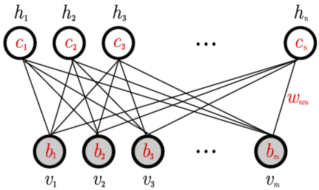
Machine learning determination of dynamical parameters: The Ising model case

A collaboration with:
Guido Cossu, Del Debbio, Tommaso Giani, Ava Khamseh, Michael Wilson

CENTRE FOR THEORETICAL PHYSICS Higgs
igmm INSTITUTE OF GENETICS & MOLECULAR MEDICINE

November 26, 2018

RBM



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

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

Summary of probability

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

$$p_{\text{RBM}}(\mathbf{v}|\theta) = \mathrm{Tr}_h \, p_{\text{RBM}}(\mathbf{v}, h|\theta) = \sum_h p_{\text{RBM}}(\mathbf{v}, h|\theta)$$

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

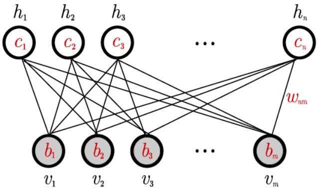
$$\mathcal{E}(\mathbf{v}|\theta) = -\sum_j b_j v_j - \sum_i \log\left[1 + e^{c_i + \sum_j w_{ij} v_j}\right]$$

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

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

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

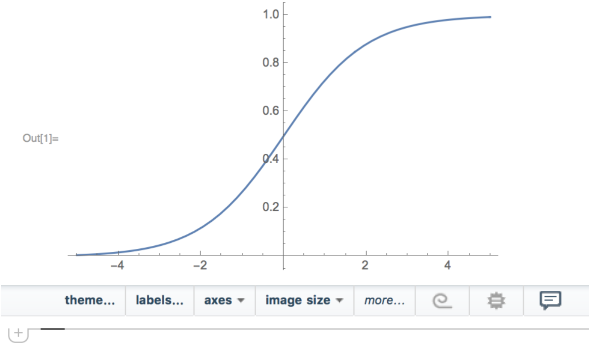
Activation probability



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

In[1]:= `Plot[1 / (1 + Exp[-x]), {x, -5, 5}]`

Out[1]=



Log-likelihood & KL divergence

$$D_{\mathrm{KL}}\left(q_{\mathrm{data}}(\mathbf{v})||p_\theta(\mathbf{v})\right) = \sum_{\mathbf{v}} q_{\mathrm{data}}(\mathbf{v}) \log\left(\frac{q_{\mathrm{data}}(\mathbf{v})}{p(\mathbf{v})}\right)$$

$$= \sum_{\mathbf{v}} \left(q_{\mathrm{data}}(\mathbf{v}) \log\left(q_{\mathrm{data}}(\mathbf{v})\right) - q_{\mathrm{data}}(\mathbf{v}) \log\left(p_\theta(\mathbf{v})\right)\right)$$

Max likelihood $\Longleftrightarrow$ min KL divergence

$$\frac{\partial \log \mathcal{L}(\theta|\mathbf{v})}{\partial w_{ij}} = p(h_i = 1|\mathbf{v})v_j - \left\langle p(h_i = 1|\mathbf{v}')v_j\right\rangle_{p(\mathbf{v}')} \ .$$

$\Longrightarrow$ Contrastive Divergence

Contrastive Divergence

Estimate $p(h_i = 1|\mathbf{v})v_j - \langle p(h_i = 1|\mathbf{v}')v_j \rangle_{p(\mathbf{v}')}$, $\forall i, j$ as:

$$p(h_i = 1|\mathbf{v}^{(\mathbf{0})})v_j - p(h_i = 1|\mathbf{v}^{(\mathbf{k})})v_j^{(k)}$$

- k : Gibbs sampling step, typically set $k = 1$
- Initialised with a training example $\mathbf{v}^{(\mathbf{0})}$
- Each step t involves sampling $h^{(t)} \sim p_{\text{RBM}}(h_i = 1|\mathbf{v}^{(t)}, \theta)$, then sampling $v^{(t+1)} \sim p_{\text{RBM}}(v_j = 1|\mathbf{h}^{(t)}, \theta)$

