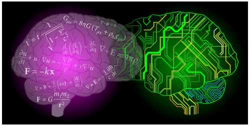
ML and Physics

Lec III: Physics for ML



Ghadir Jafari

University of Cape Town, Cape Town, South Africa

Farhangian University, Tehran, Iran

Physics Help ML

- Concepts and techniques used in ML origins in physics
- **Optimization** techniques: Improving algorithms and proposing different loss functions
- **Modeling and NN architecture**: **energy-based models** inspired by statistical physics
- **Theory of deep learning**: Physics May help understanding the theory of deep learning

Improving GD algorithm

- Drawback of GD:
- Large training inputs
- $\rightarrow$ long time and resource
- Local minimum
- Sensitive to choices of the learning rates, initial conditions,

Stochastic GD with Momentum

- stochastic $\rightarrow$ mini-batch:

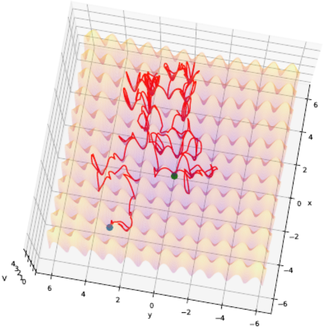
$$\mathcal{C}(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - f\left(x_i\right)\right)^2$$

- Adding Momentum

$$\begin{aligned}\boldsymbol{\omega} &\rightarrow \boldsymbol{\omega}' = \mu\boldsymbol{\omega} - \eta\boldsymbol{\nabla}C \\ \boldsymbol{\theta} &\rightarrow \boldsymbol{\theta}' = \boldsymbol{\theta} + \boldsymbol{\omega}'\end{aligned}$$

Other Cost Functions

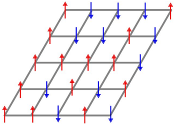
- Energy Conserving cost functions [1]
- Inspired by BI action
- Reaching global minimum



[1] 2201.11137 : G. B. Luca, E. Silverstein

The Ising Model

- Statistical model of Magnet materials



- The partition function

$$Z_\beta = \sum_\sigma e^{-\beta H(\sigma)}$$

- With Hamiltonian

$$H(\sigma) = -\sum_{\langle i\, j \rangle} J_{ij}\sigma_i\sigma_j - \mu \sum_j h_j \sigma_j,$$

- The probability of an configuration

$$P(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_\beta},$$

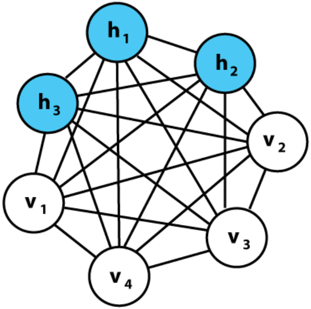
Hopfield Network

- Hopfield model (Network)

$$H(\sigma) = -\sum_{i,j} J_{ij}\sigma_i\sigma_j - \mu \sum_j h_j\sigma_j,$$

- Physics: Temperature β, magnetic field and couplings $\longrightarrow \langle M \rangle$
- ML: The inverse

Boltzmann Machines



Boltzmann Machines

- Boltzmann Machines:

$$\text{Input } v_i \rightarrow P\left(v_i\right) = \exp\left[-\mathcal{E}\left(v_i\right)\right]$$

- Energy Function $\mathcal{E}\left(v_i\right)$:

$$\mathcal{E}\left(v_i\right) \equiv \sum_i a_i v_i + \sum_{i \neq j} w_{ij} v_i v_j$$

- Visible v_i and *hidden* variables h_i
- Optimization $\Rightarrow$ Probabilities

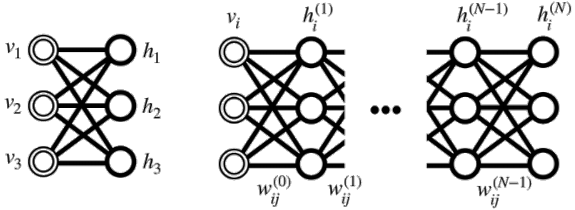
Restricted and Deep Boltzmann Machines

- Restricted Boltzmann Machines

$$\mathcal{E}\left(v_i, h_i\right) \equiv \sum_{i}\left(a_i v_i + b_i h_i\right) + \sum_{ij} w_{ij} v_i h_j$$

- Deep Boltzmann Machines

$$\mathcal{E} \equiv \sum_{i,j} w_{ij}^{(0)} v_i h_j^{(1)} + \sum_{k=1}^{N-1}\left[\sum_{i,j} w_{ij}^{(k)} h_i^{(k)} h_j^{(k+1)}\right]$$



