Derivation of an Effective Visible Energy in a Symmetry-Encoding RBM for the Ising Model

Emanuel Casiano-Diaz

March 24, 2022

Previously, Kipton discussed an energy model for an RBM-like structure that respects symmetries in the two-dimensional Ising model. This energy model is:

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

We are interested in obtaining a formula for the effective visible energy of a spin configuration, also known as a "free-energy" in the literature. The first step, is to marginalize out the hidden layer dependence of the full probability distribution:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}}\sum_{\boldsymbol{h}} P(\boldsymbol{\sigma},\boldsymbol{h}) = \frac{1}{\mathcal{Z}}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{\sigma},\boldsymbol{h})}$$

Substituting the symmetry-encoding energy into the exponent:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}}\sum_{\boldsymbol{h}} e^{\sum_{i,j,\alpha} W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha}\sigma_i h_{i,\alpha} + B\sum_i \sigma_i + \sum_{i,\alpha} c_\alpha h_{i,\alpha}}$$

The second term in the exponent is independent of the hidden layer. As such, the part of the exponential depending on this term can be factored out from the sum over hidden vectors:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i}\sum_{\boldsymbol{h}}[e^{\sum_{i,j,\alpha} W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha}\sigma_i h_{i,\alpha} + \sum_{i,\alpha} c_\alpha h_{i,\alpha}}]$$

The summations over i and α in the exponent can be factored out, alongside the hidden neuron value $h_{i,\alpha}$, which only depends on these two indices:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i}\sum_{\boldsymbol{h}}[e^{\sum_{i,\alpha} h_{i,\alpha} z_{i,\alpha}}]$$

where the substitution $z_{i,\alpha} \equiv c_\alpha + \sum_j W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha}\sigma_i$ has been performed. Letting the total number of neurons in the hidden layer be M and recalling that each of these neurons can only take on the values 1 or 0, the sum can be rewritten as:

$$\sum_{\boldsymbol{h}} \rightarrow \prod_{\boldsymbol{i}} \prod_{\alpha=1}^{M} \sum_{h_{i,\alpha}=0}^{1}$$

Additionally, recall that an exponential raised to a sum is the same as a product of exponentials raised to each term:

$$e^{\sum_i a_i} = \prod_i e^{a_i}$$

Using the two lines above, the marginalized probability distribution becomes:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i} [\prod_i \prod_{\alpha=1}^{M} \sum_{h_{i,\alpha}=0}^{1}][\prod_i \prod_{\alpha=1}^{M} e^{h_{i,\alpha} z_{i,\alpha}}] = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i} [\prod_i \prod_{\alpha=1}^{M} \sum_{h_{i,\alpha}=0}^{1} e^{h_{i,\alpha} z_{i,\alpha}}]$$

Since the hidden neuron values are either 0 or 1, the summation over $h_{i,\alpha}$ can be easily expanded:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i} \prod_i \prod_{\alpha=1}^{M} (1 + e^{z_{i,\alpha}})$$

Taking the natural logarithm of the factor in the parentheses and then exponentiating:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i} \prod_{\boldsymbol{i}} \prod_{\alpha=1}^{M} e^{\ln(1+e^{z_{i,\alpha}})} = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i} e^{\sum_{i,\alpha} \ln(1+e^{z_{i,\alpha}})}$$

Combining the two exponentials into one and recalling that $z_{i,\alpha} = c_\alpha + \sum_j W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} \sigma_i$:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{B\sum_i \sigma_i + \sum_{i,\alpha} \ln(1+e^{c_\alpha + \sum_j W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} \sigma_i})}$$

Notice that the exponent of the marginalized probability distribution now depends only on the spins $\boldsymbol{\sigma}$:

$$P(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} e^{-\varepsilon(\boldsymbol{\sigma})}$$

Where $\varepsilon(\boldsymbol{\sigma})$ is the effective visible energy of the spins, which in this case are the visible layer, and it is:

$$\varepsilon(\boldsymbol{\sigma}) = -B\sum_i \sigma_i - \sum_{i,\alpha} \ln(1 + e^{c_\alpha + \sum_j W_{(\boldsymbol{i}-\boldsymbol{j}),\alpha} \sigma_i})$$

This is also known as the "free energy" and it will be used to track the training of the model, by computing the difference between input and reconstructed spins, also known as the reconstruction error.