Richiesta assegno di ricerca

Referente: Daniele Tantari
Durata: 18 mesi
Costo: 42000 euro.

Cofinanziamento: 36000 euro
Progetto di riferimento: **PRIN22_TANTARI, "Statistical Mechanics of Learning Machines: from algorithmic and information-theoretical limits to new biologically inspired paradigms", 20229T9EAT – CUP J53D23003640001**

Proposta Commissione:

- Prof. Daniele Tantari (associato), Università di Bologna;
- Prof. Emanuele Mingione (ricercatore rtd-b), Università di Bologna;
- Dott. Enrico Malatesta (ricercatore rtd-a), Università Bocconi;

Supplente: -Prof. Giacomo De Palma (associato), Università di Bologna.

Titolo progetto di ricerca: **INFORMATION BARRIERS IN SELF-SUPERVISED LEARNING**

Descrizione del progetto:

The training of a neural network (NN) is a high-dimensional statistical inference problem requiring large amounts of data. In this sense the first barrier for a NN to work efficiently is precisely the lack of data. In a mathematical formalization, given a certain model for the environment and one for the machine, the main question is how much data, sampled from the environment, are necessary for the machine to efficiently learn a representation of the environment and thus make good predictions. What is crucial is that, regardless of any technological or computational barrier, there exists a fundamental sharp threshold below which it is just impossible for a particular machine to learn anything about the environment. In the Statistical Mechanics (SM) framework, it has been characterized as a phase transition from a regime in which it is possible for the machine to learn and generalize well, to one in which the information present in the dataset is insufficient. The objective of the project is to develop a solid theory characterizing these information-theoretical barriers and how they depend on the model for the environment, the one for the machine and their interplay. An useful approach toward a mathematical formalization of the problem is called teacher-student scenario: a NN with a given architecture (teacher) defines the model for the environment, while another one (student) has to learn something about this environment by leveraging a dataset provided by the teacher. In this controlled setting it is possible to precisely define both the accuracy in reproducing the training set and the generalization error. An information barrier leading to learning reliability phase transitions is the sum of two ingredients. 1)purely data scarcity: even in the best case of a student that uses the same model of the teacher, the amount of examples is not enough for the student to generalize well and infer the correct teacher parameters; 2)model's misspecifications: even with a large training set, if the student doesn't know the teacher's model and uses a different one for learning, it may happen that this model's gap becomes also an information gap. In typical situations both contribute to defining an information limit for learning, in particular their effects do not combine trivially. Aim of this project's objective is to analyze both situations, which in the SM jargon refer to problems on and out the Nishimori line[N01]. Despite most of ML applications are based on supervised approaches, which are task specific and require labeled data, with the growth of the number of tasks and their complexity there is an increasing attention towards unsupervised (recently called self-supervised) approaches, in which the label is the data itself or part of it and the ultimate machine's task is trying to learn the generative model of the dataset. The idea is that a self-supervised training can lead to a machine that has learnt the general rules of the

