Statistical mechanics of energy-constrained learning

Master Internship Project - R. Monasson - LPENS

October 2025

Motivation

Recent breakthroughs in Artificial Intelligence require an exponential growth ($\times 4$ per year) of computing power and, therefore, of energy consumption, raising huge ecological and social concerns. Little research has been done si far in the field of machine learning to address this pressing problem. Inspiration can be drawn from the brain of organisms, which have evolved under strong metabolic constraints for hundreds of millions of years. Recently, computational neuroscientists proposed empirical learning rules to curb energy consumption and applied them to few data sets or contexts [1, 2]. The purpose of this internship is to develop statistical mechanics tools to reach a deep understanding of the learning dynamics induced by those rules and, eventually, to improve them.

Energy-efficient learning rules

Standard dynamics for the learning of a neural net parametrized by J is gradient descent of its loss $L(J)$: the update $\Delta J = J(t+1) - J(t)$, where t denotes the step of the learning dynamics, is proportional to $-\partial L/\partial J$. Sparsification of this rule has been proposed to decrease the energy budget of learning with artificial neural networks [1, 2]. In practice, the gradient is multiplied, element-wise ($\odot$), with a probabilistic mask M with 0-1 entries, see Figure 1.

Choices for this mask include:

- fully random with independent entries;
- rank/column deletions to mimic transient removal of neurons as in dropout [3];
- preservation of coherent paths connecting inputs to outputs, called subnets in [1].

Furthermore, non-linear transformation, e.g. clipping of the update can prevent large modifications to the interactions, expected to be costly.

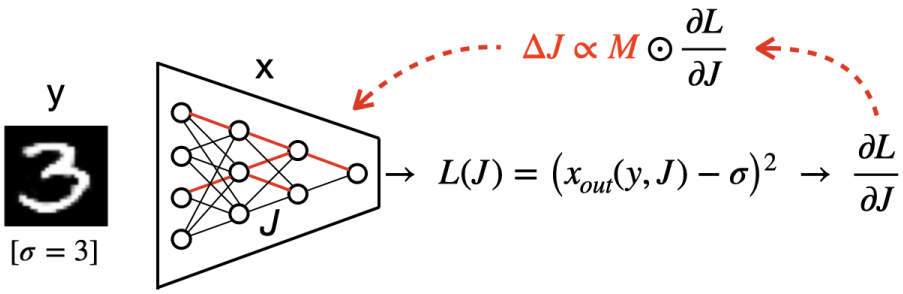

Figure 1: **Energy-efficient training.** A network is trained, here for classifying MNIST digits. The gradient of the loss, after masking many connections (shown in red), defines the updates of the basic network parameters.

Theoretical Analysis

We will characterize the training dynamics induced by these rules. The starting point will be the generating function for the neural and interaction network dynamics, informally written as

$$Z(g_J) = \int \prod_t dJ(t) \prod_t \delta \left(\Delta J(t) + \eta \, M(t) \odot \frac{\partial L}{\partial J} \right) \, e^{\sum_t g_J(t) \cdot J(t)} \qquad (1)$$

where the source term g_J allows one to compute observables of interest; η represents the learning rate.

Exponential representations of the Dirac distributions by means of auxiliary fields will lead to a Martin-Siggia-Rose (MSR) path integral formulation [4], which will be averaged over the random mask; notice that considering other fields x, representing the activities of the units, will be necessary to explicitly write the loss L (Figure 1).

When the measure dJ is Gaussian, this formalism is similar to dynamical mean-field theory, which has been applied to study the dynamics of randomly connected neural networks [5] and of some simple learning problems corresponding to a single-layer neural network [6]. We plan here to consider more complex feedforward networks, as in [1], or even unsupervised architectures, such as Boltzmann machines.

We will compare the proposed rules according to their performance (value of the loss after training, time to convergence) and their energy consumption due to the modifications of the connections and to the activity of the network across the training phase.

References

1. van Rossum MCW, Pache A. Competitive plasticity to reduce the energetic costs of learning. PLoS Computational Biology 20(10): e1012553 (2024)

2. Malkin J, O'Donnell C, Houghton C, Aitchison L. Signatures of Bayesian inference emerge from energy efficient synapses. eLife 12:RP92595 (2023)

3. Srivastava N et al. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15, 1929-58 (2014)

4. Hertz JA, Roudi Y, Sollich P. Path integral methods for the dynamics of stochastic and disordered systems. J. Phys. A: Math. Theor. 50, 033001 (2017)

5. Helias M, Damen D. Statistical field theory for neural networks. Lecture Notes in Physics 970, Springer, Heidelberg (2020)

6. Mignacco F et al. Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. 34th Conference on Neural Information Processing Systems, NeurIPS (2020)