Machine learning determination of dynamical interactions
from the Ising model to biological data

Guido Cossu

Director of Research at Ascent Robotics Inc.

Honorary Fellow at Higgs Centre, Edinburgh University

Ava Khamseh

Institute of Genetics & Molecular Medicine, Edinburgh University

in collaboration with: Luigi del Debbio, Tommaso Giani, Michael Wilson

Ascent

ascent.ai

Who are Ascent Robotics?

Founded in 2016, we are an independent Japanese AI and robotics innovator developing cutting-edge AI software to create a more autonomous future.

We have attracted leading AI talent from around the world and are working with a number of leading Japanese and Global Robotics and Auto OEMs to introduce "intelligent" elements to their technology.

This includes software to enable fully-autonomous cars and industrial robots.

Ascent

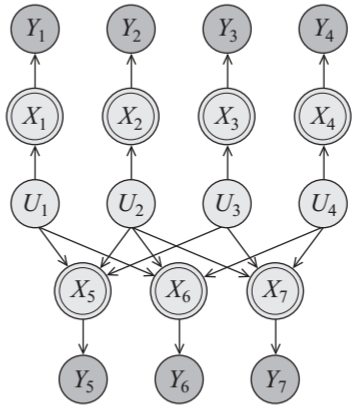
ascent.ai

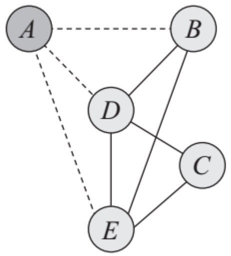
Who are Ascent Robotics?

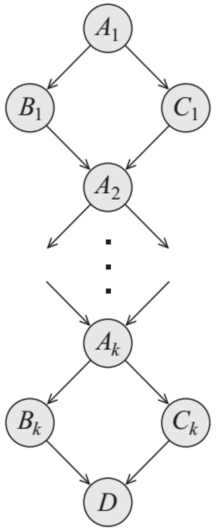
- ✓ More than 60 top class engineers and scientists
- ✓ Multidisciplinary approach to AI and robotics
- ✓ International environment
- ✓ Strong research focus
  - ✓ Decision-making, inference models, motion planning, computer vision, …
- ✓ Extending the network of academic collaborations

Motivation

- Couplings between different genes from gene expression data
  - Arbitrary couplings of any order
- Address simpler problem: Ising model, pairwise couplings
- Learn distribution from data → extract couplings

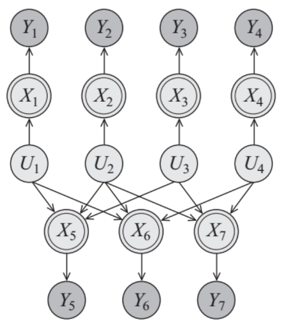


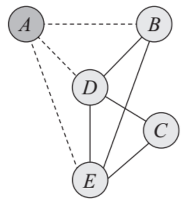


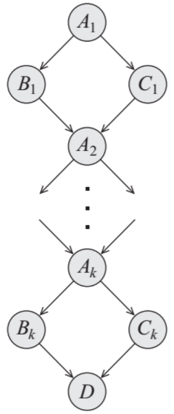


Probabilistic graphical models

- Conditional relations between random variables
- Directed and undirected graphs
- Undirected: Markov Random Fields (MRF)
- Energy based models (EBM)



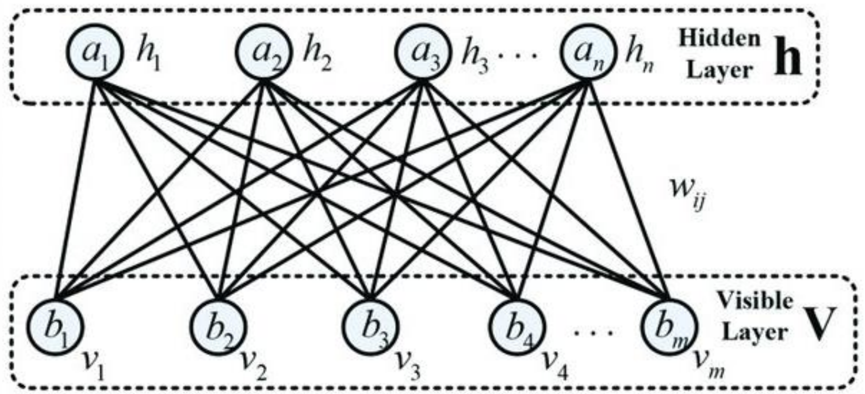




$$p(\mathbf{v}) = \frac{e^{-E(\mathbf{v})}}{Z}$$

Restricted Boltzmann Machine (RBM)

