Translation invariant RBM for generating Ising configurations

Kipton Barros

March 13, 2022

Following notation from "Learning Thermodynamics with Boltzmann Machines", Torlai and Melko (2016).

Probability of an Ising configuration $[\sigma_1, \sigma_2 \dots \sigma_N]$ follows a Boltzmann distribution $P[\sigma, h] \propto \exp(-E[\sigma, h])$, where

$$E[\sigma, h] = -\sum_{i=1}^{N} \sum_{\alpha=1}^{M} W_{i\alpha} \sigma_i h_\alpha - \sum_{i=1}^{N} b_i \sigma_i - \sum_{\alpha=1}^{M} c_\alpha h_\alpha. \tag{1}$$

The h_α represent activations in a "hidden" layer. I purposely used a greek symbol α to index components of the hidden layer, distinct from roman symbol i which indices sites. It is likely that there will be many more hidden layer components than sites, $M \gg N$.

RBM provides an approach to jointly sample (σ, h), and we really only care about the σ samples (i.e., we can discard the h data).

A problem with Eq.(1) is that it does not capture the translation invariance of the Ising model. The problem is the explicit site dependence i appearing in both b_i and $W_{i\alpha}$. Ideally, all sites i should be on equal statistical footing in the model. What we *may* have, however, are weights W that depend on the "vector displacements" between lattice sites.

To build a model that is explicitly translation invariant, we extend the hidden layer to have an explicit site dependence, $h_\alpha \to h_{i,\alpha}$, and we define weights $W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha}$ that depend on the vector displacement $\boldsymbol{i} - \boldsymbol{j}$ between sites. The bold symbol $\boldsymbol{i}$ is intended to denote the coordinates of site index i. On a 2D lattice, one should think of $(\boldsymbol{i} - \boldsymbol{j})$ as a 2-component displacement vector. The energy becomes

$$E[\sigma, h] = -\sum_{i,j,\alpha} W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} \sigma_i h_{j,\alpha} - B \sum_i \sigma_i - \sum_{i,\alpha} c_\alpha h_{i,\alpha}.$$

Note that we have replaced the bias coefficients b_i with a single scalar B that can be physically interpreted as the external field.

This model is explicitly translation invariant, in that if we shift:

$$h_{i,\alpha} \to h'_{i,\alpha} = h_{i+\Delta,\alpha}$$
$$\sigma_i \to \sigma_i^{'} = \sigma_{i+\Delta},$$

then the energy is unchanged, $E[\sigma', h'] = E[\sigma, h]$.

Typically a lattice model will have symmetries beyond translation invariance. For the Ising model on the square lattice, there is also a 4-fold rotational invariance, plus a reflection invariance. Both of these symmetries can be incorporated by further constraining the weight matrix. Specifically, we can use weight tying, such that reflection invariance is encoded as

$$W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} = W_{(\boldsymbol{j}-\boldsymbol{i}),\alpha}.$$

Rotational invariance requires

$$W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} = W_{R(\boldsymbol{i}-\boldsymbol{j}),\alpha},$$

where R is a 2×2 matrix representing a 90 degree rotation in the plane, i.e.

$$R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The great advantage of encoding symmetries into the model structure is that much less training data is required to reach the same accuracy—we don't need data to "teach" the model about translation, rotation, and reflection invariance.