Machine Learning, Renormalization Group and Phase Transition

Luo Di

May 8, 2017

Abstract

In this essay, we review recent research on the interaction between machine learning, renormalization group and phase transition. There are two important questions to be addressed. The first one is how to understand and improve machine learning algorithms from the perspective of physics while the second one is how to apply machine learning to study physics problems. For the first direction, an attempt to construct a mapping between variational renormalization group and deep learning is presented. For the second direction, an unsupervised learning algorithm has been demonstrated to study the 2D Ising Model phase transition. Discussions on other possibilities between machine learning and physics have also been included.

1 Introduction

In recent years, machine learning has become a fascinating topic in both science and engineering. Practically, it has been applied to many different fields, such as image recognition [1] , business analysis [2], and drug design [4]. Theoretically, the mechanisms behind certain machine learning algorithms, such as neural network, have not been fully understood. It has attracted great interests in understanding and designing machine learning algorithms.

One common feature between physics and machine learning is that they both investigate systems with many degree of freedoms. Physics studies many-body interaction while machine learning handles data in high dimension. In this paper, we will illustrate several recent studies on the interaction between machine learning and physics. There are two big questions of interest. The first one is how to understand and improve machine learning algorithms from the perspective of physics. The second question is how to apply machine learning to solve physics problems. In particular, we focus on the topic of machine learning, renormalization group and phase transition. On one hand, renormalization group, which is a fundamental methodology in physics, provides an attempt to understand machine learning. On the other hand, machine learning provides a new way for studying phase transition problem in physics.

The rest of the paper is organized as follows. Section 2 introduces an attempt to understand machine learning from the perspective of renormalization group. Section 3 illustrates a new way of studying 2D Ising model phase transition using machine learning. Section 2 and Section 3 provides a flavor to address the two big questions mentioned above. Further discussion is given in Section 4 on other possibilities between machine learning and physics, followed by a conclusion.

2 Deep Learning and Renormalization Group

Renormalization group is a natural and fundamental way to handle complex systems with many degrees of freedom in physics. In machine learning community, deep learning algorithms are powerful tools to extract important features from a large amount of data. Conceptually, renormalization group and deep learning share the similar spirit of simplifying the problem, which makes people wonder whether there is certain relation between the two. Recently, Pankaj Mehta and David J. Schwab have proposed an argument to interpret the Restricted Boltzmann machine (RBMs) [7] algorithm through the variational renormalization group method [11]. The material in this section is written based on the paper by Pankaj Mehta and David J. Schwab [12]. We will first introduce variational renormalizatin group and Restricted Boltzmann Machines. After that we will establish the mapping between the two and examine their relation.

2.1 Variational Renormalization Group

Variational renormalization group was introduced as a method to perform renormalization group by Kadanoff [11]. Consider a N binary spin system $\{v_i\}$ on a lattice, where $\{v_i\} = \pm 1$ for each spin. The partition function and the probability distribution for configuration

$\{v_i\}$ are given as follows.

$$Z = \mathrm{Tr}_{\{v_i\}} e^{-H(\{v_i\})} \tag{1}$$

$$P(\{v_i\}) = \frac{e^{-H(\{v_i\})}}{Z} \tag{2}$$

In general, the Hamiltonian $H(\{v_i\})$ has the following form.

$$H(\{v_i\}) = -\sum_i K_i v_i - \sum_{i,j} K_{ij} v_i v_j - \sum_{i,j,k} K_{ijk} v_i v_j v_k + ... \tag{3}$$

Renormalization group is a mapping to transform the original Hamiltonian to a new Hamiltionain with a different set of coupling constants $\{K'_{i_1 i_2 i_3 ...}\}$ and coarse grained variables $\{h_i\}$.

$$H^{RG}(\{h_i\}) = -\sum_i K'_i h_i - \sum_{i,j} K'_{ij} h_i h_j - \sum_{i,j,k} K'_{ijk} h_i h_j h_k + ... \tag{4}$$

It can be shown that after many iterations of RG mappping, certain coupling constants will disappear while some remain [6]. The remaining coupling constants are called relevant operators, which will play important roles. To realize the conceptual idea of RG, we need a concrete RG mapping. Variational renormalization group is one of the schemes to perform RG transformation, which can be implemented numerically. The variational renoramlization group introduces an operator $T_\lambda(\{v_i, h_j\})$ based on a set of parameters $\{\lambda\}$. The Hamiltonian after each RG transformation is constructed as follows.

