

# Free energy of convolutional RBM

Emanuel Casiano-Diaz

July 5, 2022

In Ref. [1], the authors propose a convolutional Restricted Boltzmann Machine (CRBM) to generate samples of physical models, such as Ising and Kitaev models. The energy of the CRBM is:

$$E(v, h) = - \sum_{k,i,j} h_{ij}^k (W^k * v)_{ij} - \sum_k h_{\text{bias}}^k \sum_{i,j} h_{ij}^k - v_{\text{bias}} \sum_{i,j} v_{ij} \quad (1)$$

where  $i, j$  are row and column indices and  $k$  is a kernel (or convolutional filter) index. The operation  $(W^k * v)$  is a two-dimensional convolution between the  $k$ -th kernel and the visible layer. The visible neuron values are  $v_{ij} = 0, 1$  and the hidden neuron values are  $h_{ij}^k = 0, 1$ . Note that in this model there are  $k$  layers of hidden neurons. The visible layer bias is  $v_{\text{bias}}$ , which is constant, and the hidden layer bias, which depends on the hidden layer, is  $h_{\text{bias}}^k$ .

The authors propose that, by summing over the hidden layer in Eq. (1), the following free-energy of the visible layer is obtained:

$$F(v) = -v_{\text{bias}} \sum_{i,j} v_{ij} - \sum_{k,i,j} \log(1 + e^{(v * W^k)_{ij} + h_{\text{bias}}^k}) \quad (2)$$

The goal of the CRBM training is to obtain network weights and biases that make Eq. (2) equal to the physical energy of the desired model. Here, we want to derive in detail how to get Eq. (2) from Eq. (1).

## Derivation

The weights of CRBM configurations follow a Boltzmann distribution:

$$P(v, h) = \frac{1}{\mathcal{Z}_{\text{CRBM}}} e^{-E(v, h)} \quad (3)$$

where  $\mathcal{Z}_{\text{CRBM}}$  is the normalization constant for the CRBM model (which in general is not equal to the normalization constant of the physical model). The marginalized probability of the visible layer can be obtained by summing over the hidden layer degrees of freedom:

$$\begin{aligned} \mathcal{Z}_{\text{CRBM}} \cdot P(v) &= \sum_h P(v, h) \\ &= \sum_h e^{\sum_{k,i,j} h_{ij}^k (W^k * v)_{ij} + \sum_k h_{\text{bias}}^k \sum_{i,j} h_{ij}^k + v_{\text{bias}} \sum_{i,j} v_{ij}} \end{aligned} \quad (4)$$

The last term in the exponent is independent of the hidden layer and the corresponding exponential can be factored out:

$$\mathcal{Z}_{\text{CRBM}} \cdot P(v) = e^{v_{\text{bias}} \sum_{i,j} v_{ij}} \sum_h e^{\sum_{k,i,j} h_{ij}^k (W^k * v)_{ij} + \sum_{k,i,j} h_{\text{bias}}^k h_{ij}^k} \quad (5)$$

Recalling that the hidden neurons can take values  $h_{ij}^k = 0, 1$ , the summation over hidden layer degrees of freedom can be written more explicitly as:

$$\sum_h \rightarrow \prod_k \prod_i \prod_j \sum_{h_{ij}^k=0}^1 \quad (6)$$

Additionally, recall that the exponential of a sum is equal to a product of exponentials. Thus Eq. (5) can be rewritten as:

$$\begin{aligned} \mathcal{Z}_{\text{CRBM}} \cdot P(v) &= e^{v_{\text{bias}} \sum_{i,j} v_{ij}} \left[ \prod_k \prod_i \prod_j \sum_{h_{ij}^k=0}^1 \right] \left[ \prod_k \prod_i \prod_j \right] e^{h_{ij}^k (W^k * v)_{ij} + h_{\text{bias}}^k h_{ij}^k} \\ &= e^{v_{\text{bias}} \sum_{i,j} v_{ij}} \left[ \prod_k \prod_i \prod_j \sum_{h_{ij}^k=0}^1 \right] e^{h_{ij}^k (W^k * v)_{ij} + h_{\text{bias}}^k h_{ij}^k} \end{aligned} \quad (7)$$

where the identity  $\prod_x f_x \prod_x g_x = \prod_x f_x g_x$  has been used. Expanding over the hidden neuron values:

$$\begin{aligned} \mathcal{Z}_{\text{CRBM}} \cdot P(v) &= e^{v_{\text{bias}} \sum_{i,j} v_{ij}} \left[ \prod_k \prod_i \prod_j \right] \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right) \\ &= e^{v_{\text{bias}} \sum_{i,j} v_{ij}} \left[ \prod_k \prod_i \prod_j \right] e^{\log \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right)} \\ &= e^{v_{\text{bias}} \sum_{i,j} v_{ij}} e^{\sum_{k,i,j} \log \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right)} \\ \mathcal{Z}_{\text{CRBM}} \cdot P(v) &= e^{v_{\text{bias}} \sum_{i,j} v_{i,j} + \sum_{k,i,j} \log \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right)} \\ \implies P(v) &= \frac{1}{\mathcal{Z}_{\text{CRBM}}} e^{v_{\text{bias}} \sum_{i,j} v_{i,j} + \sum_{k,i,j} \log \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right)} \end{aligned} \quad (8)$$

The negative of the exponent is the free-energy of the visible layer:

$$F(v) = -v_{\text{bias}} \sum_{i,j} v_{ij} - \sum_{k,i,j} \log \left( 1 + e^{(W^k * v)_{ij} + h_{\text{bias}}^k} \right) \quad (9)$$

which is the same as the free-energy in Eq. (2) ■

## References

- [1] Daniel Alcalde Puente and Ilya M. Eremin. Convolutional restricted boltzmann machine aided monte carlo: An application to ising and kitaev models. *Physical Review B*, 102(19), nov 2020.