

Restricted Boltzmann Machine Wavefunction and its Derivatives

Khaldoon Ghanem

May 11, 2018

In Restricted Boltzmann Machines (RBM), we have one input/physical layer and one hidden/virtual layer. Each node of a layer is connected to all nodes of the other layer while not connected (directly) to any node of the same layer.

The nodes of the input layer s_i represent the spin-orbitals of a determinant where a value of $+1$ means an occupied orbital and value of -1 represents an empty one. So in total we have M input nodes, where M is the size of the single particle basis set.

The nodes of the hidden layer v_i represent virtual spin-orbitals that couple the physical orbitals of a determinant together. The number of virtual orbitals P is a parameter of the network that controls the amount of correlation present in the wavefunction. With more virtual orbitals, the network should be able to capture more of the correlation energy.

The RBM wavefunction is an exponential function of some virtual energy which is defined as a bilinear function of the physical and virtual orbitals:

$$\psi(S, V; A, B, W) = \exp \left(\sum_j \alpha_j s_j + \sum_i \beta_i v_i + \sum_{i,j} w_{i,j} v_i s_j \right)$$

where A, B and W are parameters to be trained. W represents the weights between physical and virtual orbitals, while A and B represent the offsets of the input and hidden layers, respectively.

To obtain the coefficient of a determinant, we need to keep only the physical

orbitals and sum over all the virtual ones

$$\begin{aligned}
\psi(S; A, B, W) &= \sum_V \psi(S, V; A, B, W) \\
&= \sum_{v_1=\pm 1, v_2=\pm 1, \dots} \exp \left(\sum_j \alpha_j s_j + \sum_i \beta_i v_i + \sum_{i,j} w_{i,j} v_i s_j \right) \\
&= \exp \left(\sum_j \alpha_j s_j \right) \sum_{v_1=\pm 1, v_2=\pm 1, \dots} \exp \left[\sum_i \beta_i v_i + \sum_{i,j} w_{i,j} v_i s_j \right] \\
&= \exp \left(\sum_j \alpha_j s_j \right) \prod_{i=1}^P \left[\sum_{v_i=\pm 1} \exp \left(\sum_i \beta_i v_i + \sum_{i,j} w_{i,j} v_i s_j \right) \right] \\
&= \exp \left(\sum_j \alpha_j s_j \right) \prod_{i=1}^P \left[2 \cosh \left(\beta_i + \sum_j w_{i,j} s_j \right) \right]
\end{aligned}$$

To train the network, we should be able to take the derivative of the above expression with respect to the different parameters

$$\begin{aligned}
\frac{\delta \psi}{\delta \alpha_l} &= s_l \exp \left(\sum_j \alpha_j s_j \right) \prod_{i=1}^P \left[2 \cosh \left(\beta_i + \sum_j w_{i,j} s_j \right) \right] \\
&= s_l \psi(S; A, B, W)
\end{aligned}$$

$$\begin{aligned}
\frac{\delta \psi}{\delta \beta_k} &= \exp \left(\sum_j \alpha_j s_j \right) \left[2 \sinh \left(\beta_k + \sum_j w_{k,j} s_j \right) \right] \prod_{i=1, i \neq k}^P \left[2 \cosh \left(\beta_i + \sum_j w_{i,j} s_j \right) \right] \\
&= \frac{\sinh \left(\beta_k + \sum_j w_{k,j} s_j \right)}{\cosh \left(\beta_k + \sum_j w_{k,j} s_j \right)} \psi(S; A, B, W) \\
&= \tanh \left(\beta_k + \sum_j w_{k,j} s_j \right) \psi(S; A, B, W)
\end{aligned}$$

$$\begin{aligned}
\frac{\delta \psi}{\delta w_{k,l}} &= \exp \left(\sum_i \alpha_i s_i \right) \left[2 s_l \sinh \left(\beta_k + \sum_j w_{k,j} s_j \right) \right] \prod_{i=1, i \neq k}^P \left[2 \cosh \left(\beta_i + \sum_j w_{i,j} s_j \right) \right] \\
&= s_l \frac{\sinh \left(\beta_k + \sum_j w_{k,j} s_j \right)}{\cosh \left(\beta_k + \sum_j w_{k,j} s_j \right)} \psi(S; A, B, W) \\
&= s_l \tanh \left(\beta_k + \sum_j w_{k,j} s_j \right) \psi(S; A, B, W)
\end{aligned}$$