Application of Deep Learning in AdS/CFT

Chen-Yu Liu[1,*] and Wei-Chih Huang[1,†]

[1] *Department of Physics, National Tsing Hua University*

Abstract

In this article, we will introduce deep learning, one of the most popular field in the world in recent years, and some application of it. One of the application is a correspondence between Ads/CFT and Deep Learning process. We would not drive into the theories of AdS/CFT duality directly. Instead, we shall explain each of them separately, i.e., AdS as well as CFT, for the purpose of clarity. After that, we will apply AdS/CFT to deep learning by illustrating an explicit example.

PACS numbers: 03.65.-w 04.62+v 05.50.Gh

* ef850502@gmail.com

† s104021230@nthu.edu.tw

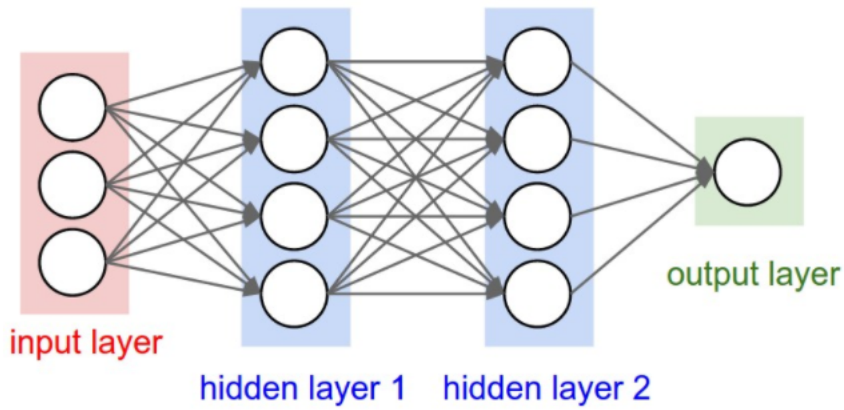


FIG. 1. A three-layer artificial neural network. source: [1]

I. DEEP LEARNING

At the time that Alpha Go beat Lee Sedol, artificial intelligence is known for its superior abilities to compete with a human in some specific field, and the principles behind artificial intelligence are deep learning, so what is deep learning and how can it be so flexible to solve the problem?

Before we come to deep learning, we may first introduce artificial neural network, a mathematical object that constructs of matrices named layers, bias, and weights. In FIG. 1, is a typical three-layer artificial neural network, including input layer (where the input dataset join the learning process), hidden layers (transport weights and biases information durning the process), and output layer (result of the learning process), so the mathematical structure between layer and layer is like:

$$L_i^{[n+1]} = f(\sum_j W_{ij}^{[n]} L_j^{[n]} + b_i^{[n]}) \tag{1.1}$$

$f(x)$ is activation function, $W_{ij}^{[n]}$ is the ij component of the weight of n-th layer, $L_j^{[n]}$ is the j component of n-th layer neuron, $b_i^{[n]}$ is the i component of n-th layer. This compute unit we call it a neuron, a neuron calculates a weighted sum of its input, adds a bias and then decides whether it should be passed or blocked by an activation function. An artificial neural network with various parameters, in theory, can approximate any continuous function [2], this gives it a very strong ability to find patterns or solve a very complex problem with unknown rules but have a lot of data.

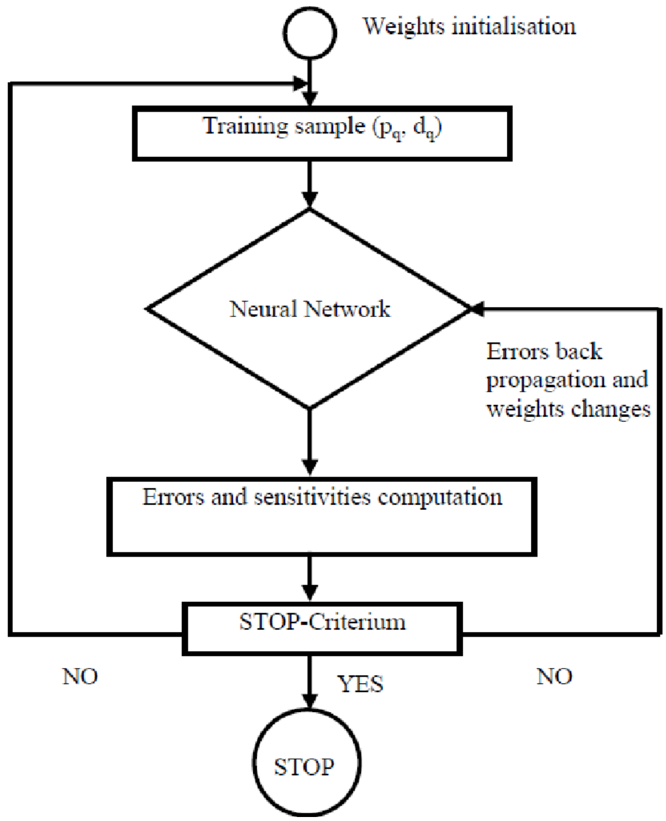


FIG. 2. Diagram of the learning process of the neural network. source: [3]

