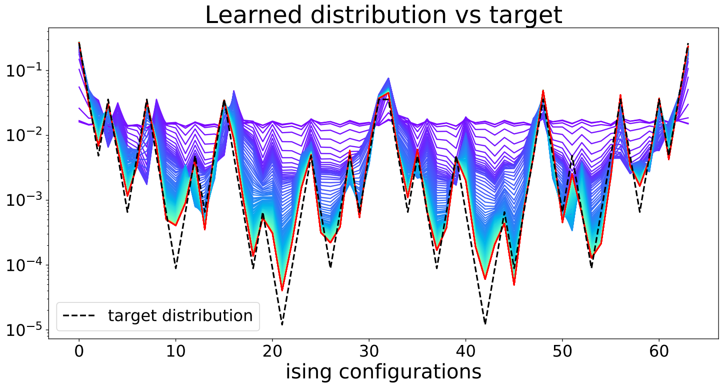
Generate a sample $D = \{s^1, s^2, \ldots\}$, $N_D \sim 10^5$, according to

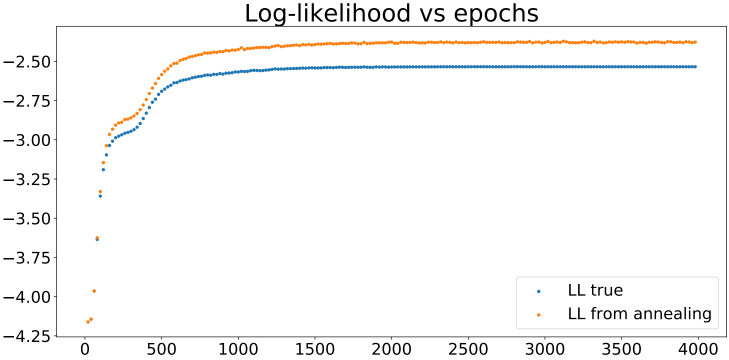
$$p_D(s) = \frac{1}{Z(J,h)}\, e^{-H_{J,h}(s)}$$

$$H_{J,h} = -\sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i \quad , \quad Z(J,h) = \sum_s e^{-H_{J,h}(s)}$$

Train the machine

Validation: Ising in 1 dimension

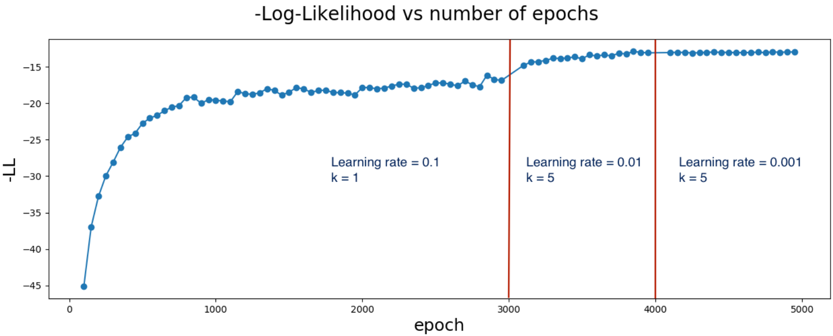




Log-likelihood vs epochs
−2.50
−2.75
−3.00
−3.25
−3.50
−3.75
−4.00
−4.25
0
500
1000
1500
2000
2500
3000
3500
4000
LL true
LL from annealing

Ising in 2 dimensions

- Contrastive Divergence steps $k = 1, 5$
- Learning parameter $\alpha = 0.1, 0.01, 0.0001$
- Batch size 200
- Training epochs 3000,1000,1000


-Log-Likelihood vs number of epochs
-LL
−15
−20
−25
−30
−35
−40
−45
0
1000
2000
3000
4000
5000
epoch
Learning rate = 0.1
k = 1
Learning rate = 0.01
k = 5
Learning rate = 0.001
k = 5

2D Ising Observables

Use the trained RBM to *generate* the spin configurations via Gibbs/Metropolis sampling