- General Boltzmann machines: all to all edges
- Restricted Boltzmann Machine
  - Nodes in same layer not connected



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

$$p(\mathbf{v},\mathbf{h}|\theta) = \frac{e^{-E_\theta(\mathbf{v},\mathbf{h})}}{Z}$$

Restricted Boltzmann Machine

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

$$p(\mathbf{v}|\theta) = \mathrm{Tr}_h\, p(\mathbf{v}, h|\theta) = \sum_h p(\mathbf{v}, h|\theta) = \frac{1}{Z} e^{-E_\theta(\mathbf{v}|\theta)}$$

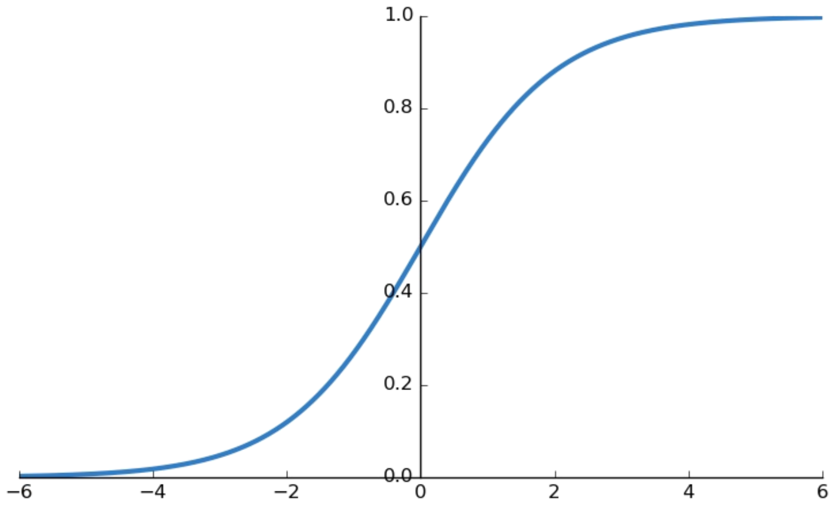
$$E_\theta(\mathbf{v}|\theta) = -\sum_j b_j v_j - \sum_i \log\left[1 + e^{c_i + \sum_j w_i v_j}\right]$$

Activation probabilities

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

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

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



Training

Maximize the Log-likelihood w.r.t. dataset S

$$\ln \mathcal{L}(\boldsymbol{\theta}|S) = \ln \prod_{i=1}^{\ell} p\left(\boldsymbol{x}_i|\boldsymbol{\theta}\right) = \sum_{i=1}^{\ell} \ln p\left(\boldsymbol{x}_i|\boldsymbol{\theta}\right)$$

Equivalent to minimization of the Kullback-Leibler divergence

$$\mathrm{KL}(q||p) = \sum_{\boldsymbol{x}\in S} q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} = \sum_{\boldsymbol{x}\in S} q(\boldsymbol{x}) \ln q(\boldsymbol{x}) - \sum_{\boldsymbol{x}\in S} q(\boldsymbol{x}) \ln p(\boldsymbol{x})$$

Use gradient ascent methods for optimization: need gradient

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

$$\frac{1}{|S|} \sum_{\mathbf{v}\in S} \frac{\partial \log p(\mathbf{v}|\theta)}{\partial w_{ij}} = \boxed{\mathbb{E}_{\text{ data }} \left[p\left(h_i = 1|\mathbf{v};\theta\right) v_j\right]} - \boxed{\mathbb{E}_{\text{ model }} \left[p\left(h_i = 1|\mathbf{v}';\theta\right) v_j'\right]}$$

Easy Hard

Importance sampling: contrastive divergence

Full Gibbs sampling of the Hard term can be expensive

Contrastive divergence:

$$p\left(h_i = 1|\mathbf{v}\right) v_j - \left\langle p\left(h_i = 1|\mathbf{v}'\right) v_j' \right\rangle_{p(\mathbf{v}')} \sim p\left(h_i = 1|\mathbf{v}^{(\mathbf{0})}\right) v_j^{(0)} - p\left(h_i = 1|\mathbf{v}^{(\mathbf{k})}\right) v_j^{(k)}$$

