Notes to check sampling

Vilde Flusgrud

February 6, 2018

1 Restricted Boltzmann Machine (RBM)

The joint probability distribution is defined as [3]

$$F_{rbm}(\mathbf{X}, \mathbf{H}) = \frac{1}{Z} e^{-\frac{1}{T_0} E(\mathbf{X}, \mathbf{H})} \tag{1}$$

where

$$Z = \int \int \frac{1}{Z} e^{-\frac{1}{T_0} E(\mathbf{x}, \mathbf{h})} \, d\mathbf{x} \, d\mathbf{h} \tag{2}$$

It is common to ignore T_0 by setting it to one.

2 Gaussian-Binary RBM

Here we have [3]

$$E(\mathbf{X}, \mathbf{H}) = \sum_i^M \frac{(X_i - a_i)^2}{2\sigma_i^2} - \sum_j^N b_j H_j - \sum_{i,j}^{M,N} \frac{X_i w_{ij} H_j}{\sigma_i^2} \tag{3}$$

If $\sigma_i = \sigma$ then

$$E(\mathbf{V}, \mathbf{H}) = \frac{||\mathbf{X} - \mathbf{a}||^2}{2\sigma^2} - \mathbf{b}^T \mathbf{H} - \frac{\mathbf{X}^T \mathbf{W} \mathbf{H}}{\sigma^2} \tag{4}$$

OBS: in the eq above, check the factor two's in the denominator with σ.

3 The Wave Function

To find the marginal probability $F_{rbm}(X)$ we set:

$$F_{rbm}(\mathbf{X}) = \sum_{\mathbf{h}} F_{rbm}(\mathbf{X}, \mathbf{h}) \tag{5}$$

$$= \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{X}, \mathbf{h})} \tag{6}$$

This is used to represent the wave function:

$$\Psi(\mathbf{X}) = F_{rbm}(\mathbf{X}) \tag{7}$$

$$= \frac{1}{Z} \sum_{\{h_j\}} e^{-E(\mathbf{X},\mathbf{h})} \tag{8}$$

$$= \frac{1}{Z} \sum_{\{h_j\}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2} + \sum_j^N b_j h_j + \sum_{i,j}^{M,N} \frac{X_i w_{ij} h_j}{\sigma^2}} \tag{9}$$

$$= \frac{1}{Z} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_j^N (1 + e^{b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}}) \tag{10}$$

$$\tag{11}$$

3.1 Alternative 1: $\psi = \sqrt{F_{rbm}}$

This is used by Carleo in his seminar lectures for simple systems where the ground state wave function is positive definite and Gibbs sampling is used rather than Metropolis. (Though by mistake he uses this only in the conditional probabilites used in the Gibbs sampling, while the wave function used in the local energy is still $\Psi(\mathbf{X}) = F_{rbm}(\mathbf{X})$ as in the previous section.

He says however in his email that the idea was to do the below version.)

$$|\Psi(\mathbf{X})|^2 = F_{rbm}(\mathbf{X}) \tag{12}$$

$$\Rightarrow \Psi(\mathbf{X}) = \sqrt{F_{rbm}(\mathbf{X})} \tag{13}$$

$$= \frac{1}{\sqrt{Z}} \sqrt{\sum_{\{h_j\}} e^{-E(\mathbf{X},\mathbf{h})}} \tag{14}$$

$$= \frac{1}{\sqrt{Z}} \sqrt{\sum_{\{h_j\}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2} + \sum_j^N b_j h_j + \sum_{i,j}^{M,N} \frac{X_i w_{ij} h_j}{\sigma^2}}} \tag{15}$$

$$= \frac{1}{\sqrt{Z}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{4\sigma^2}} \sqrt{\sum_{\{h_j\}} \prod_j^N e^{b_j h_j + \sum_i^M \frac{X_i w_{ij} h_j}{\sigma^2}}} \tag{16}$$

$$= \frac{1}{\sqrt{Z}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{4\sigma^2}} \sqrt{\prod_j^N \sum_{h_j} e^{b_j h_j + \sum_i^M \frac{X_i w_{ij} h_j}{\sigma^2}}} \tag{17}$$

$$= \frac{1}{\sqrt{Z}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{4\sigma^2}} \prod_j^N \sqrt{e^0 + e^{b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}}} \tag{18}$$

$$= \frac{1}{\sqrt{Z}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{4\sigma^2}} \prod_j^N \sqrt{1 + e^{b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}}} \tag{19}$$

$$\tag{20}$$

3.2 Alternative 2: $h_j \in \{-1, 1\}$, not $\{0, 1\}$

Used when representing the spin lattice system (intuitive because the visible spin values in that case take the values -1 and 1 as well):

$$\Psi(\mathbf{X}) = F_{rbm}(\mathbf{X}) \tag{21}$$

$$= \frac{1}{Z} \sum_{\{h_j\}} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2} + \sum_j^N b_j h_j + \sum_{i,j}^{M,N} \frac{X_i w_{ij} h_j}{\sigma^2}} \tag{22}$$