Importance sampling

Original distribution: $p_1(\mathbf{v}) = p_1^*(\mathbf{v})/Z_1$
Simple distribution: $p_0(\mathbf{v}) = \frac{1}{Z_0} p_0^*(\mathbf{v})$

$$Z_1 = \int p_1^*(\mathbf{v}) d\mathbf{v} = Z_0 \int p_0(\mathbf{v}) \frac{p_1^*(\mathbf{v})}{p_0^*(\mathbf{v})} d\mathbf{v} \ ,$$

measure its estimator

$$\hat{Z}_1 = \frac{Z_0}{M} \sum_{m=1}^{M} \frac{p_1^*(\mathbf{v}^{(m)})}{p_0^*(\mathbf{v}^{(m)})} \quad \text{s.t. } : \mathbf{v}^{(m)} \sim p_0 \ ,$$

$$\hat{\text{Var}}[\hat{Z}_1] = \frac{\hat{Z}_1^2}{M^2} \sum_{m=1}^{M} \left[\frac{p_1(\mathbf{v}^{(m)})}{p_0(\mathbf{v}^{(m)})} - 1 \right]^2$$ is large if p_0 and p_1 not close!

Annealed Importance sampling

- Bridge the original distribution p_1 and the simple distribution p_0
- Introduce intermediate closer distribution $p_{\beta_0}, p_{\beta_1}, \cdots, p_{\beta_n}$ s.t. $0 = \beta_0 < \beta_1 < \cdots < \beta_{n-1} < \beta_n = 1$. Estimate Z_1/Z_0 via:

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

- Geometric mean:

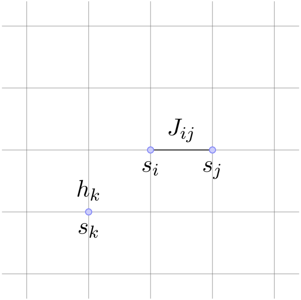
$$p_{\beta_j} \propto p_1^*(\mathbf{v})^{\beta_j} p_0^*(\mathbf{v})^{1-\beta_j} \ .$$

$$p_1^*(\mathbf{v})^{\beta_j} p_0^*(\mathbf{v})^{1-\beta_j} = e^{-\beta E_1(\mathbf{v})} e^{-(1-\beta)E_0(\mathbf{v})} = e^{-E_0} e^{-\beta(E_1 - E_0)} \ .$$

Ising configurations as training input

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

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

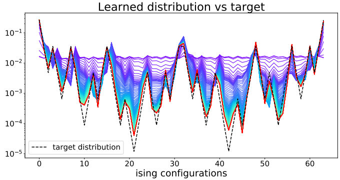


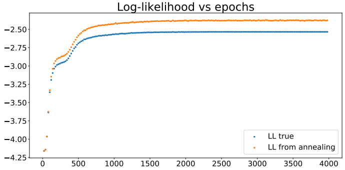
MC simulation of the 2D Ising model at various temperatures

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

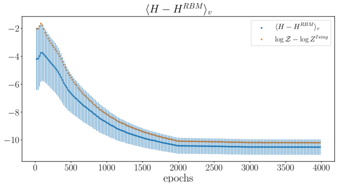
Onsager 1944

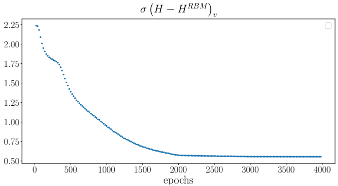
Validation in 1-Dimension





$$D_{\mathrm{KL}} = \sum_{\mathbf{v}} q_{\mathrm{Ising}}(\mathbf{v}) \left[\left(F_{v}\{H\} - F_{vh}\{E\} \right) - \left(H(\mathbf{v}) - H_{\lambda}^{\mathrm{RBM}}(\mathbf{v}) \right) \right]$$