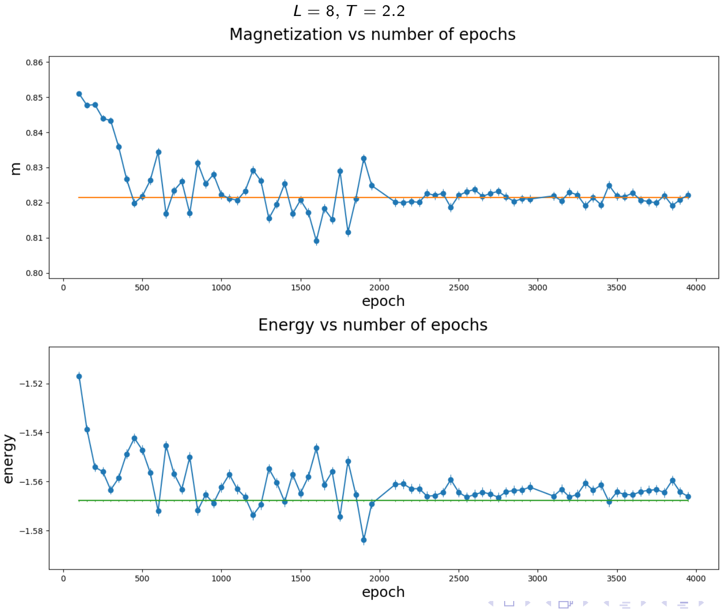
$$\langle m \rangle = \frac{1}{L^2} \left\langle \left| \sum_{i=1}^{L^2} s_i \right| \right\rangle ,$$

$$\langle \chi \rangle = \frac{L^2}{T} \left\langle \langle m^2 \rangle - \langle m \rangle^2 \right\rangle ,$$

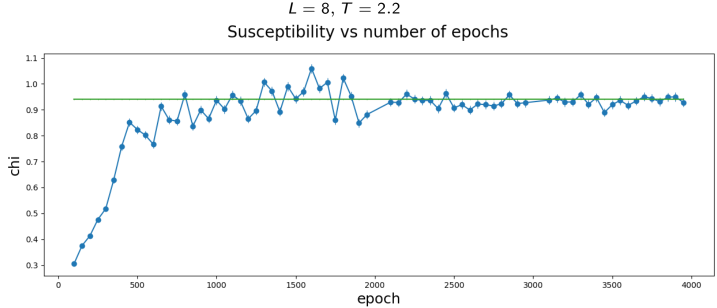
$$\langle E \rangle = -\frac{1}{L^2} \left\langle \sum_{\langle i,j \rangle} s_i s_j \right\rangle ,$$

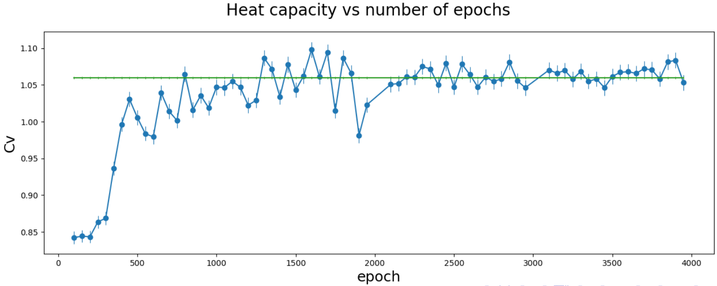
$$\langle c_v \rangle = \frac{L^2}{T^2} \left\langle \langle E^2 \rangle - \langle E \rangle^2 \right\rangle .$$

Results: Magnetisation and energy

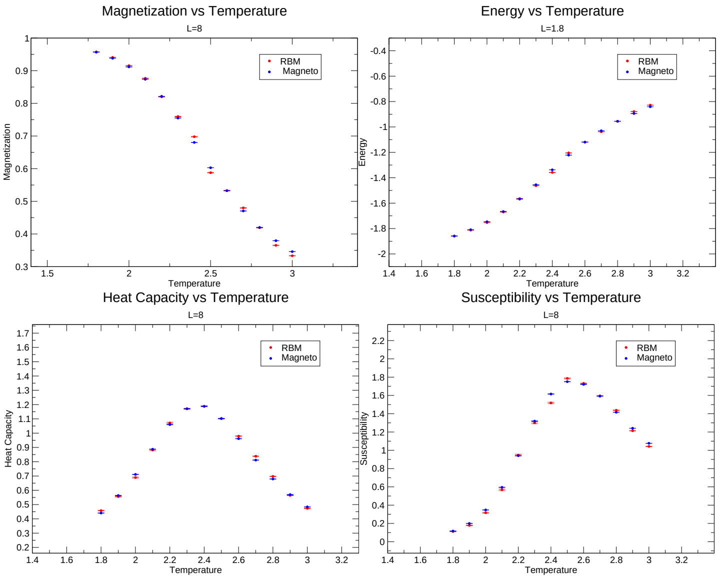


Results: Susceptibility and heat capacity





Results: Observables vs Temperature



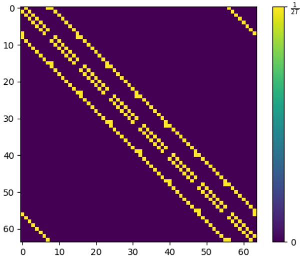
Extracting the couplings

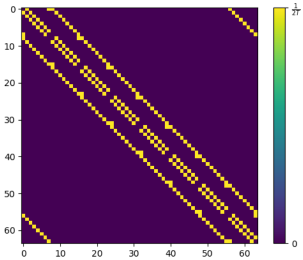
Cumulant generating function:

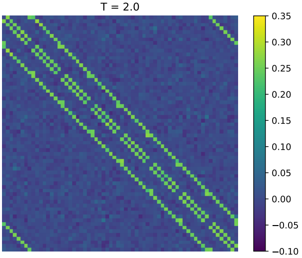
$$K_i(t) := \log \sum_{h_i} q_i(h_i) e^{t_i h_i} = \sum_n \kappa_i^{(n)} \frac{t^n}{n!} \quad , \quad \kappa_i^{(n)} = \partial_t^n K_i|_{t=0}$$

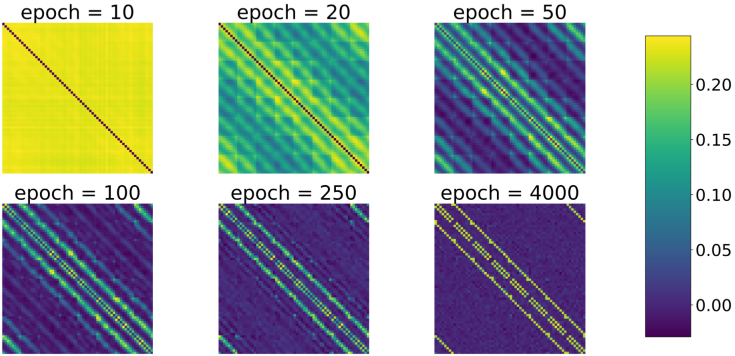
Let $q_i(h_i) = e^{b_i h_i}/Z$, then the h-marginalised energy:

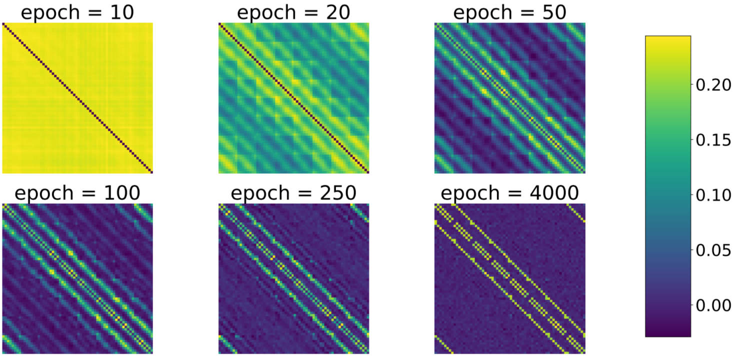
$$\begin{aligned}
\mathcal{E}(\mathbf{v}) &= -\sum_j b_j v_j - \sum_i K_i \left(\sum_j W_{ij} v_j \right) \\
&= -\sum_j b_j v_j - \sum_j \left(\sum_i \kappa_i^{(1)} W_{ij} \right) v_j \\
&\quad -\frac{1}{2} \sum_{jk} \left(\sum_i \kappa_i^{(2)} W_{ik} W_{ij} \right) v_j v_k + \cdots
\end{aligned}$$