We can define a loss function in the following form:

$$Loss = \frac{1}{N}\sum_{i=1}^{N} |data^i - prediction^i|^2 \tag{1.2}$$

Which calculate the difference between the data and the prediction of the neural network, N is the number of the data. with the learning process in FIG. 2, the neural network get the proper parameters for the approximation, obviously, the accuracy of the approximation is proportional to the number of the data, the more data you have, the more accurate the parameters of the neural network is.

There are some different types of training the artificial neural network:

A. Supervised learning

The aim of supervised learning is to find a function that match the sample dataset, so the loss function is the difference between the prediction of the neural network and the correct answer provided by the dataset. Usually, this method is used to solve problems like

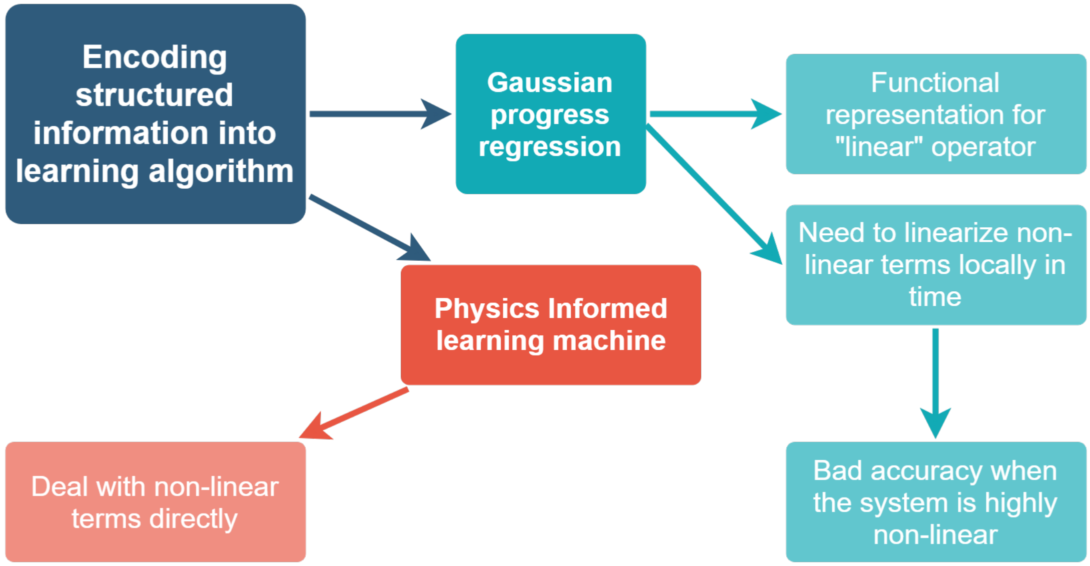


FIG. 3. An example to solve physics problem with deep learning.

classification (pattern finding) and function approximation, for example, hand writing and speech recognition.

B. Unsupervised learning

In general, an estimation problem like statistical distribution or filtering can be deal with unsupervised learning. With some data is given, we can define any function of the data and the neural output as a loss function.

II. APPLICATION OF DEEP LEARNING IN PHYSICS

Deep learning method involved in many research field, including physics, in the course of analyzing complex physical, biological or engineering systems, the cost of data acquisition is prohibitive, and we are inevitably faced with the challenge of drawing conclusions and making decisions under partial information. In this small data regime, the vast majority of state-of-the art machine learning techniques (e.g., deep/convolutional/recurrent neural networks) are lacking robustness and fail to provide any guarantees of convergence [4, 5]. An example is shown at FIG. 3

Consider a parameterized & non-linear PDE:

$$u_t + N[u; \lambda] = 0 \tag{2.1}$$

where N is a non-linear operator parameterized by λ.