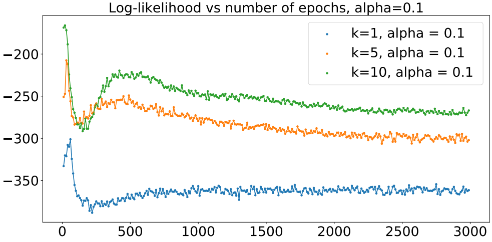


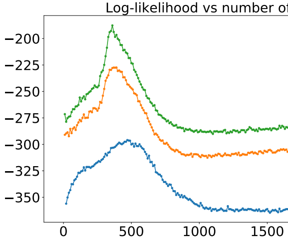


Training parameters for 2D Ising model

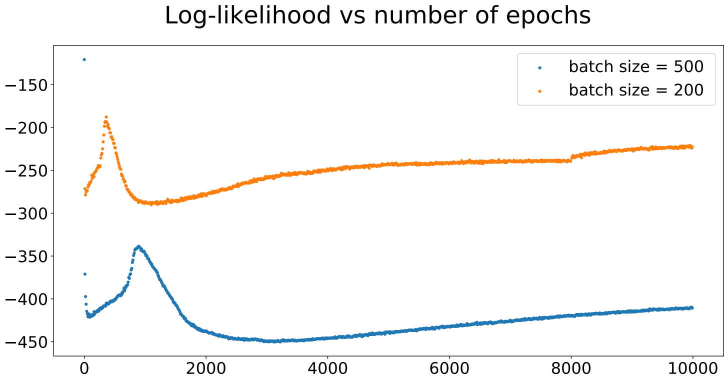
- Contrastive Divergence steps $k = 1, 5$
- Learning parameter $\alpha = 0.1, 0.01, 0.0001$
- Batch size 200
- Training epochs 3000,1000,1000
- Use the trained RBM to *generate* the spin configs via Gibbs/Metropolis sampling

Choice of hyperparameters: Contrastive divergence k

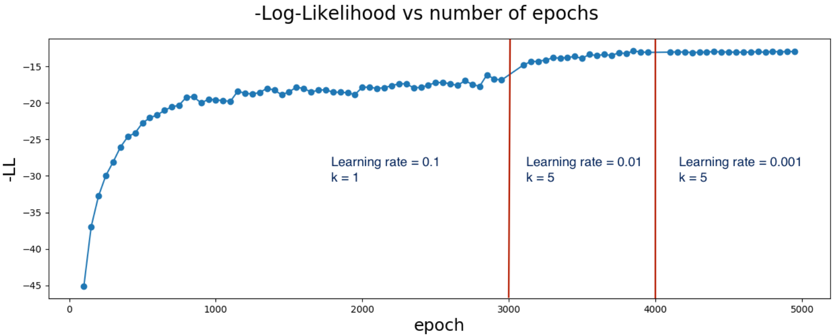




Choice of hyperparameters: Batch size



Results: Log-likelihood vs epoch

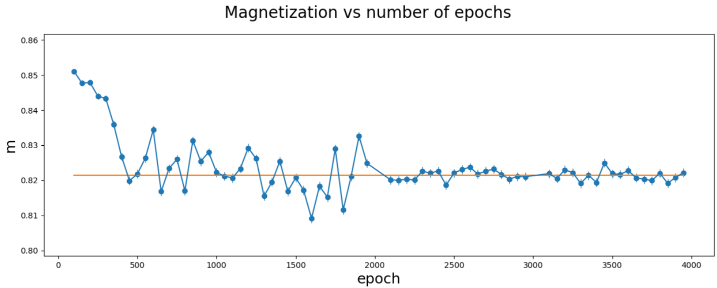


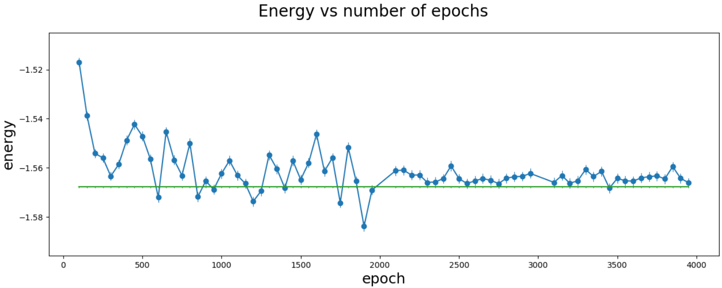
2D Ising Observables

$$
\begin{aligned}
\langle m \rangle &= \frac{1}{L^2} \left\langle \left| \sum_{i=1}^{L^2} s_i \right| \right\rangle, \\
\langle \chi \rangle &= \frac{L^2}{T} \left\langle \langle m^2 \rangle - \langle m \rangle^2 \right\rangle, \\
\langle E \rangle &= -\frac{1}{L^2} \left\langle \sum_{\langle i,j \rangle} s_i s_j \right\rangle, \\
\langle c_v \rangle &= \frac{L^2}{T^2} \left\langle \langle E^2 \rangle - \langle E \rangle^2 \right\rangle.
\end{aligned}
$$