- k : Gibbs sampling step, typically set $k = 1$
- Initialised with a training example $\mathbf{v}^{(0)}$
- Each step t involves sampling $h(t) \sim p(h_i = 1|\, v(t); \theta)$, then sampling $v(t+1) \sim p(v_j = 1|\, h(t); \theta)$
- Other algorithms: Persistent CD, parallel tempering

Annealed importance sampling

To monitor the log-likelihood we need an estimate of the partition function Z

Target distribution: $p_1(\mathbf{v}) = p_1^*(\mathbf{v})/Z_1$

Simpler distribution: $p_0(\mathbf{v}) = \frac{1}{Z_0} p_0^*(\mathbf{v})$

Estimate by importance sampling $Z_1 = \int p_1^*(\mathbf{v}) d\mathbf{v} = Z_0 \int p_0(\mathbf{v}) \frac{p_1^*(\mathbf{v})}{p_0^*(\mathbf{v})} d\mathbf{v}$

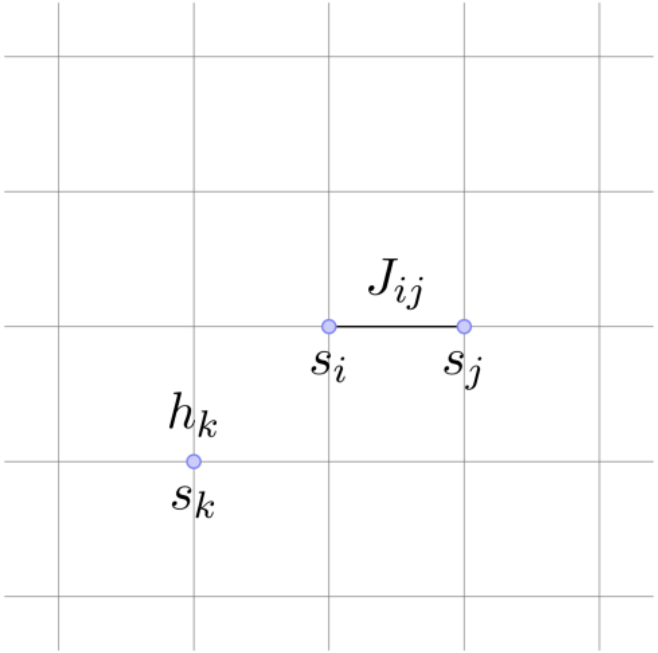
Can introduce biases, thus estimate by successive ratios

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

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

Ising model

- Nearest neighbour couplings
- Onsager 1944, analytical solution for 2D. 2nd order phase transition.



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

Generate dataset with standard MonteCarlo Markov Chain at various temperatures

Literature review

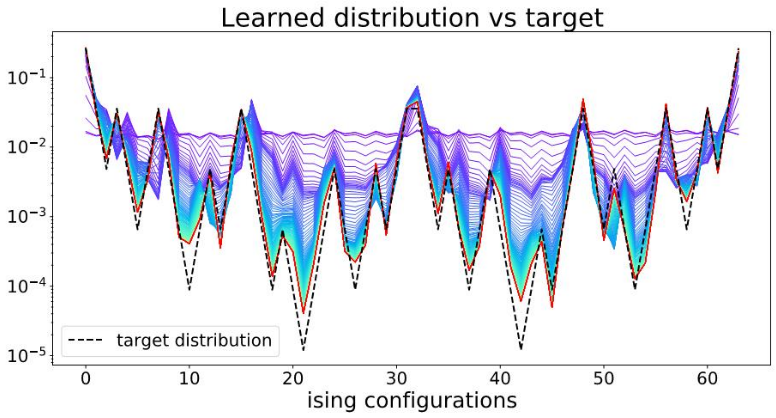
An exact mapping between the variational renormalization group and deep learning, P. Mehta and D. J. Schwab [2014]

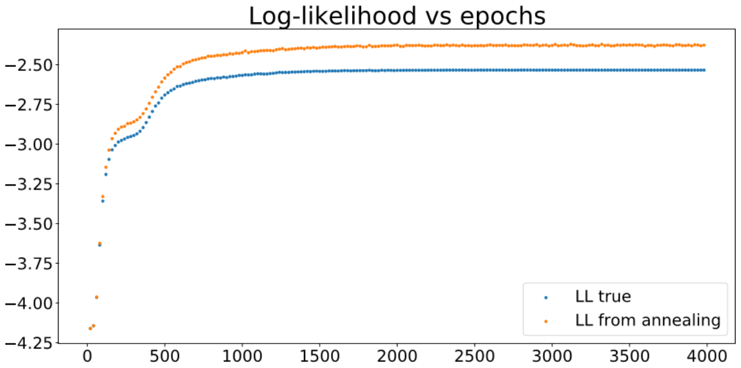
Learning thermodynamics with Boltzmann machines, G. Torlai and R. G. Melko [2016]

