The Variational Playground for Machine Learning in Statistical Physics

Solving Statistical Mechanics using Variational Autoregressive Networks
Authors: Dian Wu, Lei Wang, and Pan Zhang
arXiv:1809.10606 (2018)

Recommended with a Commentary by Giuseppe Carleo, Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

The past decade has seen a prodigious rise of machine learning (ML) applications to several domains, including computer vision, automation, cybersecurity, and much more. The high reputation of ML techniques, and most chiefly deep learning [1], comes from the *fact* that they have successfully solved key problems in those applied areas, substantially improving on existing technology. Given the generality of ML-based approaches, in recent years we are witnessing a massive wave of ML applications to fundamental sciences, including physics. This momentum, however, often makes the task of navigating through new literature rather hard. In this respect, it is often challenging to grasp what the *real potential* of ML applications to physics is, once some of the inevitable hype associated with the rise of this new field is safely kept under control. One criterion I find useful for this purpose, is to focus on whether the proposed ML technology allows to successfully, and convincingly address existing challenging problems, improving on existing computational techniques. In this sense, what makes ML applications to pure science clearly gaugeable, is that it is possible to devise very stringent criteria to establish the success or failure of ML techniques applied to the solution of challenging problems. For example, whereas it can be somehow subjective to establish a baseline performance for image classification, it is much less subjective to establish how well a machine must perform to improve on existing techniques in the realm of computational physics. In this sense, ML applications targeting the direct solution of challenging physical problems are natural candidates to assess the advances in the field, and try to answer to some of the skeptical, "it's all-hype" voices.

Wu, Wang, and Zhang consider the ambitious problem of *solving statistical mechanics*. Take for example a classical spin system, described by the Boltzmann probability density

$$p(s_1 \ldots s_N) \quad = \quad \frac{e^{-\beta E(s_1 \ldots s_N)}}{Z}, \tag{1}$$

where $E(s_1 \ldots s_N)$ is the energy of a given instance of spin configurations $\mathbf{s} \equiv s_1 \ldots s_N$, $\beta = 1/k_b T$ is the inverse temperature, and Z is the partition function. Among the several

challenges posed by complex, interacting classical systems, stand out the problems of computing the expectation value of physical quantities, as well as the free energy $F = -\frac{1}{\beta} \log Z$. While powerful techniques such as Monte Carlo sampling are routinely applied to solve these tasks in important applications, there are however innumerable cases in which efficiently exploring the configuration space is plagued by large entropic barriers, and even the most advanced Monte Carlo techniques fail. These can be for example a major source of concern when studying complex systems with several metastable states, such as proteins or glasses. This is definitely a field where some help from ML would be highly appreciated, and where the performance of new ML-powered techniques can be conveniently measured.

Wu, Wang, and Zhang make an important first step in this direction, introducing an approach that nicely connects the latest advances in ML as a mean to solve challenging problems in statistical physics. The key for this connection is the variational formulation of statistical mechanics. Consider for example a family of variational probability distributions $q_\theta(\mathbf{s})$, parametrized by a (possibly large) set of parameters θ. It is well known that the model free energy

$$F_\theta \;=\; \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \left[E(\mathbf{s}) + \frac{1}{\beta} \log q_\theta(\mathbf{s}) \right], \qquad (2)$$

satisfies the bound $F_\theta \geq F$. Minimizing F_θ with respect to the parameters θ then leads to a reformulation of statistical mechanics in terms of a high-dimensional optimization problem. This variational formulation of statistical mechanics is especially useful when a simplified model q_θ is adopted, for which the variational free energy is easily computed. A separable ansatz $q_\theta(\mathbf{s}) = \Pi_i \theta_i(s_i)$ would result, for example, in the mean field approximation. An outstanding historical problem for this formulation is that, however, providing accurate yet computationally manageable ansatz probability distributions is a hard task.

Particularly innovative in Ref. [2] is the use of a highly expressive family of probability distributions q_θ, built on artificial neural networks consisting of a large number of parameters. Specifically, Wu and coworkers adopt one of the most exciting methodological developments the ML community has put forward in the past years: *autoregressive models*. Auto-regressive networks are a family of probability distributions

$$q_\theta(\mathbf{s}) \;=\; \Pi_i^N q_\theta^i(s_i | s_1 \dots s_{i-1}), \qquad (3)$$