$$= \frac{1}{Z} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_j^N \sum_{h_j} e^{b_j h_j + \sum_i^M \frac{X_i w_{ij} h_j}{\sigma^2}} \tag{23}$$

$$= \frac{1}{Z} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_j^N (e^{-b_j - \sum_i^M \frac{X_i w_{ij}}{\sigma^2}} + e^{b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}}) \tag{24}$$

$$= \frac{1}{Z} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_j^N 2\cosh(b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}) \tag{25}$$

$$= \frac{1}{Z} e^{-\sum_i^M \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_j^N 2\cosh(b_j + \sum_i^M \frac{X_i w_{ij}}{\sigma^2}) \tag{26}$$

$$\tag{27}$$

3.3 Quantum Monte Carlo expressions

In order to optimize the RBM parameters α ($\mathbf{a}$, $\mathbf{b}$ and $\mathbf{W}$) we apply quantum Monte Carlo methods. Using the variational principle we approximate the ground state wave function by minimizing the local energy E_L. Its gradient with respect to the RBM parameters is given by [2]

$$G_i = \frac{\partial \langle E_L \rangle}{\partial \alpha_i} = 2(\langle E_L \frac{1}{\Psi} \frac{\partial \Psi}{\partial \alpha_i} \rangle - \langle E_L \rangle \langle \frac{1}{\Psi} \frac{\partial \Psi}{\partial \alpha_i} \rangle) \tag{28}$$

where $\alpha_i = a_1, ..., a_M, b_1, ..., b_N, w_{11}, ..., w_{MN}$.

The local energy is given by the expression

$$E_L = \frac{1}{\Psi} \hat{\mathbf{H}} \Psi \tag{29}$$

Where $\hat{\mathbf{H}}$ is the Hamiltonian, which is dependent on the physical system under investigation. The expectation value of E_L is obtained during sampling by

$$\langle E_L \rangle = \frac{1}{N} \sum_n^N (E_L)_n \tag{30}$$

where N is the number of samples.

We use that $\frac{1}{\Psi}\frac{\partial \Psi}{\partial \alpha_i} = \frac{\partial \ln \Psi}{\partial \alpha_i}$ and find

$$\ln \Psi(\mathbf{X}) = -\ln Z - \sum_m^M \frac{(X_m - a_m)^2}{2\sigma^2} + \sum_n^N \ln(1 + e^{b_n + \sum_i^M \frac{X_i w_{in}}{\sigma^2}}) \tag{31}$$

Giving

$$\frac{\partial}{\partial a_m} \ln \Psi = \frac{1}{\sigma^2}(X_m - a_m) \tag{32}$$

$$\frac{\partial}{\partial b_n} \ln \Psi = \frac{1}{e^{-b_n - \frac{1}{\sigma^2}\sum_i^M X_i w_{in}} + 1} \tag{33}$$

$$\frac{\partial}{\partial w_{mn}} \ln \Psi = \frac{X_m}{\sigma^2(e^{-b_n - \frac{1}{\sigma^2}\sum_i^M X_i w_{in}} + 1)} \tag{34}$$

4 The system and the Hamiltonian

4.1 The Harmonic Oscillator

The Hamiltonian for the Harmonic Oscillator system is [1]

$$\hat{\mathbf{H}} = \sum_p^P (-\frac{1}{2}\nabla_p^2 + \frac{1}{2}\omega^2 r_p^2) + \sum_{p<q} \frac{1}{r_{pq}} \tag{35}$$

where the first summation term represents the standard harmonic oscillator part and the latter the repulsive interaction between two electrons. Natural units ($\hbar = c = e = m_e = 1$) are used, and P is the number of particles.

Thus we get (D being the number of dimensions)

$$E = \frac{1}{\Psi}\mathbf{H}\Psi \tag{36}$$

$$= \frac{1}{\Psi}(\sum_p^P (-\frac{1}{2}\nabla_p^2 + \frac{1}{2}\omega^2 r_p^2) + \sum_{p<q} \frac{1}{r_{pq}})\Psi \tag{37}$$

$$= -\frac{1}{2}\frac{1}{\Psi}\sum_p^P \nabla_p^2\Psi + \frac{1}{2}\omega^2\sum_p^P r_p^2 + \sum_{p<q} \frac{1}{r_{pq}} \tag{38}$$

$$= -\frac{1}{2}\frac{1}{\Psi}\sum_p^P\sum_d^D \frac{\partial^2\Psi}{\partial x_{pd}^2} + \frac{1}{2}\omega^2\sum_p^P r_p^2 + \sum_{p<q} \frac{1}{r_{pq}} \tag{39}$$

$$= \frac{1}{2}\sum_p^P\sum_d^D(-(\frac{\partial}{\partial x_{pd}}\ln \Psi)^2 - \frac{\partial^2}{\partial x_{pd}^2}\ln \Psi + \omega^2 x_{pd}^2) + \sum_{p<q} \frac{1}{r_{pq}} \tag{40}$$

$$\tag{41}$$

5

Now if each visible node in the Boltzmann machine represents one coordinate of one particle, this can be written as

$$E = \frac{1}{2} \sum_m^M \left(-\left(\frac{\partial}{\partial v_m} \ln \Psi\right)^2 - \frac{\partial^2}{\partial v_m^2} \ln \Psi + \omega^2 v_m^2\right) + \sum_{p<q} \frac{1}{r_{pq}} \tag{42}$$

