Log-likelihood computation for restricted Boltzmann machines (RBMs)

Marco Baiesi

March 5, 2025

The log-likelihood $\ell_\theta(x) = \ln p_\theta(x)$ per data point x, averaged over M data points, gives the log-likelihood of data,

$$\mathcal{L} = \frac{1}{M} \sum_{m \le M} \ell_\theta(x^{(m)})$$

Knowing $\mathcal{L}$ is useful to assess the goodness of the model.

In the case of RBMs, we have $p(x,z) = e^{-E(x,z)}/Z$, $p(x) = \sum_z p(x,z)$, and thus

$$\ell_\theta(x) = \ln \sum_z e^{-E(x,z)} - \ln \underbrace{\sum_{x'} \sum_z e^{-E(x',z)}}_{\text{partition function } Z} \tag{1}$$

The main difficulty in computing a $\ell_\theta(x)$ is summing up the Boltzmann weights of all possible configurations in Z. With D visible units and L hidden units, there are 2^{D+L} possible configurations. In our example with $D = 784$ and $L = 12$, the computation of 2^{796} weights is unfeasible. However, we have a helpful formula for the sums for the two-level variables $x_i, z_\mu \in \{0,1\}$. It takes advantage of the energy function

$$\begin{aligned} E(x,z) &= -\sum_i a_i x_i - \sum_\mu b_\mu z_\mu - \sum_i \sum_\mu x_i w_{i\mu} z_\mu \\ &= -\sum_i H_i(z) x_i - \sum_\mu b_\mu z_\mu \end{aligned}$$

with

$$H_i(z) = a_i + \sum_\mu w_{i\mu} z_\mu$$

and of the corresponding Boltzmann weight

$$e^{-E(x,z)} = \underbrace{\prod_\mu e^{b_\mu z_\mu}}_{G(z)} \prod_i e^{H_i(z) x_i}$$

Fixing z and defining a reduced partition function $Z(z) = \sum_x e^{-E(x,z)}$, we progressively sum its contributions,

$$\begin{aligned}
Z(z) &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_M} e^{-E(x,z)} \\
&= G(z)\sum_{x_1}\sum_{x_2}\cdots\sum_{x_M}\prod_i e^{H_i(z)x_i} \\
&= G(z)(1+e^{H_1(z)})\sum_{x_2}\cdots\sum_{x_M}\prod_{i\geq 2} e^{H_i(z)x_i} \\
&= G(z)(1+e^{H_1(z)})(1+e^{H_2(z)})\sum_{x_3}\cdots\sum_{x_M}\prod_{i\geq 3} e^{H_i(z)x_i} \\
&= \cdots \\
&= G(z)\prod_i (1+e^{H_i(z)})
\end{aligned}$$

Such quantity is easy to compute, if one avoids overflows from the multiplication of many numbers $1+e^{H_i(z)}$ larger than 1. Let us introduce a number q close to the average value of $1+e^{H_i(z)}$ over i's, and define

$$\tilde{Z}(z) = q^{-D} Z(z) = G(z)\prod_{i=1}^{D}\frac{1+e^{H_i(z)}}{q}$$

which remains stable, as it multiplies numbers of order 1. We therefore replace $Z = q^D \tilde{Z}$ in the formulas.

Since L is small in our RBMs, we can thus compute a sum over 2^L possible z's, specifically

$$\begin{aligned}
Z &= \sum_z Z(z) \\
&= q^D \sum_z G(z)\prod_{i=1}^{D}\frac{1+e^{H_i(z)}}{q}
\end{aligned}$$

which leads to

$$\ln Z = D\ln q + \ln\left[\sum_z G(z)\prod_{i=1}^{D}\frac{1+e^{H_i(z)}}{q}\right]$$

This expression can be used in (1) to compute the single log-likelihood. Its average over $x^{(m)}$'s in the dataset gives $\mathcal{L}$.

Note that this derivation works for the Bernoulli variables $\{0,1\}$. It needs a modification for the spin variables $\{-1,1\}$.