

GROUP2505 – Log-likelihood analysis for Restricted Boltzmann Machines

Elias Maria Bonasera, Alberto Casellato, Nicola Garbin, and Francesco Pazzocco
(Dated: March 27, 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.

zzzzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzz
z zzzzzzzzzzzz zzzzz zzzzzzzzzz zzzzzzzzzzzzz zzzzz zzzzzzzzzzz zzzzzzz z zzzzzzzzzzzz zzzzz zzzzzzzzzzz
zzzzzzzzzzzzzzzz zzzzz zzzzzzzzzzz zzzzzzz z zzzzzzzzzzzz zzzzz zzzzzzzzzzz
zzzzzzzz z zzzzzzzzzzzz zzzzz zzzzzzzzzzz.

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.

2. RESTRICTED BOLTZMANN MACHINES

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 $\{v_i\}_{i=1,\dots,D}$ is symmetrically connected to a set of L binary hidden units μ of state $\{h_\mu\}_{\mu=1,\dots,L}$; the continuous 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 . One of the most adopted 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 :

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

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

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_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 set to 2 if real data is used to fix $\{v_i\}_{i=1,\dots,D}$ in the first place.

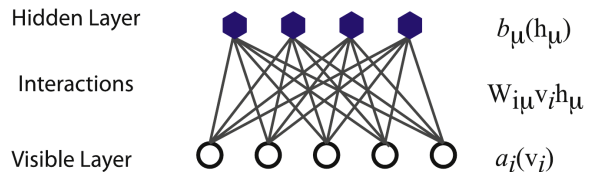


FIG. 1. A Restricted Boltzmann Machine (RBM) is made up of visible units, denoted as v_i , and hidden units, represented as h_μ . These units engage with one another through interactions characterized by the weights $w_{i\mu}$. Notably, there are no direct interactions among the visible units or among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

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

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

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

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

3. METHODS

3.1. 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 all possible state combinations of the system, a possible approach is to split the summation separately over v and h .

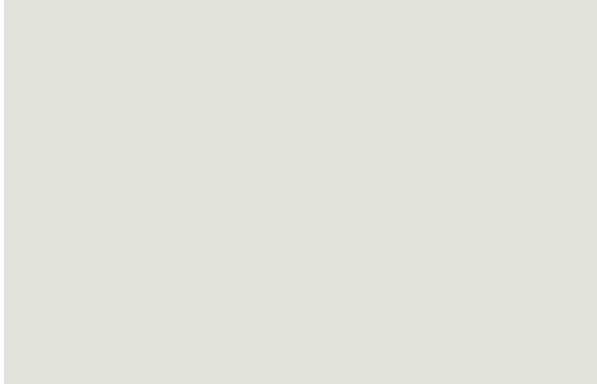


FIG. 2. Description...

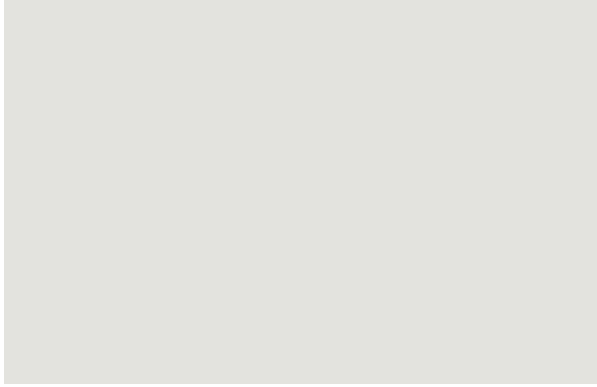


FIG. 3. Description...

Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ. ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z
ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

5. 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.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ. ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z
ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z
ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ
Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ
ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ
ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z
ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

ZZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z ZZZZZZZZZZZZ
ZZZZ ZZZZZZZZZZ ZZZZZZZZZZZZZZZZZ ZZZZZ ZZZZZZZZZZZ ZZZZZZZZ Z
ZZZZZZZZZZZZZ ZZZZ ZZZZZZZZZZ.

APPENDIX

$$\theta_{t+1} = \theta_t - \gamma \cdot \eta_t \cdot \text{sign}(\theta) \quad (1)$$

$$\theta_{t+1} = \theta_t - \eta_t \nabla_{\theta} E(\theta) \quad (2)$$

$$\begin{cases} \mathbf{g}_t = \nabla_{\theta} E(\theta) \\ \mathbf{s}_t = \beta \mathbf{s}_{t-1} + (1 - \beta) \mathbf{g}_t^2 \\ \theta_{t+1} = \theta_t - \eta_t \frac{\mathbf{g}_t}{\sqrt{\mathbf{s}_t + \epsilon}} \end{cases} \quad (3)$$

-
- [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