Approximate $u(x,t)$ by deep neural network and define PINN (Physics-Informed-Neural-Network):

$$f := u_t + N[u; \lambda] \tag{2.2}$$

And the mean square loss that we are going to minimize:

$$MSE = MSE_u + MSE_f \tag{2.3}$$

where:

$$MSE_u = \frac{1}{N_u} \sum_{i=1}^{N_u} |u(t_u^i, x_u^i) - u^i|^2 \tag{2.4}$$

$$MSE_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2 \tag{2.5}$$

Where eq.(2.4) let the neural network construct an approximation that contains the features in training data, and eq.(2.5), is PINN, that penalizes the solutions that do not satisfy physical conditions. For example, Burger's equation along with Dirichlet boundary conditions:

$$u_t + uu_x - (0.01/\pi)u_x x = 0, x \in [-1, 1], t \in [0, 1], \tag{2.6}$$

$$u(0, x) = -\sin(\pi x), \tag{2.7}$$

$$u(t, -1) = u(t, 1) = 0 \tag{2.8}$$

define $f(t, x)$ to be given by:

$$f := u_t + uu_x - (0.01/\pi)u_x x \tag{2.9}$$

the result is shown at FIG. 4

After data driven solution, we look to data driven discovery. As before, we use Burger's equation as an example:

$$u_t + \lambda_1 uu_x - \lambda_2 u_x x = 0 \tag{2.10}$$

and define PINN:

