

## GROUP2505 – Log-likelihood analysis for Restricted Boltzmann Machines

Elias Maria Bonasera, Alberto Casellato, Nicola Garbin, and Francesco Pazzocco (781238?)  
(Dated: March 25, 2025)

[illegible]

## 1. INTRODUCTION

The main goal of this assignment is to explore how the performance of an RBM changes for different choices of the hyperparameters of the model, using the MNIST digits as the database; in particular using the log-likelihood evaluation we explore the trend of the model during the learning process as the number of epochs increases.

Restricted Boltzmann Machines (RBM) are a powerful kind of generative models designed to accomplish training processes relatively fast. In RBMs, a set of  $D$  binary visible units  $i$  of state  $v_i$  is symmetrically connected to a set of  $L$  binary hidden units  $\mu$  of state  $h_\mu$ ; the continuous weight  $w_{i\mu}$  quantifies the strength between unit  $i$  and unit  $\mu$  (see Figure 1). RBMs use an energy function to define the probability distribution over the input data. In the training process the energy of the configuration is minimized by adjusting the parameters  $\theta$ . The most common training algorithm is contrastive divergence which allows to approximate the gradient of the likelihood to update the parameters. During this process, a cyclic Gibbs sampling is performed setting the visible units given the hidden ones and vice versa, according to the following probabilities:

$$p(h_\mu = 1 \mid \mathbf{v}) = \sigma(b_\mu + \sum_i v_i w_{i\mu}) \quad (1)$$

$$p(v_i = 1 \mid \mathbf{h}) = \sigma(a_i + \sum_{\mu} h_{\mu} w_{i\mu}) \quad (2)$$

where  $\sigma(x) = 1/(1 + e^{-x})$  is the logistic sigmoid function,  $a_i$  is the bias of the  $i$ -th visible unit and  $b_\mu$  is the bias of the  $\mu$ -th hidden unit; they act shifting the sigmoid function  $\sigma(x)$ . The absence of links among units of the same type simplifies the training process. Moreover, the number of iterations of Eq. 1 and Eq. 2 can be set to 1 if real data is used to fix  $\mathbf{v}$  in the first place.

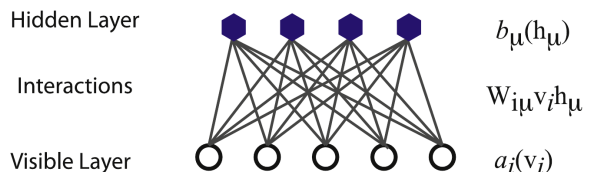


FIG. 1. A Restricted Boltzmann Machine (RBM) is made up of visible units, denoted as  $v_i$ , and hidden units, represented as  $h_\mu$ . These units engage with one another through interactions characterized by the weights  $W_{i\mu}$ . Notably, there are no direct interactions among the visible units or among the hidden units themselves.

The goodness of the models is evaluated by computing the log-likelihood  $\mathcal{L}$  of the data:

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^M l_{\theta}(v^{(m)}) \quad (3)$$

$$l_{\theta}(v^{(m)}) = \log \sum_h e^{-E(v,h)} - \log Z \quad (4)$$

where  $M$  is the number of data points,  $Z$  is the partition function,  $E(v, h) = G(h) \prod_i e^{H_i(h)v_i}$  is the energy function, with  $G(h) = \prod_\mu e^{b_\mu h_\mu}$  and  $H_i(h)$  defined as  $H_i(h) = a_i + \sum_\mu w_{i\mu} h_\mu$ , and, in Eq. 4, the index  $h$  of the summation represents each possible state of the hidden layer.

## 2. METHODS

### 2.1. Theoretical implementation

Regarding the computation of the partition function  $Z$  (in Eq. 4), its implementation depends on the set of values which each unit can be set on. Since the direct computation of the  $Z$  function is unfeasible, due to the summation of on all possible state combinations of the system, a possible approach is to split the summation separately over  $v$  and  $h$ .



ZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZ  
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZ.

[illegible]

ZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ  
Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZ ZZZZZ  
ZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZ  
ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ  
Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ. ZZZZZZZZZZZZZZZZ ZZZZZ  
ZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZ

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ  
Z ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZ ZZZZZ  
ZZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ  
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z  
ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ  
Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ  
ZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ ZZZZ ZZZZZZZZZ  
ZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z  
ZZZZZZZZZZ ZZZZ ZZZZZZZZZ.

Discuss the key aspects that we can take home from this work.

Check if your text is light, swift, and correct in exposing its passages.

[illegible][illegible]

ZZZZZZZZZZZZZZZZZZZZ ZZZZZZ ZZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ  
ZZZZZ ZZZZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZZZ ZZZZZZ ZZZZZZZZZZZZ ZZZZZZZZ Z  
ZZZZZZZZZZZZZZZZZZZZ ZZZZZZ ZZZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
 ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ  
 Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZ ZZZZZ  
 ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ  
 ZZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
 ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z  
 ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZZ ZZZZZZZ Z ZZZZZZZZZZZ  
ZZZZ ZZZZZZZZZZZ ZZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ Z  
ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZ.

- 
- [1] B. Franklin, J. Here There **10**, 20–40 (1800).  
[2] A. Einstein, Int. J. There Here **20**, 125–133 (1910).

# Assignment score grid

<b>Structure:</b> the exposition follow a logic order	8
<b>Clarity:</b> the text is brief enough, avoids complicated sentences and specifies all concepts and links	8
<b>Depth:</b> the text is not a shallow repetition of notions, there emerges a good understanding	8
<b>Rigor:</b> the analysis of the results is precise, quantitatively, and convincing	8
<b>Innovation:</b> new methods/ideas are introduced; conclusions beyond what introduced in the class	4