$$e^{-H^{RG}_\lambda(\{h_i\})} = \mathrm{Tr}_{\{v_i\}} e^{T_\lambda(\{v_i,h_j\}) - H(\{v_i\})} \tag{5}$$

After choosing certain form of $T_\lambda(\{v_i, h_j\})$, we need to minimize the following quantity to zero as close as possible by variation of the parameter set $\{\lambda\}$.

$$log(\mathrm{Tr}_{\{v_i\}} e^{-H(\{v_i\})}) - log(\mathrm{Tr}_{\{h_i\}} e^{-H^{RG}_\lambda}(\{h_i\})) \tag{6}$$

Notice that when the following condition holds, the above quantity becomes zero and the RG transformation is called exact.

$$\mathrm{Tr}_{\{h_i\}} e^{T_\lambda(\{v_i,h_j\})} = 1 \tag{7}$$

Later we will see how to choose the form of $T_\lambda(\{v_i, h_j\})$ so that the variational RG corresponds to the Restricted Boltzamann Machines.

2.2 Restricted Boltzmann Machines

Restricted Boltzmann Machines (RBMs) is one of the popular deep learning algorithms. The simplest RBMs has two layers, one visible layer with input $\{v_i\}$ and one hidden layer with output $\{h_j\}$ as Figure 1 shows.

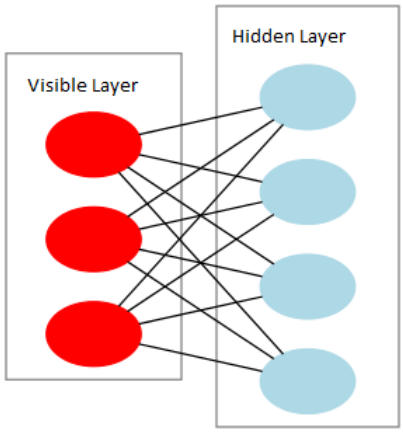

Figure 1: Restricted Boltzmann Machines with a visible layer and one hidden layer

In general, we can keep adding new hidden layers after one hidden layer, which results in a deep learning structure. There is a set of parameter $\lambda = \{b_j, w_{ij}, c_i\}$ stored between two layers in RBMs structure. A energy function for RBMs can be expressed using the above parameters.

$$E(\{v_i\}, \{h_j\}) = \sum_j b_j h_j + \sum_i c_i v_i + \sum_{ij} w_{ij} v_i h_j \tag{8}$$

The goal for RBMs is to return a probability distribution $p_\lambda(\{v_i\})$ as close as possible to the input data probability distribution $P(\{v_i\})$. Notice that a joint probability distribution for RBMs can be defined as follows.

$$p_\lambda(\{v_i\}, \{h_j\}) = \frac{e^{-E(\{v_i\},\{h_j\})}}{Z} \tag{9}$$

We sum over $\{v_i\}$ and $\{h_j\}$ respectively to attain the marginal probability distribution $p_\lambda(v_i)$ and $p_\lambda(h_j)$.

$$p_\lambda(\{v_i\}) = \mathrm{Tr}_{\{h_j\}}\, p_\lambda(\{v_i\}, \{h_j\}) \tag{10}$$

$$p_\lambda(\{h_j\}) = \mathrm{Tr}_{\{v_i\}}\, p_\lambda(\{v_i\}, \{h_j\}) \tag{11}$$

The values for the parameter set $\lambda = \{b_j, w_{ij}, c_i\}$ are attained through the minimization of the Kullback-Leibler divergence.

$$D_{KL}(P(\{v_i\}) \| p_\lambda(\{v_i\})) = \sum_{\{v_i\}} P(\{v_i\}) log(\frac{P(\{v_i\})}{p_\lambda(\{v_i\})}) \tag{12}$$

When $D_{KL}(P(\{v_i\}) \| p_\lambda(\{v_i\})) = 0$, the distribution $p_\lambda(\{v_i\})$ produced by RMBs is the same as the data distribution $P(\{v_i\})$. Before we establish the mapping between variational