$$f := u_t + \lambda_1 uu_x - \lambda_2 u_x x \tag{2.11}$$

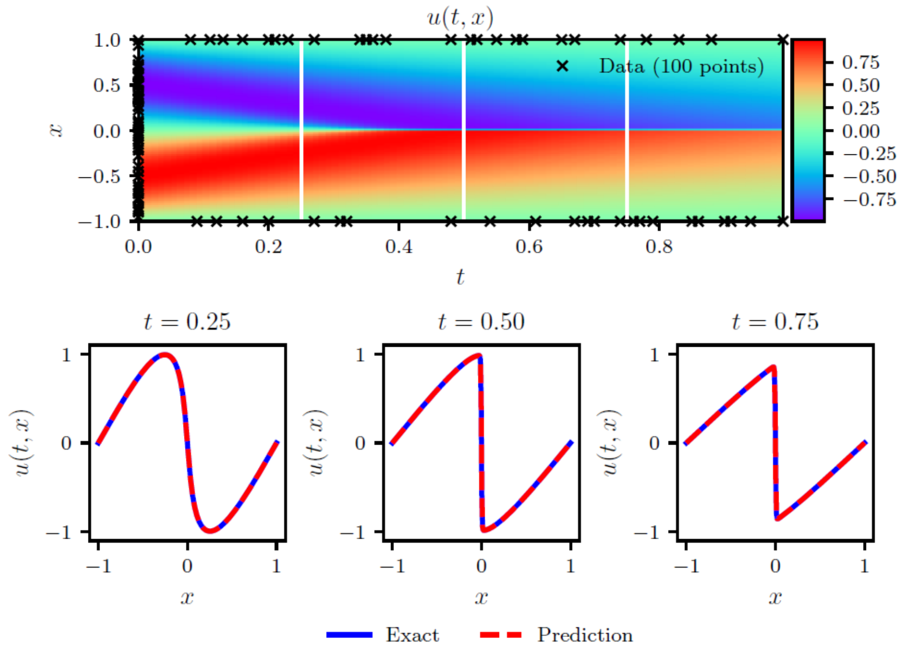


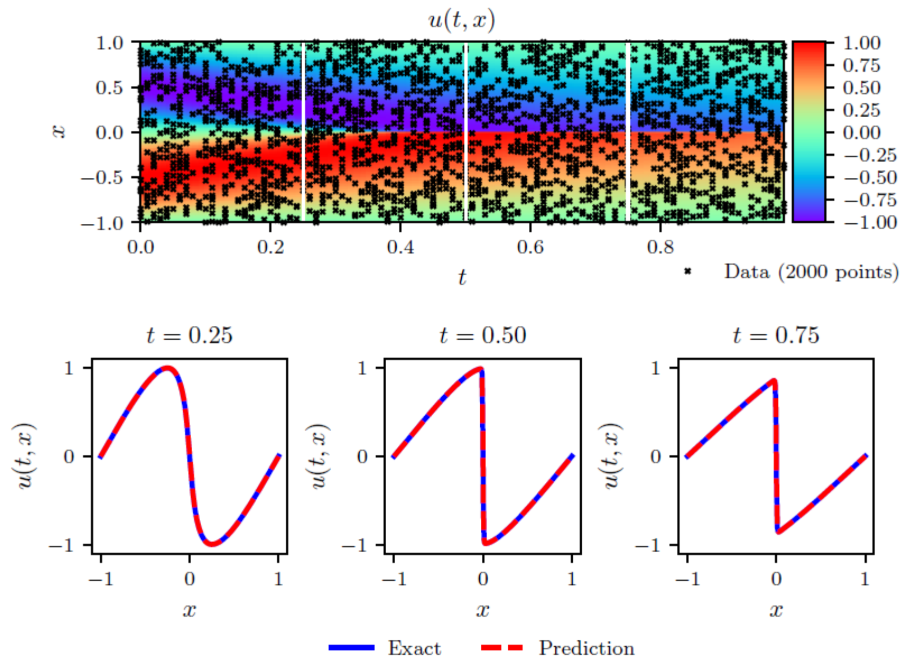
FIG. 4. Burgers' equation: Top: Predicted solution u(t; x) along with the initial and boundary training data. In addition we are using 10,000 collocation points generated using a Latin Hypercube Sampling strategy. Bottom: Comparison of the predicted and exact solutions corresponding to the three temporal snapshots depicted by the white vertical lines in the top panel. The relative L2 error for this case is $6.7*10^{-14}$. Model training took approximately 60 seconds on a single NVIDIA Titan X GPU card.

And the mean square loss that we are going to mininize:

$$MSE = MSE_u + MSE_f \tag{2.12}$$

the result is shown at FIG. 5

Physics informed neural networks, a new class of universal function approximators that is capable of encoding any underlying physical laws that govern a given data-set, and can be described by partial differential equations. The modification result in the good accuracy with relatively small training data set. The way they encoding structured information of the problem is modifying the loss function, in theory, can be any constraint that we are target to. It can be an efficient method to optimize our result in any NN-related problem.



Correct PDE	$u_t + uu_x - 0.0031831u_{xx} = 0$
Identified PDE (clean data)	$u_t + 0.99915uu_x - 0.0031794u_{xx} = 0$
Identified PDE (1% noise)	$u_t + 1.00042uu_x - 0.0032098u_{xx} = 0$

FIG. 5. Burgers equation: Top: Predicted solution $u(t;x)$ along with the training data. Middle: Comparison of the predicted and exact solutions corresponding to the three temporal snapshots depicted by the dashed vertical lines in the top panel. Bottom: Correct partial differential equation along with the identified one obtained by learning λ_1 and λ_2.

III. INTRODUCTION TO ADS/CFT

In this section, we would not drive into the theories of AdS/CFT duality directly. Instead, we shall explain each of them separately, i.e., AdS as well as CFT, for the purpose of clarity.

A. Anti-de Sitter Space(AdS)

AdS, anti-de Sitter space, hypothesizes that the curvature of the universe stays a negative constant. Equivalently, one can formulate Einstein equation in AdS as follow

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{3.1}$$

where $g_{\mu\nu}$ is metric tensor, $R_{\mu\nu}$ is Ricci tensor, R is Ricci scalar, $T_{\mu\nu}$ is energy momentum tensor, G is gravitational constant, and Λ is cosmological constant, which is assumed to be negative in AdS space(here we use the unit $c = 1$ for simplicity). Mathematically speaking, anti-de Sitter space is a maximally symmetric Lorentzian manifold with negative constant curvature. Generically, we have n-dimensional vacuum($T_{\mu\nu}$) solution of metric tensor for (3.1).

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2 d\Omega_{n-2}^2, \tag{3.2}$$

where $f(r) = 1 + \frac{r^2}{\alpha^2}$ for some constant α. It has the following geometric properties

$$R_{\mu\nu\alpha\beta} = \frac{-1}{\alpha^2}(g_{\mu\alpha}g_{\nu\beta} - g_{\mu\beta}g_{\nu\alpha}) \tag{3.3}$$

$$R_{\mu\nu} = \frac{-(n-1)}{\alpha^2}g_{\mu\nu} \tag{3.4}$$

$$R = \frac{-n(n-1)}{\alpha^2} \tag{3.5}$$

B. Conformal Field Theory(CFT)

First we define conformal transformation. Conformal transformation is local change of scale, transforming metric tensor as

$$\tilde{g}_{\mu\nu} = \omega^2(x)g_{\mu\nu}, \tag{3.6}$$

or equivalently

$$d\tilde{s}^2 = \omega^2(x)ds^2, \tag{3.7}$$

where tilde symbol $\sim$ represents the transformed coordinate, which is called conformal frame, and $\omega^2(x)$ is certain nonvanishing function. Note that the angle between two vectors is invariant while the length is not invariant under conformal transformation.(FIG. (6)) In QFT, conformal field is a quantum field that is invariant under conformal transformation.