Deep Learning the Ising Model Near Criticality, A.Morningstar and R. G. Melko [2017]

Validation in 1D

6 spins: 2^6=64 states





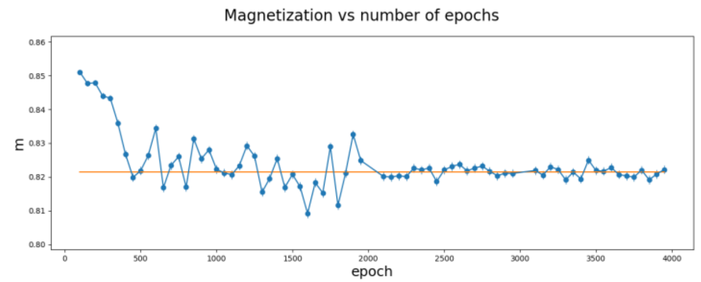
Ising 2D observables

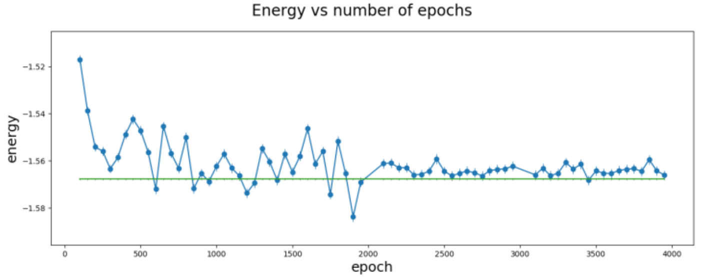
$$\langle m \rangle = \frac{1}{L^2} \left\langle \left| \sum_{i=1}^{L^2} s_i \right| \right\rangle$$ Magnetization

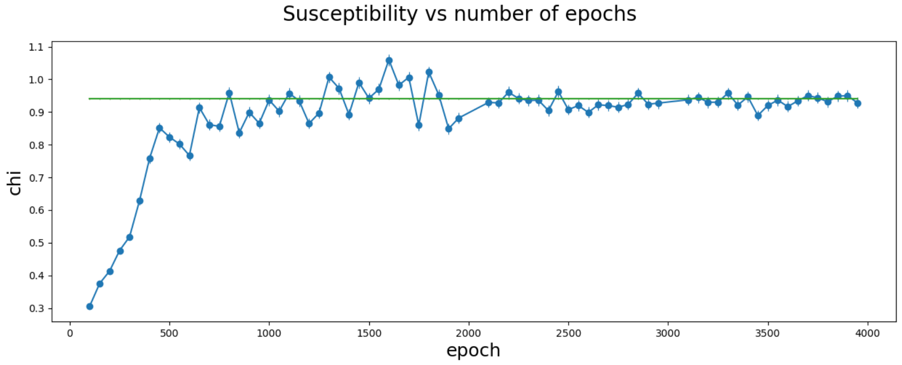
$$\langle \chi \rangle = \frac{L^2}{T} \left(\langle m^2 \rangle - \langle m \rangle^2 \right)$$ Susceptibility

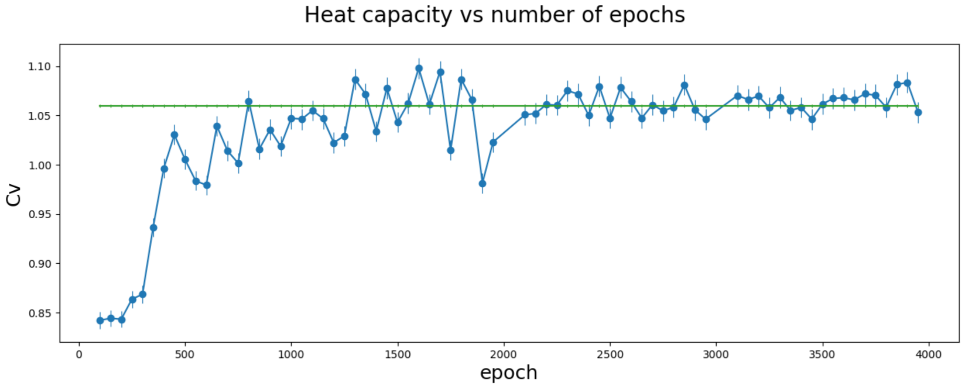
$$\langle E \rangle = -\frac{1}{L^2} \left\langle \sum_{\langle i,j \rangle} s_i s_j \right\rangle$$ Total energy

$$\langle c_v \rangle = \frac{L^2}{T^2} \left(\langle E^2 \rangle - \langle E \rangle^2 \right)$$ Specific heat