where the full probability is factored into a product of conditional probabilities (see for example [4, 5, 6]) . These conditional probabilities, $q_\theta^i(s_i | s_1 \dots s_{i-1})$, are themselves expressed in terms of artificial neural networks. The key advantages of this ansatz are two. First of all, because of the form (3), it is possible to sample exactly (i.e. without performing Markov Chain Monte Carlo) from q_θ. The sampling procedure boils down to generating the first spin s_1 at random according to the probability $q_\theta^1(s_1)$, then, with the first spin fixed, the second one is generated at random according to $q_\theta^2(s_2 | s_1)$, until the last one is sampled conditioned to all the previous, already drawn, values. Second, the flexibility of the neural networks guarantees that highly non-trivial correlations among the spin variables can be efficiently encoded, going substantially beyond mean field models. A simple example is obtained considering a single-layer feed-forward network, with outputs $\hat{s}_i = \text{sigmoid}(\sum_{j<i} W_{i,j} s_j)$. Here $\text{sigmoid}(z) = 1/(1 + \exp(-z))$ is a non-linear activation function, loosely inspired by neurons physiology.
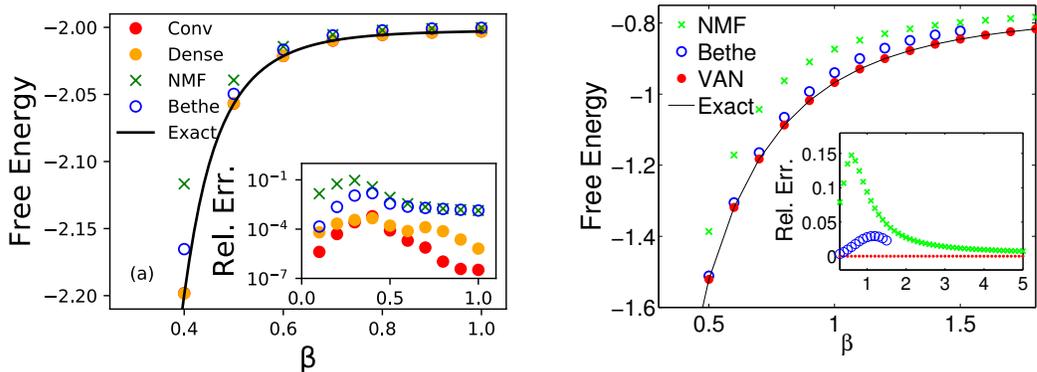
Figure 1: Free energy calculations using the VAN approach as a function of inverse temperature β, from Ref. [2]. (Left) Ising model on a 16x16 square lattice. The relative error on the free energy is plotted in the inset, as compared to mean field (green crosses) the Bethe approximation (blue empty dots). (Right) Disordered Sherrington-Kirkpatrick model for 20 spins, for a fully-connected VAN with only about 100 free parameters.

In this case, the variational parameters are the weights $W_{i,j}$, the condition $j < i$ in the sum guarantees the auto-regressive property, and the probabilities are $q_W^i(s_i|s_1 \ldots s_{i-1}) = \hat{s}_i \delta_{s_i,+1} + (1 - \hat{s}_i)\delta_{s_i,-1}$.

Once the ansatz probability is fixed, the learning algorithm devised by Wu and coworkers is a version of the REINFORCE algorithm used in variational inference (see [3] for an introduction). In brief, at each learning step a (possibly large) batch of samples from the current q_θ is drawn, and the gradient $\nabla_\theta F_\theta$ is approximated as an expectation value of a suitable quantity over those samples. The gradient information is then used to update the parameters in the probability distribution, re-sample new spin configuration, etc until convergence is achieved. The authors overall dub this approach to statistical mechanics as VAN (Variational Autoregressive Networks). This effectively constitutes a self-learning approach, in which no external data is used but the information about the energy of the system. A similar learning rule is used in the context of many-body quantum applications, where the wave-function amplitudes are parametrized with artificial neural-networks [7].

In their paper, Wu and coworkers apply the VAN technology to several spin models, and also considering several neural-network architectures with the autoregressive property. To start with, a sanity check on the 2-dimensional Ising model (Figure 1, left) reveals that neural-network based models are powerful enough to deliver very accurate predictions (a part over ten thousand accuracy or better) on the free energy. This is in line with what expected from the flexibility of artificial neural networks as functional approximations. The power of the autoregressive property, on the other hand, is mostly chiefly exposed when the Boltzmann weight is dominated by multiple modes, possibly separated by large free-energy barriers. Those are typically the most challenging cases for Markov Chain Monte Carlo, and those where an exact sampling can be highly beneficial. Applications to a prototypical spin glass model, (Figure 1, right) as well as to other degenerate ground-states, show that also in this case the VAN approach remains very effective.

3

Overall, this technique provides a very interesting addition to the arsenal of computational techniques already available to solve statistical physics models. The limitation of this approach, that future studies will need to confront with, is that the optimization (learning) process for the neural-network parameters is an intrinsically hard task. In this sense, the original complexity of solving for the free energy of the underlying statistical physics problem is moved to the complexity of solving a high-dimensional optimization problem. In practice, this kind of optimization problem is routinely efficiently solved in modern ML applications, and applications to statistical physics will also try to benefit from the knowledge and technology developed for industry applications.

References

[1] LeCun, Y., Bengio, Y. & Hinton, G. E. *Deep learning.* Nature, 521, 436 (2015).

[2] Wu, D., Wang, L. & Zhang, P. *Solving Statistical Mechanics using Variational Autoregressive Networks.* arXiv:1809.10606 (2018).

[3] Sutton,R. S. & Barto,A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).

[4] Frey, B. J. *Graphical Models for Machine Learning and Digital Communication* (MIT Press, 1998).

[5] Bengio, Y., & Bengio, S. *Modeling high-dimensional discrete data with multi-layer neural networks.* In Advances in Neural Information Processing Systems 12, pages 400–406. (MIT Press, 2000).

[6] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu. *Pixel recurrent neural networks.* In International Conference on Machine Learning, pages. 1747– 1756 (2016).

[7] Carleo, G. & Troyer, M. *Solving the quantum many-body problem with artificial neural networks.* Science 355, 602 (2017).