RG and RBMs, it is useful to introduce the notation of RBMs variational Hamiltonian for visible layer and hidden layer, $e^{-H_\lambda^{RBM}}$ and $e^{-H_\lambda^{RBM}}$, which satisfy

$$\frac{e^{-H_\lambda^{RBM}(\{v_i\})}}{Z} = p_\lambda(\{v_i\}) = \frac{\mathrm{Tr}_{\{v_i\}}\, e^{-E(\{v_i,h_j\})}}{Z} \tag{13}$$

$$\frac{e^{-H_\lambda^{RBM}(\{h_j\})}}{Z} = p_\lambda(\{h_j\}) = \frac{\mathrm{Tr}_{\{h_i\}}\, e^{-E(\{v_i,h_j\})}}{Z} \tag{14}$$

2.3 A Mapping Between Variational RG and RBMs

The crucial step to establish the mapping between variational RG and RBMs is to choose the following form for the variational operator $T_\lambda(\{v_i, h_j\})$.

$$T_\lambda(\{v_i, h_j\}) = -E(\{v_i\}, \{h_j\}) + H(\{v_i\}) \tag{15}$$

It implies two important results. The first one is that the probability distribution $p_\lambda(\{h_j\})$ in the hidden layer is the same as the Boltzmann distribution given by the renormalized Hamiltonian $H_\lambda^{RG}(\{h_j\})$, which is proved as follows.

$$p_\lambda(\{h_j\}) = \frac{e^{-H_\lambda^{RBM}(\{h_j\})}}{Z} = \frac{\mathrm{Tr}_{\{v_i\}}\, e^{-E(\{v_i,h_j\})}}{Z} \tag{16}$$

$$\frac{e^{-H_\lambda^{RG}(\{h_i\})}}{Z} = \frac{\mathrm{Tr}_{\{v_i\}}\, e^{T_\lambda(\{v_i,h_j\})-H(\{v_i\})}}{Z} = \frac{\mathrm{Tr}_{\{v_i\}}\, e^{-E(\{v_i,h_j\})}}{Z} \tag{17}$$

It further implies that renormalized Hamiltonian $H_\lambda^{RG}(\{h_j\})$ will be the same as the RBMs Hamiltonian in the hidden layer. Therefore, each RG transformation produces a same Hamiltonian as the one that RBMs produces in one hidden layer.

$$H_\lambda^{RG}(\{h_j\}) = H_\lambda^{RBM}(\{h_j\}) \tag{18}$$

The second result is that the exact variational RG condition is the same as $D_{KL} = 0$. It can be proved that

$$e^{T_\lambda(\{v_i,h_j\})} = p_\lambda(\{h_j\}|\{v_i\}) e^{H(\{v_i\})-H_\lambda^{RBM}(\{v_i\})} \tag{19}$$

The exact variation RG conditon in *Equ.*(7) implies that

$$H(\{v_i\}) = H_\lambda^{RBM}(\{v_i\}) \tag{20}$$

$$p(\{v_i\}) = \frac{e^{-H_\lambda(\{v_i\})}}{Z} = \frac{e^{-H_\lambda^{RBM}(\{v_i\})}}{Z} = p_\lambda(\{v_i\}) \tag{21}$$

The minimization condition in each RG step is similar to the minimization in RBMs. Under the exact RG condition, the data distribution $p(\{v_i\})$ can be reproduced by the variational distribution $p_\lambda(\{v_i\})$ so that $D_{KL} = 0$. Similarly, $D_{KL} = 0$ will imply the exact RG condition from *Equ.*(19). Based on the above two results, a mapping between

a variational RG and RBMs is established. To further demonstrate the idea, the authors implement RBMs on the 2D Ising model.

$$H(\{v_i\}) = -J \sum_{ij} v_i v_j \tag{22}$$

The authors first use standard equilibrium Monte Carlo techniques to generate 20,000 samples on a periodic 40×40 2D Ising model at temperature $J = 0.408$, which is below the critical temperature $J_c = 0.4352$. A RBMs with four layers, which are of size 40×40, 20×20, 10×10, 5×5 are constructed to extract information from the Ising Model. It is shown that local spin structure emerges from the training of RBMs, which is a suggestion that RBMs may perform block spin renormalization.

Even though the authors claim that they have constructed an exat mapping between variational RG and deep learning, there are still questions about the relation between renormalization group and machine learning. On one hand, there are many different machine learning algorithms. More than 20 different structures of neural network have been recorded up to now [9] and Restricted Boltzmann Machines is just one of them. On the other hand, an important idea behind RG is to simplify the model instead of data. The essence of RG transformation is to establish a equation between the coupling constants and see how the coupling constants evolve in the RG iteration. In most discussion of machine learning algorithms, even though they are capable of extracting important variables of the problem, such as dimension reduction, seldom discuss how machine learning really simplifies the model and whether machine learning is able to track the flow of coupling constants and detect the universality class. The issues are worth further investigations.