IV. ADS/CFT DUALITY

AdS/CFT duality claims the equivalence between two theories, strongly-coupled 4D gauge theory and gravity theory in 5D AdS spacetime.(FIG. (7)) Sometimes it is called Maldacena

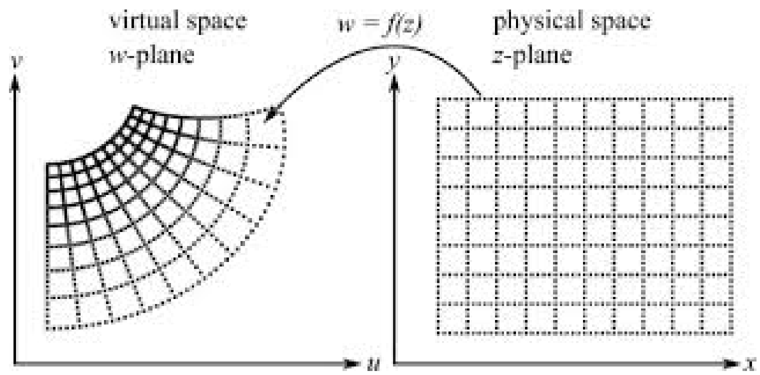


FIG. 6. conformal transformation

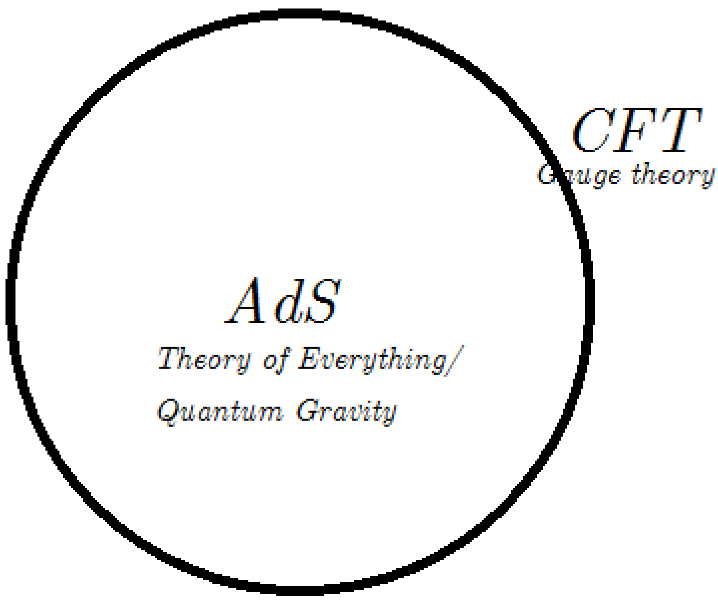


FIG. 7. AdS/CFT duality

duality gauge/gravity duality, or holographic theory. The gauge theory can explain all of the force except gravity, such as strong force, weak force, and electromagnetic force. That is, one has the following gauge group for standard model

$$SU(3) \times SU(2) \times U(1). \tag{4.1}$$

$SU(3)$ describes strong interaction with eight different strong force carriers gluon, which is also known as quantum chromodynamics. $SU(2) \times U(1)$ describes electroweak interaction with 3 different force carriers, which are photons, W bosons, and Z bosons, The last force, gravity, its corresponding force carrier, graviton, is hypothetical because it is generally believed that interaction of single graviton is too weak to be detected in any kind of viable experiment.

In AdS/CFT duality, we can use the equivalence between gauge theory and gravity theory even though they behave greatly different. One of the theory is weak-couple whereas the other is strong-couple. As a result, we shall use weak-coupled gravity theory to understand strong-coupled theory. In this way, the calculation would be highly simplified and easy to carry out. In general, we are particularly interested in black holes. Due to Hawking radiation, black holes have the notion of temperature, so one can construct the thermodynamical system for black holes. Thus it is easy to calculate that the black hole entropy is proportional to its are. In common thermodynamical system, the entropy for the system is usually proportional to volume. Roughly speaking, "area" in five dimensional space is "volume" in four dimensional space. This implies that if four dimensional field theory can ever describe black hole, then black hole must live in five dimensional space. Equivalently, we have the following equation

$$Z_{gague} = Z_{AdS}, \tag{4.2}$$

where Z_{gague} and Z_{AdS} are partition functions of gauge theory and gravity theory, respectively.