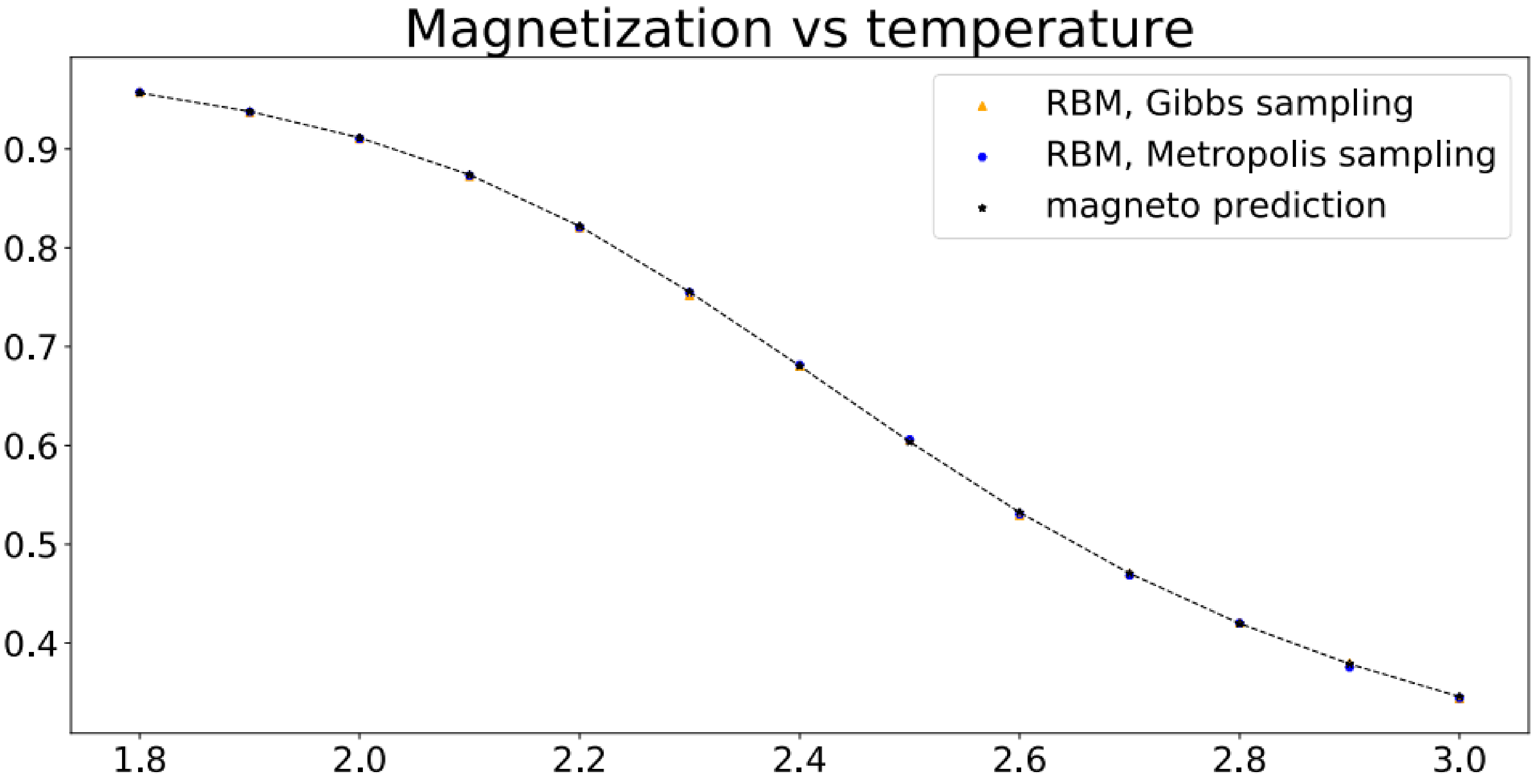






Plots for T= 2.2 (8^2 lattice)

RBM vs temperature



N-point interactions

Observe that in general $\mathbf{E(v)}$ can be rewritten via cumulants

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

If we recover the Ising model we can read all the couplings in this way