3 Machine Learning for Phase Transition Study

3.1 Approach from Machine Learning

The power of machine learning lies in its ability to detect pattern from a large amount of data. In particular, for unsupervised learning, the algorithm is able to detect the underlying pattern of the data without extra information of the system given. It is natural to ask the question whether it is possible for machine learning to come up with the ordered parameter from data and detect the phase transition in physics problems.

Compared to the traditional theoretical modelling approach, which writes down an equation and solves it, the machine learning approach for phase transition study adopts a totally different philosophy. There are two important steps. The first one is to generate a large amount of data for the system using Monte Carlo or other simulation techniques. The second step is to apply unsupervised learning algorithms to analyze the data and figure out the ordered parameter and other phase transition information. Notice that the machine learning approach can be self-contained theoretically, which doesn't rely on experimental data. The success of the approach depends on both the data quality and the ability of the learning algorithm.

In the following subsection, we present a case study for 2D Ising Model Phase Transition using unsupervised machine learning algorithm from a recent paper [8]. In the paper, it

examines two unsupervised learning algorithms, PCA [10] and autoencoder [7], to various models, including the square and triangular-lattice Ising models, the Blume-Capel model, a highly degenerate biquadratic-exchange spin-one Ising (BSI) model, and the 2D XY model. For simplicity, we only focus on studying 2D Ising Model using PCA.

3.2 Principle Component Analysis

Principle Component Analysis (PCA) is a simple unsupervised learning algorithm. PCA has been applied in many different settings and it is closely related to the singular valued decomposition [10]. The basic idea of PCA is to extract the major information from the data set by projecting the data to certain important basis. More details can be referred to [10]. To apply PCA to the Ising model, we implement the following steps.

The first step is to generate the configuration matrix S, where each row stores a listing of spin configuration from a Monte Carlo simulation. The total number of rows of S, which is also the total number of configurations, is denoted by M. Usually $M = nt$, where t is the number of different temperatures and n is the number of configurations under the same temperature. A 'centered data matrix' X is constructed by subtracting each column of S by the column mean value $m_j = (1/M)\sum_i S_{ij}$.

The second step is to look for the eigenvalues and eigenvectors of the matrix $X^T X$ as follows.

$$X^T X w_n = \lambda_n w_n \tag{23}$$

We only keep the first k largest eigenvalues λ_k and the corresponding eigenvectors w_k. The principle component p_{ij} is calculated as

$$p_{ij} = S_i w_j \tag{24}$$

The 'quantified principle component' is defined as the averages of p_{ij}

$$< |p_i| >= \frac{1}{n}\sum_i |p_{ij}| \tag{25}$$

3.3 Case Study: 2D Ising Model Phase Transition

PCA is applied to study the 2D Ising Model phase transition. The details for data generation and results are presented as follows.

Notice that the critical temperature for 2D Ising model is T_c=2.269. The data is generated using Monte Carlo simulation from $T = 2.0$ to 2.8 under $t = 40$ different temperatures with increment $\Delta T = 0.02$. For each temperature, $n = 10,000$ uncorrelated spin configurations are generated and thus the number of rows for the configuration matrix S is $M = 400,000$. The system size is denoted as L and calculation has been repeated for $L = 20, 30, 40, 50$.

PCA is implemented on the configuration matrix S. The first two largest eigenvalues p_1, p_2 are kept and the corresponding eigenvectors are chosen as principle axis. The results are summarized in the following figures. In Figure 2 [8], the data points above T_c concentrate but the data points below T_c separate, which indicates symmetry breaking for 2D Ising Model.

Figure 3 [8] shows that $|p_1|/L$ drops as T increases from 2.0 to 2.8. In particular, as the system size L increases, it becomes clear that there is a sharp drop around T_c. It indicates that p_1 is actually the magnetization ordered parameter. Figure 4 [8] shows that $|p_2|/L$ first increases and then drops from $T = 2.0$ to 2.8. The peak around T_c becomes clearer as the system size L increases. It implies that p_2 is actually the susceptibility ordered parameter. We can further calculate T_c by plotting the peak location versus $1/L$ as Figure 5 [8]. It is shown that the critical temperature is $T_c = 2.278 \pm 0.015$.