In the next section, we will expand the notion of AdS/CFT duality to the field of computer science.

V. ADS/CFT AND DEEP LEARNING

The basic motivation of the relationship between Ads/CFT and deep learning is obvious and simple: given the data from an experiment which is conducted under some external fields, can one model it holographically? In this section, we try to utilize the well-known data processing technology, *deep learning*, to analyze a certain type of model(FIG. (8))

Specifically, we study a data-driven holographic gravity modeling of strongly coupled quantum systems. Although conventional holographic modeling starts with a given bulk gravity metric, we provide a method that solves the inverse problem: given data of a boundary QFT calculate a suitable bulk metric function, assuming the existence of a black hole horizon. In the following, we consider a deep neural network representation of a scalar field

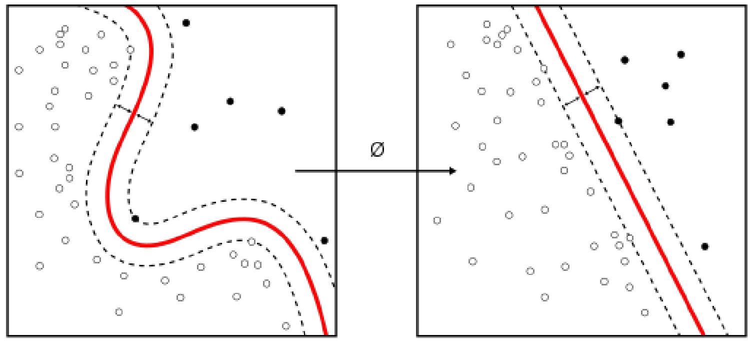


FIG. 8. deep learning

equation in $d+1$ dimensional curved spacetime. The input data is at the boundary of AdS, and the output binomial data is the black hole horizon condition.

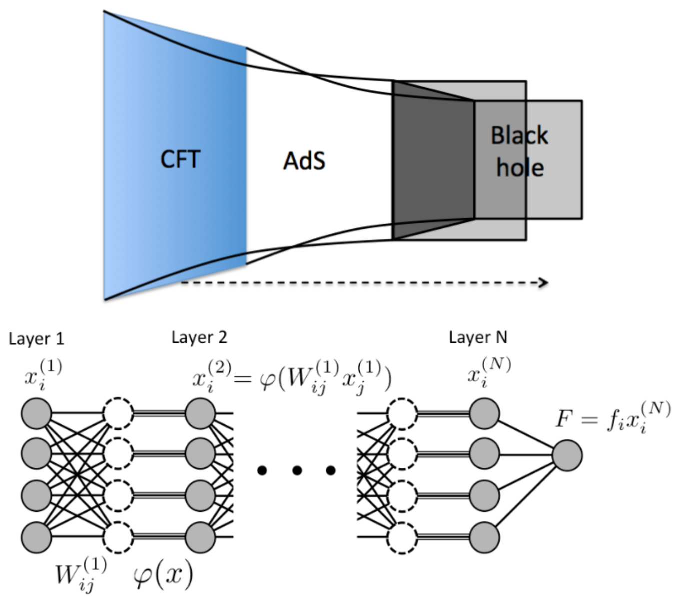


FIG. 9. the AdS/CFT and the deep learning

In [6], we use the data of magnetic response of $Sm_{0.6}Sr_{0.4}MnO_3$ because it has strong quantum fluctuations. We set up the neural network as(FIG. (9))

$$y(x^{(1)}) = f_i \varphi(W_{ij}^{(N-1)} \varphi(W_{jk}^{(N-2)} \dots \varphi(W_{lm}^{(1)} x_m^{(1)}))), \tag{5.1}$$

where x transforms as $x_i \to W_{ij} x_j$(linear), and $x_i \to \varphi(x_i)$(nonlinear). Now suppose a scalar field theory in $d+1$ dimensional curved spacetime is written as

$$S = \int d^{d+1}x \sqrt{-det(g)} \left[-\frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2} m^2 \phi^2 - V(\phi) \right], \tag{5.2}$$

where the metric is given by

$$ds^2 = -f(\eta)dt^2 + d\eta^2 + g(\eta)(dx_1^2 + \cdots + dx_{d-1}^2), \tag{5.3}$$

with the asymptotic AdS boundary condition $f \approx g \approx e^{\frac{2\eta}{L}}$ when $\eta \to \infty$, where L is AdS radius, and black hole horizon boundary condition is $f \approx \eta^2, g \approx constant$ when $\eta \approx 0$. Then one can get the classical equation of motion for $\phi(\eta)$ as follows(FIG. (10)).