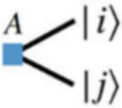
Tensor Networks

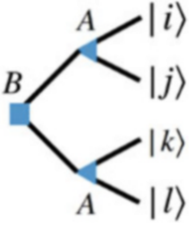
- Four particle state

$$|\Psi\rangle = \psi_{ijkl}|i\rangle|j\rangle|k\rangle|l\rangle$$

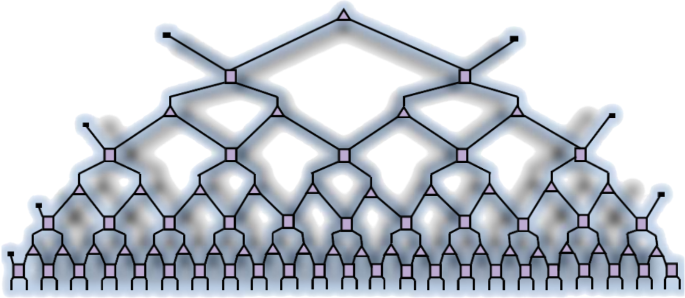
- Represent the state instead by

$$|\Psi\rangle_{TN} = B_{mn}A_{mij}A_{nkl}|i\rangle|j\rangle|k\rangle|l\rangle$$





Tensor Network States



- relationship between the restricted Boltzmann machines and the tensor networks[1]

[1] arXiv:1701.04831, J. Chen, S. Cheng, H. Xie, L. Wang, T. Xiang

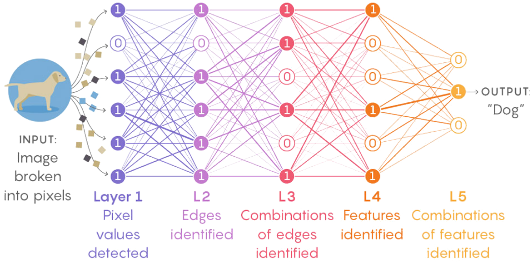
Tensor Networks for Machine Learning

- Both NN and TT could represent quantum states
- Using a tensor network directly as machine learning model architecture
- Supervised Learning with Quantum-Inspired Tensor Networks [1]
- Tensor networks for unsupervised machine learning [2]

[1] arXiv:1605.05775 , E. M. Stoudenmire, D. J. Schwab
[2] arXiv:2106.12974 , J. Liu, S. Li, J. Zhang, P. Zhang

A Theory For deep Learning

- Relation of microscopic aspect and macroscopic techniques?
- Coarse Graining:



The $N \to \infty$ Limit

- Main hyper-parameters Of NN $\Rightarrow N_\ell$ And L
- The output distribution converge to multivariate Gaussian distributions for finite L infinite width limit $N_\ell \to \infty$ [1]
- No Learning! ☹
- The NN is no longer deep $\frac{L}{N} \to 0$
- Large depth L and large width N , with their ratio L/N held fixed
- finite-width-corrected nearly-Gaussian model

[1] arXiv:1711.00165 , J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, J. Sohl-Dickstein

Effective Field Theory Approach To NNs[1,2]

NGP / finite NN	Interacting QFT
input x	external space or momentum space point
network output $f_L(x)$	interacting field ϕ_Λ
non-Gaussianities	interactions
non-Gaussian coefficients	coupling strengths
log probability	effective action S_Λ

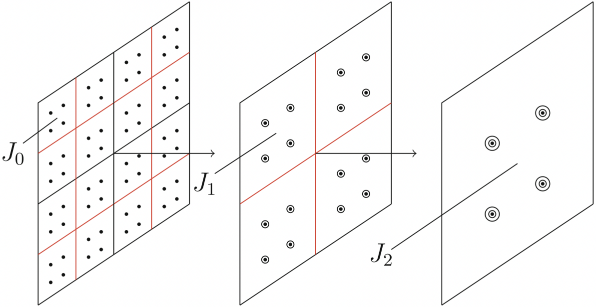
- Perturbative methods in QFT (Expansion in $\frac{L}{N}$)
- Interactions as perturbations around free asymptotic states

[1] arXiv:2106.10165 , D. A. Roberts, S. Yaida, B. Hanin, CUP

[2] arXiv:2008.08601 , J. Halverson, A. Maiti, K. Stoner

RG and DNN

- The RG



- Real space renormalization

RG to NN

- RG in Ising

$$\mathbf{H}\left[\{v_i\}\right] = -\sum_i K_i v_i - \sum_{ij} K_{ij} v_i v_j$$

- The Effective (coarse-grained) Hamiltonian

$$\mathbf{H}^{RG}\left[\left\{h_j\right\}\right] = -\sum_i \tilde{K}_i h_i - \sum_{ij} \tilde{K}_{ij} h_i h_j$$

- Variational RG:

$$e^{-\mathbf{H}^{RG}_{\lambda}\left[\left\{h_j\right\}\right]} \equiv \mathrm{Tr}_{v_i}\, e^{\mathbf{T}_{\lambda}\left(\{v_i\},\left\{h_j\right\}\right) - \mathbf{H}(\{v_i\})}$$