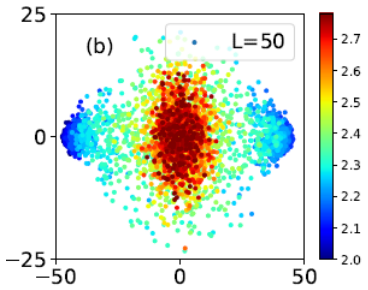

Figure 2: Data projection to the two largest quantified principle components

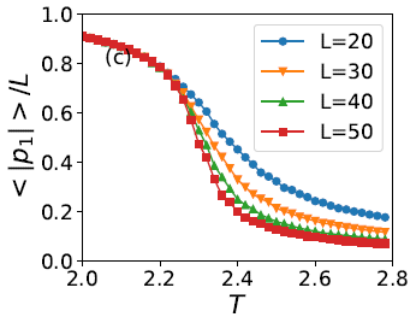

Figure 3: The first quantified principle component vs. temperature

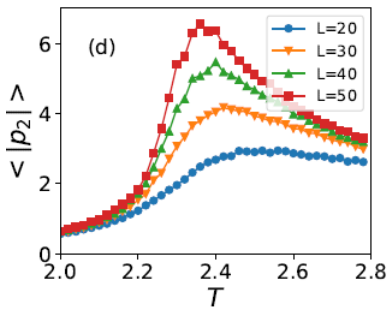

Figure 4: The second quantified principle component vs. temperature

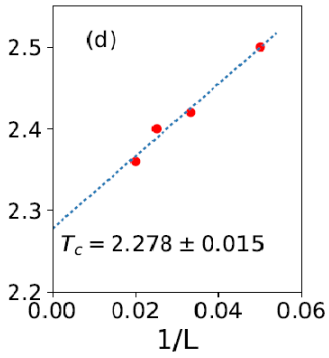

Figure 5: Calculation of critical temperature

This is an example of applying machine learning to phase transition study. Notice that PCA successfully detects the two most important ordered parameters, magnetization and susceptibility. We can further identify the nature of the phase transition and calculate the critical temperature using the information provided by PCA.

4 Further Discussion

In the above two sessions, we mention two examples on the interaction between machine learning and physics, with focus on renormalization group and phase transition study. However, more investigations are going on in this interesting field. To better explore the interplay between physics and machine learning, it is important to integrate ideas from mathematics, physics and information theory. It is found by people in the information geometry community that most models for complex system have the property to be 'sloppy', which means that the model is controlled by a relatively small number of parameter combination [16]. This is similar to the idea from RG that most models only have a few relevant operators. Besides the conventional data analysis by machine learning, people have also developed algorithms to inference natural laws and models based on data. The algorithm by Michael Schmidt and Hod Lipson is able to automatically discover Hamiltonians, Lagrangians and equations of motions using physical system data, such as simple harmonic oscillators and chaotic double-pendula [14]. Ideas from information theory and machine learning can be powerful tools for theoretical physics research as well.

Besides understanding machine learning from physics, another interesting question to ask is whether it is possible to design machine learning algorithms using physics ideas. In the paper by Serena Braddea and William Bialek [3], they propose a new way to analyze neural activity data in the retina using renormalization group idea, where the PCA algorithm doesn't work well. More inspirations come from quantum mechanics. In recent years, the advance of quantum information provides the important information perspective to understand quantum mechanics. Powerful algorithms based on the ideas of information and entanglement are proposed to solve many body problem, such as tensor network. Since machine learning is also dealing with information and data in high dimension, it is natural to think about bridging the two fields. Recently, people have developed several machine learning algorithms using ideas in quantum mechanics, such as supervised learning with tensor network

[13], reinforcement learning using quantum Boltzmann machines [5]. It will be exciting to see new algorithms for machine learning generating from physics ideas in the future.

To close our discussion, we would like to mention an application of machine learning to physics experiment. In the above two sessions, we mainly focus on the interplay between theoretical physics and machine learning. As data analysis is an important part for experimental physics, machine learning will be a powerful tool for experimental physicists in the future. The research of quantum state tomography (QST) is to reconstruct the complete quantum state based on limited measurement. It is a challenge subject and recently people have designed a machine learning approach to QST. The key idea is to represent a quantum state by Restricted Boltzmanns Machines, which can be trained with experimental data and help to reconstruct the complex many-body states [15]. It is believed that the technique will benefit the experimental study of quantum device, such as quantum computer, quantum simulator and quantum microscopes.