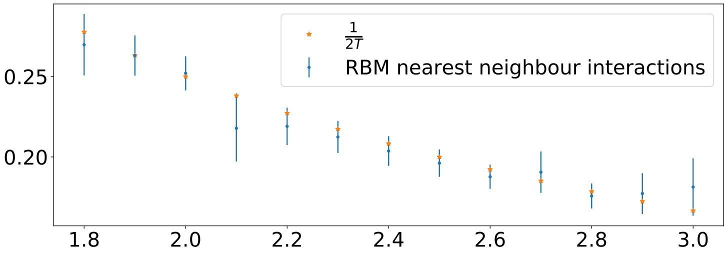

0
10
20
30
40
50
60
0
10
20
30
40
50
60
1/2T
0


0
10
20
30
40
50
60
0
10
20
30
40
50
60
1/2T
0


T = 2.0
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10

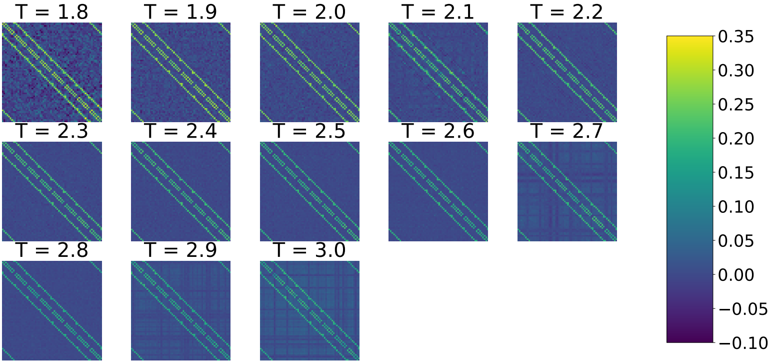

epoch = 10
epoch = 20
epoch = 50
epoch = 100
epoch = 250
epoch = 4000
0.20
0.15
0.10
0.05
0.00

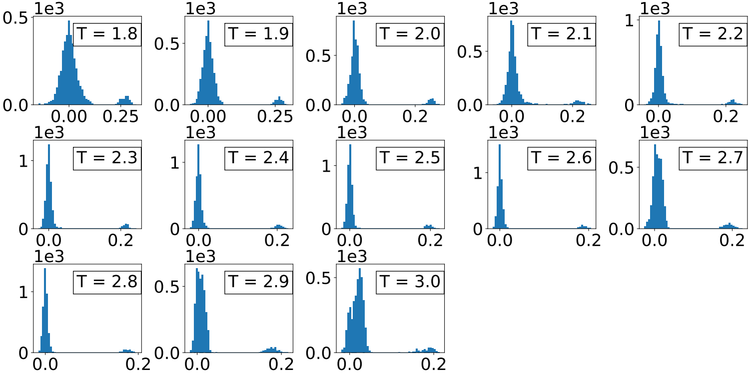

epoch = 10
epoch = 20
epoch = 50
epoch = 100
epoch = 250
epoch = 4000
0.20
0.15
0.10
0.05
0.00

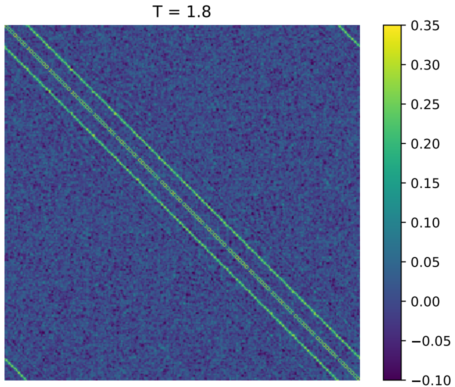

$\frac{1}{2T}$
RBM nearest neighbour interactions
0.25
0.20
1.8
2.0
2.2
2.4
2.6
2.8
3.0

Summary

- we have implemented a binary RBM
- we have trained it over a set of Ising states at different temperatures
- we have computed the log-likelihood along the training and tuned the hiperparameters involved in the training to ensure a monotonic behaviour
- we have computed the moments of the learnt distribution sampling states from it
- we have derived closed expressions for the couplings in the hamiltonian in terms of the parameters of the trained machine


T = 1.8
T = 1.9
T = 2.0
T = 2.1
T = 2.2
T = 2.3
T = 2.4
T = 2.5
T = 2.6
T = 2.7
T = 2.8
T = 2.9
T = 3.0
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10


T = 1.8
T = 1.9
T = 2.0
T = 2.1
T = 2.2
T = 2.3
T = 2.4
T = 2.5
T = 2.6
T = 2.7
T = 2.8
T = 2.9
T = 3.0
1e3


T = 1.8
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10

Couplings

$$H_{j_1 j_2} = \frac{1}{8} \sum_i \ln \frac{(1 + e^{c_i + W_{ij_1} + W_{ij_2}})(1 + e^{c_i})}{(1 + e^{c_i + W_{ij_1}})(1 + e^{c_i + W_{ij_2}})} \quad (1)$$

$$H_{j_1 .. j_N} = \frac{1}{N!} \sum_{l=0}^{N} (-1)^l \sum_{\alpha_1 < \ldots < \alpha_{N-l}} K_i(W_{i,j_{\alpha_1}} + \ldots + W_{i,j_{\alpha_{N-l}}}) \quad (2)$$