$$\partial_\eta \pi + h(\eta)\pi - m^2\phi - \frac{\delta V[\phi]}{\delta \phi} = 0, \tag{5.4}$$

where $\pi = \partial_\eta \phi$. We discretize the above equations.

$$\phi(\eta + \Delta\eta) = \phi(\eta) + \Delta\eta\ \pi(\eta), \tag{5.5}$$

$$\pi(\eta + \Delta\eta) = \pi(\eta) - \Delta\eta \left(h(\eta)\pi(\eta) - m^2\phi(\eta) - \frac{\delta V[\phi]}{\delta \phi} \right). \tag{5.6}$$

The parameter η is discretized as $\eta^{(n)} = (N - n + 1)\delta\eta$. Thus the linear part of neural network is given by

$$W^{(n)} = \begin{bmatrix} 1 & \delta\eta \\ \Delta\eta\ m^2 & 1 - \Delta\eta\ h(\eta^{(n)}) \end{bmatrix}. \tag{5.7}$$

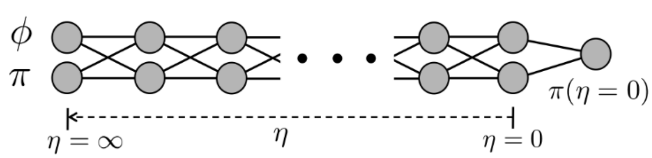


FIG. 10. deep neural network

The activation function is

$$\varphi(x_1) = x_1, \tag{5.8}$$

$$\varphi(x_2) = x_2 + \Delta\eta \frac{\delta V[x_1]}{\delta x_1}. \tag{5.9}$$

We can use the boundary conditions in AdS/CFT as initial conditions of data for iteration with the AdS radius L, asymptotically $h(\eta) \approx \frac{d}{L}$. In QFT with external field value J and its response $\langle \mathcal{O} \rangle$, the linear map becomes

$$\phi(\eta_i) = Je^{-\eta_i \Delta_-} + \langle \mathcal{O} \rangle \frac{e^{-\eta_i \Delta_+}}{\Delta_+ - \Delta_-}, \tag{5.10}$$

$$\pi(\eta_i) = -J\Delta_- e^{-\eta_i \Delta_-} - \langle \mathcal{O} \rangle \frac{\Delta_+ e^{-\eta_i \Delta_+}}{\Delta_+ - \Delta_-}, \tag{5.11}$$

where $\Delta_\pm = \frac{d}{2} \pm \sqrt{\frac{d^2}{4} + m^2 L^2}$ and η_i represents the initial value(input) of η, which $\eta_i \approx \infty$ is the regularized cutoff of the asymptotic AdS spacetime. Then we shall use (5.10) to convert the data of QFT to the input data of neural network. In fact, we input the data at $\eta = \eta_i$ and let it propagate in neural network until $\eta = 0$. If the input data is positive, i.e., $\eta_i > 0$, then the output data should satisfy the boundary condition of the black hole horizon, i.e.,

$$0 = F = \left[\frac{2}{\eta}\pi - m^2 \phi - \frac{\delta V(\phi)}{\delta \phi} \right]_{\eta=\eta_f}, \tag{5.12}$$

where $\eta = \eta_f \approx 0$ is the horizon cutoff(we may assign a small value to it for numerical analysis).

In the following, we apply the method introduced above to a specific example, which is AdS Schwarzschild black hole. We work in $d = 3$, unit $L = 1$, and the metric is

$$h(\eta) = 3 \coth(3\eta), \tag{5.13}$$

where we discretize η by $N = 10$ layers with $\eta_i = 1$ and $\eta_f = 0.1$. For simplicity, we let $m^2 = 1$ and $V[\phi] = \frac{1}{4}\phi^4$. We generate positive(negative) data by collecting randomly generated $\phi(\eta_i, \pi(\eta_i))$ with cutoff $|F| > 0.1(|F| < 0.1)$. Then we obtain 1000 positive and 1000 negative data plot in FIG. (11). Here we use popular Python module, PyTorch, to be