5 Conclusion

In this essay, we have reviewed several recent studies on the interaction between physics and machine learning. We mainly try to address two important questions. The first one is how to understand and improve machine learning algorithms from the perspective of physics, while the second one is how to apply machine learning to study physics problems. Two interesting examples have been examined for the above questions respectively. A mapping between variational renormalization group and Restricted Boltzmann Machines has been proposed to investigate the first question. Principle component analysis is used to study 2D Ising Model phase transition, which gives a flavor for the research in the second question. We also include discussion on other possibilities between machine learning and physics, which may point to more discoveries in the future.

References

[1] M. S. Bartlett et al. "Recognizing facial expression: machine learning and application to spontaneous behavior". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. June 2005, 568–573 vol. 2. DOI: 10.1109/CVPR.2005.297.

[2] Indranil Bose and Radha K. Mahapatra. "Business data mining: a machine learning perspective". In: *Information Management* 39.3 (2001), pp. 211–225. ISSN: 0378-7206. DOI: https://doi.org/10.1016/S0378-7206(01)00091-X. URL: http://www.sciencedirect.com/science/article/pii/S037872060100091X.

[3] S. Bradde and W. Bialek. "PCA Meets RG". In: *Journal of Statistical Physics* (Mar. 2017). DOI: 10.1007/s10955-017-1770-6. arXiv: 1610.09733.

[4] R. Burbidge et al. "Drug design by machine learning: support vector machines for pharmaceutical data analysis". In: *Computers Chemistry* 26.1 (2001), pp. 5–14. ISSN: 0097-8485. DOI: https://doi.org/10.1016/S0097-8485(01)00094-8. URL: http://www.sciencedirect.com/science/article/pii/S0097848501000948.

[5] D. Crawford et al. “Reinforcement Learning Using Quantum Boltzmann Machines”. In: *ArXiv e-prints* (Dec. 2016). arXiv: 1612.05695.

[6] N Goldenfeld. “Lectures on phase transitions and the renormalization group”. In: (1992).

[7] G. E. Hinton and R. R. Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks”. In: *Science* 313.5786 (2006), pp. 504–507. ISSN: 0036-8075. DOI: 10.1126/science.1127647. eprint: http://science.sciencemag.org/content/313/5786/504.full.pdf. URL: http://science.sciencemag.org/content/313/5786/504.

[8] W. Hu, R. R. P. Singh, and R. T. Scalettar. “Discovering Phases, Phase Transitions and Crossovers through Unsupervised Machine Learning: A critical examination”. In: *ArXiv e-prints* (Mar. 2017). arXiv: 1704.00080.

[9] Asimov Institute. “The Neural Network Zoo”. In: (2016). URL: http://asimovinstitute.org/neural-network-zoo/.

[10] I. T. Jolliffe. “Principle Component Analysis”. In: (2002). ISSN: 9780387954424. DOI: 10.1007/b98835.

[11] Leo P. Kadanoff, Anthony Houghton, and Mehmet C. Yalabik. “Variational approximations for renormalization group transformations”. In: *Journal of Statistical Physics* 14.2 (1976), pp. 171–203. ISSN: 1572-9613. DOI: 10.1007/BF01011765. URL: http://dx.doi.org/10.1007/BF01011765.

[12] P. Mehta and D. J. Schwab. “An exact mapping between the Variational Renormalization Group and Deep Learning”. In: *ArXiv e-prints* (Oct. 2014). arXiv: 1410.3831.

[13] E. Miles Stoudenmire and D. J. Schwab. “Supervised Learning with Quantum-Inspired Tensor Networks”. In: *ArXiv e-prints* (May 2016). arXiv: 1605.05775.

[14] Michael Schmidt and Hod Lipson. “Distilling Free-Form Natural Laws from Experimental Data”. In: *Science* 324.5923 (2009), pp. 81–85. ISSN: 0036-8075. DOI: 10.1126/science.1165893. eprint: http://science.sciencemag.org/content/324/5923/81.full.pdf. URL: http://science.sciencemag.org/content/324/5923/81.

[15] G. Torlai et al. “Many-body quantum state tomography with neural networks”. In: *ArXiv e-prints* (Mar. 2017). arXiv: 1703.05334.

[16] M. K. Transtrum et al. “Perspective: Sloppiness and emergent theories in physics, biology, and beyond”. In: 143.1, 010901 (July 2015), p. 010901. DOI: 10.1063/1.4923066. arXiv: 1501.07668.