Results: Magnetisation and energy

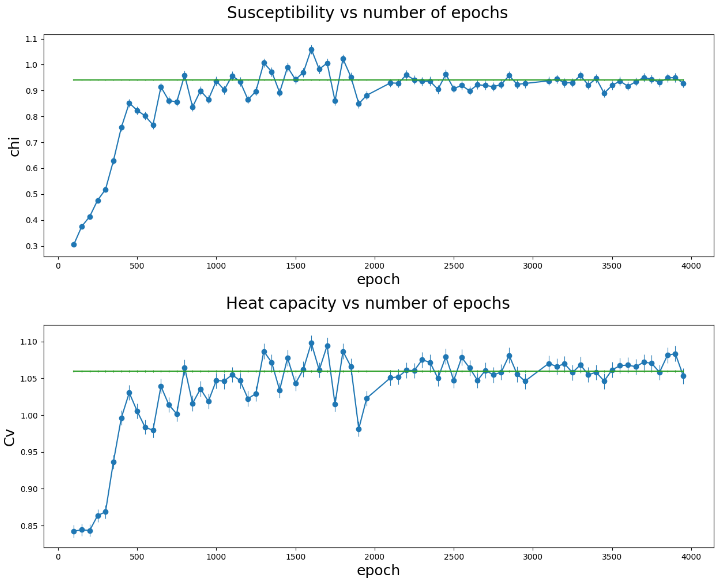
$L = 8$, $T = 2.2$



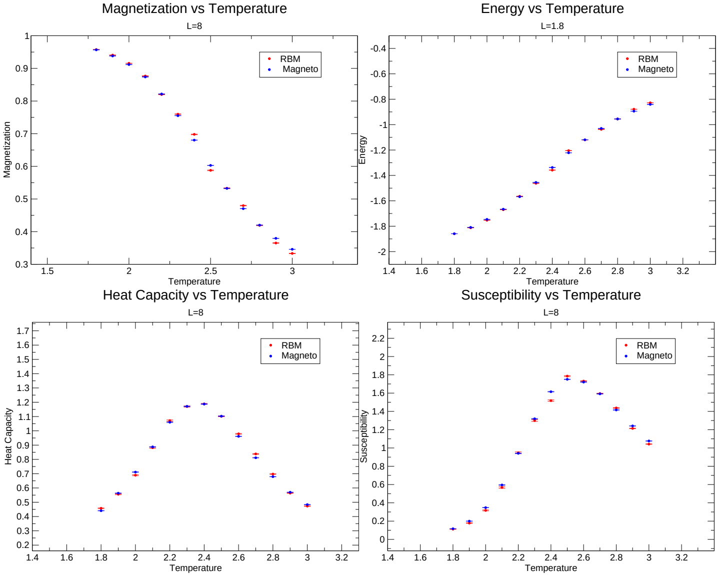


Results - Susceptibility and heat capacity

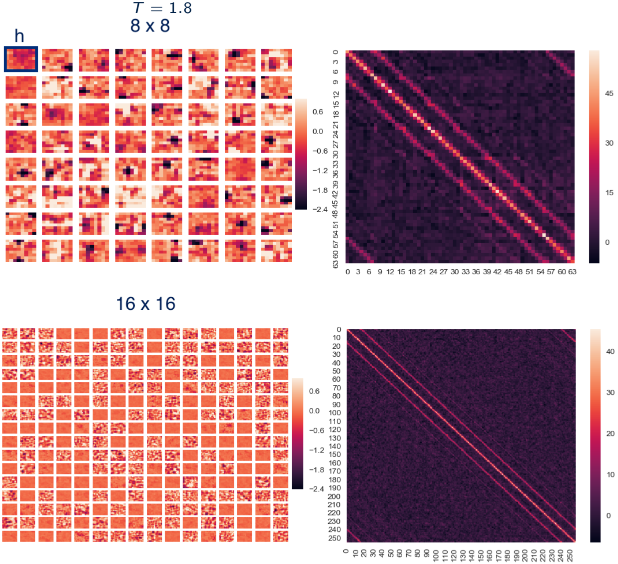
$L = 8,\ T = 2.2$



Results: Observables vs Temperature



Results: W and W^TW matrices



Extracting the couplings

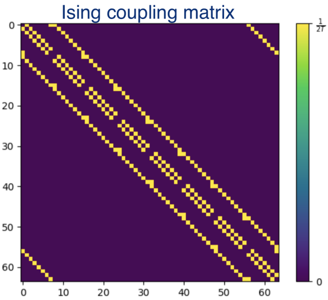
Cumulant generating function:

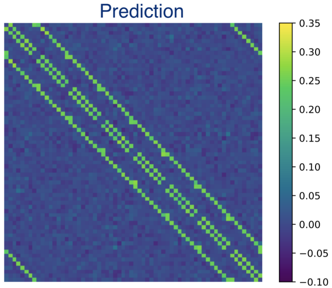
$$K_i(t) := \log \sum_{h_i} q_i(h_i) e^{t_i h_i} = \sum_n \kappa_i^{(n)} \frac{t^n}{n!} \quad , \quad \kappa_i^{(n)} = \partial_t^n K_i|_{t=0}$$

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

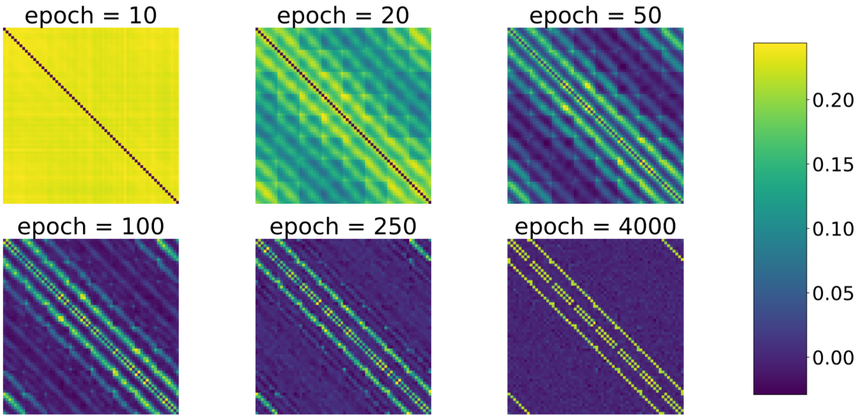
$$\begin{aligned} E(\mathbf{v}) &= -\sum_j b_j v_j - \sum_i K_i \left(\sum_j W_{ij} v_j \right) \\ &= -\sum_j b_j v_j - \sum_j \left(\sum_i \kappa_i^{(1)} W_{ij} \right) v_j \\ &\quad - \frac{1}{2} \sum_{jk} \left(\sum_i \kappa_i^{(2)} W_{ik} W_{ij} \right) v_j v_k + \cdots \end{aligned}$$

The coupling matrix

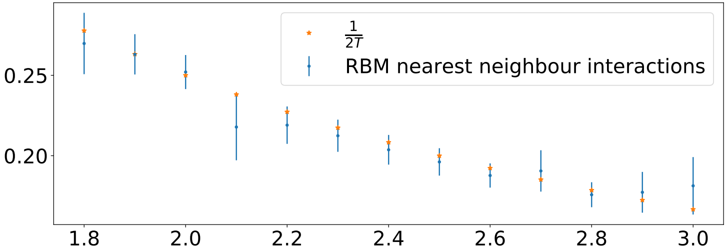




Coupling structure during the training



Predicted coupling values at all temperatures



Summary

- Defined a generative energy based binary model - RBM
- Trained on data generated from a different binary energy based model - Ising model
- Used annealed importance sampling to check convergence during training
- Validated model in 1D because partition function is tractable with sufficiently small lattice
- Used Ising observables to validate in 2D
- Used cumulative generating function to extract physical coupling from the RBM

Summary

- **Defined a generative energy based binary model - RBM**
- Trained on data generated from a different binary energy based model - Ising model
- **Used annealed importance sampling to check convergence during training**
- Validated model in 1D because partition function is tractable with sufficiently small lattice
- Used Ising observables to validate in 2D
- **Used cumulative generating function to extract physical coupling from the RBM**

Literature review

- *An exact mapping between the variational renormalization group and deep learning*, P. Mehta and D. J. Schwab [2014].
- *Learning thermodynamics with Boltzmann machines*, G. Torlai and R. G. Melko [2016]
- *Deep Learning the Ising Model Near Criticality*, A. Morningstar and R. G. Melko [2017]