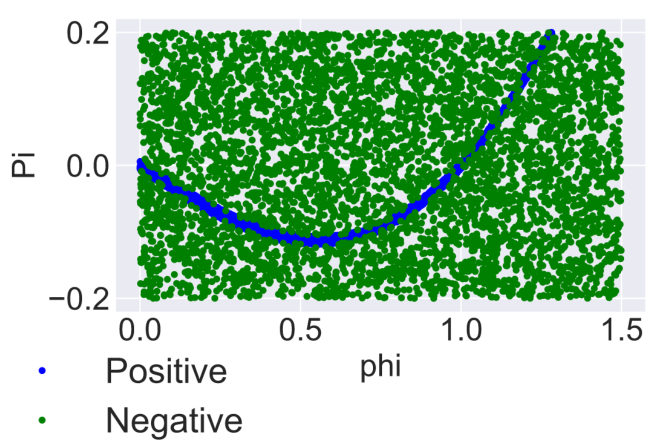


FIG. 11. the generated data

our deep learning library for our work of neural network, and the initial metric is randomly chosen. With 50 learned metric, we see he asymptotic AdS region is almost perfectly learned, as shown in FIG. (12)

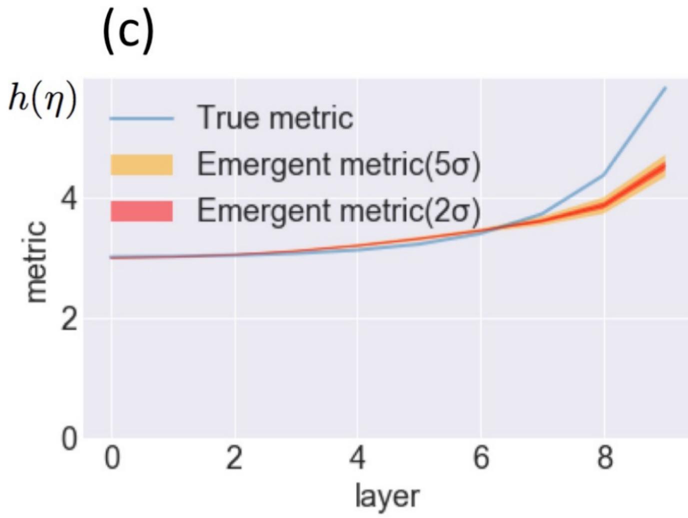


FIG. 12. learned metrics

VI. CONCLUSION

In the first part of this article, we briefly introduce deep learning in computer science by explaining artificial neural network. Furthermore, we illustrate a simple example which is commonly used in teaching. And in the second part of the article, we start the discussion by reviewing AdS/CFT in physics. Next, we practically apply the notion in deep learning to a real world problem, which is a scalar field near black hole. With the technique, we can get self-consistent result in the physics system we consider.

---

[1] `http://cs231n.github.io/convolutional-networks/#conv`

[2] Balzs Csand Csji ,Approximation with Artificial Neural Networks; Faculty of Sciences; Etvs Lornd University, Hungary (2001)

[3] Siddi Tengeleng and Nzeukou Armand, Performance of Using Cascade Forward Back Propagation Neural Networks for Estimating Rain Parameters with Rain Drop Size Distribution; Atmosphere 2014, 5, 454-472 (2014)

[4] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations, arXiv:1711.10561 (2017)

[5] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Physics Informed Deep Learning (Part II): Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations, arXiv:1711.10566 (2017)

[6] K. Hashimoto, S. Sugishita, A. Tanaka and A. Tomiya, "Deep learning and the AdS/CFT correspondence," Phys. Rev. D **98**, no. 4, 046019 (2018) doi:10.1103/PhysRevD.98.046019 [arXiv:1802.08313 [hep-th]].

[7] M. Natsuume, "AdS/CFT Duality User Guide," Lect. Notes Phys. **903**, pp.1 (2015) doi:10.1007/978-4-431-55441-7 [arXiv:1409.3575 [hep-th]].

[8] J. D. Qualls, "Lectures on Conformal Field Theory," arXiv:1511.04074 [hep-th]. 27 citations counted in INSPIRE as of 28 Nov 2018

[9] A. V. Ramallo, "Introduction to the AdS/CFT correspondence," Springer Proc. Phys. **161**, 411 (2015) [arXiv:1310.4319 [hep-th]].

[10] L. Fidkowski, V. Hubeny, M. Kleban and S. Shenker, "The Black hole singularity in AdS / CFT," JHEP **0402**, 014 (2004) doi:10.1088/1126-6708/2004/02/014 [hep-th/0306170].

[11] S. M. Carroll, "Lecture notes on general relativity," gr-qc/9712019.