Where we have that

$$\frac{\partial}{\partial x_m} \ln \Psi = -\frac{1}{\sigma^2}(x_m - a_m) + \frac{1}{\sigma^2} \sum_n^N \frac{w_{mn}}{e^{-b_n - \frac{1}{\sigma^2} \sum_i^M x_i w_{in}} + 1} \tag{43}$$

$$\frac{\partial^2}{\partial x_m^2} \ln \Psi = -\frac{1}{\sigma^2} + \frac{1}{\sigma^4} \sum_n^N \omega_{mn}^2 \frac{e^{b_n + \frac{1}{\sigma^2} \sum_i^M x_i w_{in}}}{(e^{b_n + \frac{1}{\sigma^2} \sum_i^M x_i w_{in}} + 1)^2} \tag{44}$$

We now have all the expressions neeeded to calculate the gradient of the local energy with respect to the RBM parameters $\frac{\partial \langle E_L \rangle}{\partial \alpha_i}$.

5 Sampling

The goal is to sample positions $\mathbf{x}$ according to the probability density given by our current approximation of $|\Psi(\mathbf{x})|^2$.

5.1 Gibbs sampling

It was suggested by Carleo that this method should mainly used for simple systems where the wave function is positive definite. In this method we sample from the joint probability of $\mathbf{x}$ and $\mathbf{h}$, in the form of a two step sampling process. The samples $\mathbf{x}$ by themselves then model the probability density $|\Psi(\mathbf{x})|^2$ as we wish. The updated samples are generated according to the conditional probabilities $P(X_i|\mathbf{h})$ and $P(H_j|\mathbf{x})$ respectively and accepted with the probability of 1.

5.1.1 Conditionals when using $\Psi(x) = \sqrt{F_{rbm}}$

When the wave function is represented by $\Psi = \sqrt{F_{rbm}}$ as Carleo suggested for a simple system, the condition probabilities are

$$P(X_i|\mathbf{h}) = \mathcal{N}(X_i; a_i + \mathbf{w}_{i*}\mathbf{h}, \sigma^2) \tag{45}$$

$$P(\mathbf{X}|\mathbf{h}) = \prod_i^M \mathcal{N}(X_i; a_i + \mathbf{w}_{i*}\mathbf{h}, \sigma^2) \tag{46}$$

$$= \mathcal{N}(\mathbf{X}; \mathbf{a} + \mathbf{W}\mathbf{h}, \sigma^2) \tag{47}$$

$$P(H_j|\mathbf{x}) = \frac{e^{(b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2})H_j}}{\sum_{h_j} e^{(b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2})h_j}} \tag{48}$$

$$= \frac{e^{(b_j + \frac{\mathbf{v}^T \mathbf{w}_{*j}}{\sigma^2})H_j}}{1 + e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}} \tag{49}$$

$$P(\mathbf{H}|\mathbf{x}) = \prod_j^N \frac{e^{(b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2})H_j}}{1 + e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}} \tag{50}$$

$$\tag{51}$$

Meaning

$$P(H_j = 1|\mathbf{x}) = \frac{e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}}{1 + e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}} \tag{52}$$

$$= \frac{1}{1 + e^{-b_j - \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}} \tag{53}$$

$$P(H_j = 0|\mathbf{x}) = \frac{1}{1 + e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}}} \tag{54}$$

5.2 Metropolis-Hastings sampling

In the brute force version of this, we move each position x_i with a step size determined by a unifrom distribution. The change in the system is then accpeted with an acceptance probability

$$A(\mathbf{x} \rightarrow \mathbf{x}') = \min(1, |\frac{\Psi(\mathbf{x}')}{\Psi(\mathbf{x})}|^2) \tag{55}$$

References

[1] Morten Hjorh-Jensen. *Computational Physics Lecture Notes*. University of Oslo Department of Physics, 2015.

[2] Jørgen Høgberget. Quantum monte-carlo studies of generalized many-body systems. Master's thesis, University of Oslo, 2013.

[3] Jan Melchior, Nan Wang, and Laurenz Wiskott. Gaussian-binary restricted boltzmann machines on modeling natural image statistics. *PLOS ONE*, 2017.