GROUP2505 – Log-likelihood analysis for Restricted Boltzmann Machines

Elias Maria Bonasera, Alberto Casellato, Nicola Garbin, and Francesco Pazzocco
(Dated: March 26, 2025)

Here, in the "abstract" (more or less of 8 lines), there should be a short summary of the work and of its main findings. In a paper, the abstract is important also for attracting potential readers, hence it is convenient to start it with some catchy statement. — The rest of this text has the double purpose of (a) providing a template for the assignment latex, and (b) introducing how to structure the backbone of the text and explaining some details of this assignment. The "zzz" fill the space to show a typical (but not rigid) extension of the parts. zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

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 μ of state h_μ; the continuos weight $w_{i\mu}$ quantifies the strength between unit i and unit μ (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 θ. 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:

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

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

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_μ is the bias of the μ-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 setted to 1 if real data is used to fix $\mathbf{v}$ in the first place.

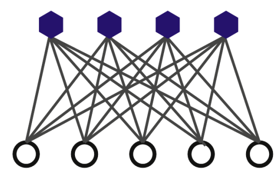

FIG. 1. A Restricted Boltzmann Machine (RBM) is made up of visible units, denoted as v_i, and hidden units, represented as h_μ. These units engage with one another through interactions characterized by the weights $W_{i\mu}$. Notably, there are no direct interactions among the visible units or among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

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

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

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

where M is the number of data points, Z is the partition function, $e^{-E(v,h)} = G(h) \prod_i e^{H_i(h)v_i}$ 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.

For the Bernulli case the units can assume the values $\{0,1\}$ and the partition function form can be manipulated to obtain the following expression:

$$Z = \sum_v \sum_h e^{-E(v,h)} = q^D \sum_h G(h) \prod_{i=1}^{D} \frac{1+e^{H_i(h)}}{q} \tag{5}$$

where q is the average of $1+e^{H_i(h)}$ $\forall i, h$ that we introduce to avoid overflow. Concerning the training process of RBM, $\mathbf{v}$ is set using real data. Weights $w_{i\mu}$ are initialized sampling values from a Gaussian distribution of mean 0 and standard deviation of $2/\sqrt{L+D}$. Biases b_μ of hidden units are initialized to 0. Thus $\mathbf{h}$ is evaluated by Eq. 1 before passed to Eq. 2 in order to compute back $\mathbf{v}$. In such evaluation it is convenient to set $a_i = \log[p_i/(1-p_i)]$; fixed the i-th visible unit, p_i is defined as the number of times such unit is on over the whole training array set, normalized over the size of the set. Referring to Eq. 2, the choice of shifting the sigmoid $\sigma(x)$ by the function $\log[p_i/(1-p_i)]$ of the average p_i ensures that the units can activate properly, even given initial weights $w_{i\mu}$ close to 0.

The initialization of biases a_i, firstly proposed by Hinton, makes hidden units able to activate and differentiate better their activation based on data, avoiding the risk of getting stuck on values near to 0.

2.2. Computation

We define two Python3 classes, one for the computation of the log-likelihood and the other for defining the RBM structure and managing all the input and output of the training process. Referring to Eq. 4, in order to address the problem of overflow of $e^{H_i(h)v_i}$, we approximate the summation of powers over h as a sum of the biggest ones. The purpose of this choice is to extract the biggest term out of the summation.

Regarding the expression of the partition function shown in Eq. 5, it changes form for Spin case, characterized by units taking values from $\{-1,1\}$. Manipulating the second member of the equation one gets:

$$Z = \sum_v \sum_h e^{-E(v,h)} = \sum_h G(h) \prod_{i=1}^{D} 2\cosh(H_i(h)) \tag{6}$$

For Eq. 6 the $2\cosh(H_i(h))$ term is approximated to just $e^{H_i(h)}$ by shifting all exponents away from negative values by adding a positive value and properly compensating that out of the summation. Thus the same approximation procedure described for Eq. 5 is applied to Eq. 6.

Here describe which tools you decided to use for solving the problem, with equations

$$A = B \tag{7}$$

or systems of equations

$$\begin{aligned} \dot{v}(t) &= -U'(x) \\ \dot{x}(t) &= v(t) \end{aligned} \tag{8}$$

and eventually with pieces of code (Jupyter allows saving

```
import random
import math
import matplotlib.pyplot as plt
```

in latex, it might produce a better output than including a figure with the code as done here).

As already mentioned, the rest of the text is filled with "zzz" to show the typical length of the corresponding sections.

zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz. zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz. zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

quantity	symbol	dimensionless
time	t	t'
momentum	p	v

TABLE I. Description of the table.

FIG. 2. Description...

FIG. 3. Description...

3. RESULTS

Describe what you found.

Cite Figure **??**(a), etc. to add information. Later also cite Figure 2 and Figure 3. Of course the number and size of figures may vary from project to project.

Cite Table I to collect useful data in a clear way.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz. zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz. zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

4. CONCLUSIONS

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.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz. zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz.

zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz zzzz zzzzzzzzz zzzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz z zzzzzzzzzzz 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

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