game, or of the environment in which it is operating, and later is able to specialize more efficiently (in terms of labeled data needed) on many different tasks. A paradigm of self-supervised learning architectures is the Deep Boltzmann Machine (DBM). In terms of generative model (or direct problem) a DBM has been intensively studied [T03,M21,BCMT15] but still there are many open problems concerning how to generalize the Parisi theory to non-convex structures like this, starting from the free energy computation to the characterization of equilibrium states. In the teacher-student scenario the training set of a student DBM is composed of configurations sampled from another teacher DBM. As shown in [BGST17], in this controlled environment, the landscape of the parameter's space can be defined in terms of a Gibbs measure which is the dual of the one defining the generative model. Phase transitions in the generative model thus become transitions between different learning regimes and are therefore related to the presence of information-theoretical barriers. For this reason it is fundamental to study the SM of DBM, starting from the generative model to its relation with the inverse problem of learning. ROLE OF TRAINING SIZE: A disordered-to-order transition can be associated with the switching from the regime in which the student can learn from the one in which the information present in the dataset is insufficient because of its scarcity. The transition is known only for a RBM [BGST17,HVH19], conjectured for a number of hidden units larger than two [DHRT21] and in any case without a rigorous proof. We aim at formalizing these results, extending them to many hidden units and generalizing the approach to the case of a DRBM, by studying its corresponding dual measure and by considering different weights distributions to mimic a dataset with a more general structure. We start from the well specified teacher-student scenario that represents a comfort-zone because the posterior associated with the inference problem is a Gibbs measure on the Nishimori line. We aim to relax this assumption and study the case where data structure and network architecture do not match. ROLE OF WIDTH: We consider RBMs with more hidden units than patterns in the data. Depending on the size of the dataset and network hyperparameters we expect an overfitting transition to a regime where the excess number of hidden units are used to distinguish examples which are actually highly correlated by deriving from the same pattern [AABD22]. We also plan to investigate how this regime changes if hidden units are organized on different layers and the network becomes deep. ROLE OF REGULARIZATION AND ACTIVATION FUNCTION: Another interesting conjecture, corroborated by numerical findings in [KH16], is the existence of a transition in the way patterns are codified: ANNs can learn by prototypes or by features, also known as compositional representation [TM17]. It seems to depend on the activation function and on the weights regularization. We want to prove the existence of such a transition and build a phase diagram especially for DRBMs, where compositional representations become hierarchically structured and other regimes may appear. ROLE OF DEPTH: Most of the previous questions need to be investigated in the case of networks with many layers. We aim at doing that starting from the observation that a DBM is a particular case of a RBM where a portion of visible units are never observed. This motivates the investigation of an intermediate class of self-supervised learning problems with RBMs where visible units are partially and randomly observed, a sort of semi self-supervised learning. REPLICA SYMMETRY BREAKING: Dealing with teacher-student scenarios with misspecifications the problem is not ensured to be replica symmetric [N01] . It is necessary to generalize the Parisi theory of Replica Symmetry Breaking [G03] to the class of non-convex NNs [M21]. This is a difficult mathematical challenge that we propose to tackle with recently developed techniques at the boundary between SM [G03,P13] and pde theory [MP20]. These activities are part of a more structured 2-years project on Statistical Mechanics of Learning Machines, funded by PRIN 2022.

References:

[AABD22] Agliari et al., Neural Networks 148 (2022): 232-253
[BCMT15] Barra et al., Annales Henri Poincaré (Vol. 16, No. 3, pp. 691-708). Springer Basel.

[BGST17] Barra et al. Physical Review E, 96(4), 042156 (2017)
[BGST18] Barra et al. Physical Review E, 97(2), 022310 (2018)
[DHRT21] Decelle et al., Scientific Reports, 11(1), 1-13. (2021)
[G03] Guerra, Communications in mathematical physics, 233(1), 1-12 (2003)
[HK14] Huang et al., Physical Review E 90.5 (2014): 052813.
[HVH19] Hou et al., Journal of Physics A: Mathematical and Theoretical, 52(41), 414001 (2019)
[KH16] Krotov et al., Ad. Neural Inf. Proc. Syst. 29 (2016)
[M21] Mourrat, Probability and Mathematical Physics, 2(2), 281-339 (2021)
[MP20] Mourrat, Panchenko, Electronic Journal of Probability, 25, 1-17. (2020)
[N01] Nishimori, H. (2001). Statistical physics of spin glasses and information processing: an introduction (No. 111). Clarendon Press.
[P13] Panchenko, Springer Science & Business Media, (2013).
[T03] Talagrand, Michel. Spin glasses: a challenge for mathematicians: cavity and mean field models. Vol. 46. Springer Science & Business Media, 2003.
[TM17] Tubiana, Monasson, Phys. Rev. Lett. 118 (2016)