- Interactions between the physical and coarse-grained degrees of freedom

RG and RBM

- Choosing (**Optimizing**) the parameters λ such that the partition functions be same

$$\mathcal{Z} = \mathrm{Tr}_{v_i}\, e^{-\mathbf{H}[\{v_i\}]} = \mathrm{Tr}_{h_i}\, e^{-\mathbf{H}^{RG}(\{h_i\})}$$

- The same procedure of training RBM with [1]

$$\mathbf{E}\left(\{v_i\}, \{h_j\}\right) = \sum_i b_j h_j + \sum_{ij} v_i w_{ij} h_j + \sum_i c_i v_i$$
$$\lambda = \left\{b_j, w_{ij}, c_i\right\}$$
$$\mathbf{T}\left(\{v_i\}, \{h_j\}\right) = -\mathbf{E}\left(\{v_i\}, \{h_j\}\right) + \mathbf{H}\left[\{v_i\}\right]$$

[1] arXiv:1410.3831, , P. Mehta, D. J. Schwab

PCA meets RG [1]

- Diagonalizing the covariance matrix
- Eliminate the modes that have small variance.
- Restricting to modes with less than some cutoff
- Project onto just a few degrees of freedom
- Momentum space renormalization

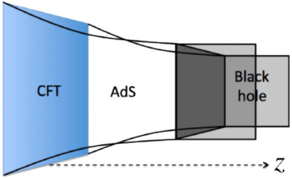
[1] arXiv:1610.09733 S. Bradde, W. Bialek

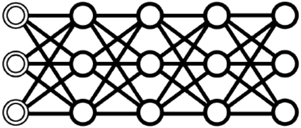
TNs, NNs, QFT, Entanglement and Geometry

- Tensor Network representation of states
- Tensor Networks and Renormalization: MERA
- Geometry of the network $\Rightarrow$ AdS/CFT correspondence
- QMBS $\Rightarrow$ QFT
- Relation of NN with entanglement and geometry?

AdS/CFT and Deep Learning[1]

- Energy scale in QFT $\sim$ Depth in NNs
- Energy scale in QFT $\sim$ Depth in AdS space
- AdS/CFT to a DBM:





[1] arXiv:1903.04951, K. Hashimoto

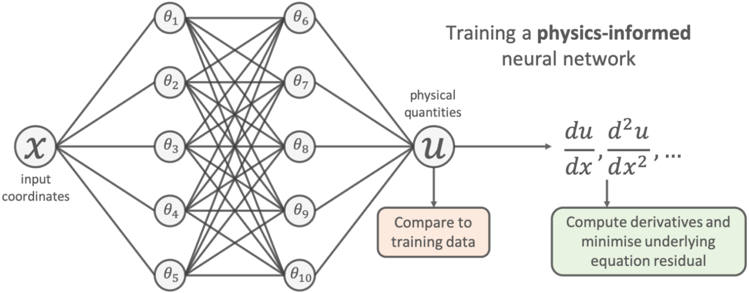
AdS/CFT and Deep Learning

AdS/CFT	Deep Boltzmann machine
Bulk coordinate z	Hidden layer label k
QFT source $J(x)$	Input value v_i
Bulk field $\phi(x,z)$	Hidden variables $h_i^{(k)}$
QFT generating function $Z[J]$	Probability distribution $P(v_i)$
Bulk action $S[\phi]$	Energy function $\mathcal{E}\left(v_i, h_i^{(k)}\right)$

$$Z_{\mathrm{QFT}}[J] = \exp\left(-S_{\mathrm{gravity}}[\phi]\right) = \sum_{h_i} \exp(-\mathcal{E}(v_i, h_i))$$

Physics-Informed Machine Learning[1]

- Include prior scientific knowledge into our machine learning



[1] nature.com/articles/s42254-021-00314-5

Physics-Informed Machine Learning

- Add the known differential equations directly into the loss function when training the neural network

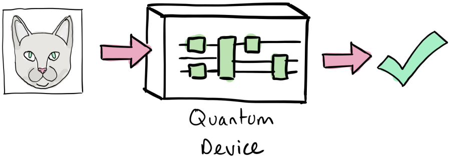
$$\mathcal{C}(\theta) = \min \frac{1}{N} \sum_{i}^{N} \left(u_{\mathrm{NN}}\left(x_i; \theta\right) - u_{\text{true}}\left(x_i\right)\right)^2 + \frac{1}{M} \sum_{j}^{M} \left(\left[m \frac{d^2}{dx^2} + \mu \frac{d}{dx} + k\right] u_{\mathrm{NN}}\left(x_j; \theta\right)\right)^2$$

- Using NN for solving differential equation[1]

[1] doi.org/10.1016/j.jcp.2018.10.045

Quantum Machine Learning

- Quantum computers can help improve ML algorithms
- Quantum networks replace classic neural networks
- Using quantum devices like neural networks
- adapt the **physical control parameters**, such as an electromagnetic field strength or a laser pulse frequency, to solve a problem



Summery

- Physics can help developments of ML
- Making better optimization algorithms
- Model selection and NN architecture
- Making process in understanding the theory beyond Deep learning