Swansea University
Prifysgol Abertawe

Quantum field-theoretic machine learning

Dimitrios Bachtis

Joint work with Profs. Gert Aarts and Biagio Lucini.

LATTICE 21
JULY 26-30 2021, ZOOM/GATHER@MIT
The 38th International Symposium on Lattice Field Theory

Can we view machine learning as part of quantum field theory?

And why?

Probability distribution

A probability distribution is a product of strictly positive and appropriately normalized factors (or potential functions) ψ:

$$p(\phi) = \frac{\prod_{c \in C} \psi_c(\phi)}{\int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}},$$

Probability distribution

A probability distribution is a product of strictly positive and appropriately normalized factors (or potential functions) ψ:

$$p(\phi) = \frac{\prod_{c \in C} \psi_c(\phi)}{\int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}},$$

1. Factors are the fundamental building blocks of probability distributions.

Probability distribution

A probability distribution is a product of strictly positive and appropriately normalized factors (or potential functions) ψ:

$$p(\phi) = \frac{\prod_{c \in C} \psi_c(\phi)}{\int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}},$$

1. Factors are the fundamental building blocks of probability distributions.

2. By controlling the factors we are able to control the probability distribution.

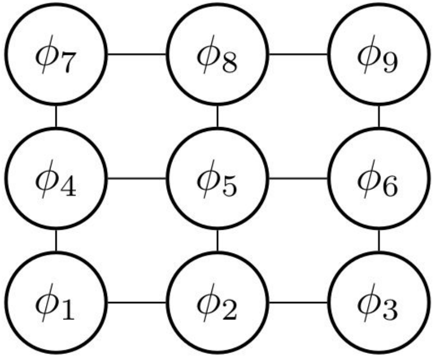
Representation

We require some form of representation to construct the probability distribution. We are going to use a finite set Λ that we express as a graph $\mathcal{G}(\Lambda, e)$ where e is the set of edges in $\mathcal{G}$.

A clique c is a subset of Λ where the points are pairwise connected. A maximal clique is a clique where we cannot add another point that is pairwise connected with ***all*** the points in the subset.

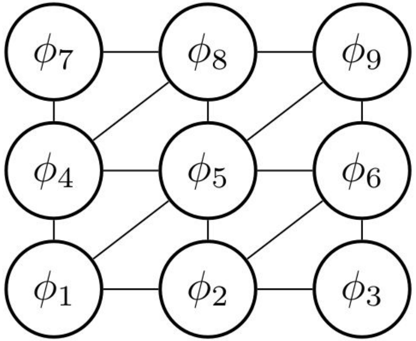
Representation

On the square lattice a maximal clique is an edge.



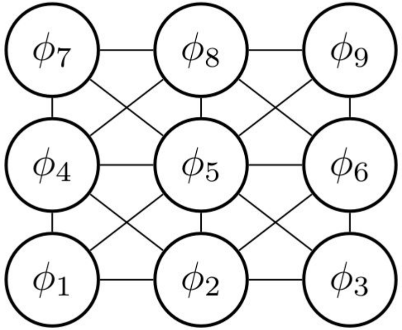
(a)

On a triangular lattice a maximal clique is a triangle.

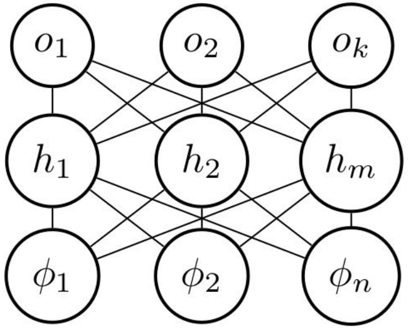


(b)

On the square lattice with both diagonals a maximal clique is a square.



(c)



(d)

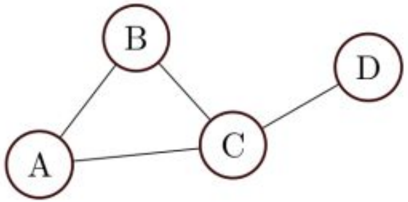
On the bipartite graph, which represents standard neural network architectures a maximal clique is an edge.

Representation

Given a graph $\mathcal{G}(\Lambda, e)$, the random variables φ_i at each point i define a Markov random field if they fulfill the local Markov property with respect to $\mathcal{G}$.

The local Markov property denotes that a random variable φ_i depends only on its neighbors and it is conditionally independent of all other random variables in the set:

$$p(\phi_i | (\phi_j)_{j \in \Lambda - \phi_i}) = p(\phi_i | (\phi_j)_{j \in n_i}).$$



Representation

Hammersley-Clifford theorem

A strictly positive distribution p satisfies the local Markov property of an undirected graph $\mathcal{G}$, if and only if p can be represented as a product of strictly positive potential functions ψ_c over $\mathcal{G}$, one per maximal clique c, i.e.

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi), \quad Z = \int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}$$

where Z is the partition function and $\boldsymbol{\varphi}$ are all possible states of the system.

Representation

There are two different directions to pursue:

1. We can devise potential functions that satisfy the Hammersley-Clifford theorem to construct a Markov random field.

Representation

There are two different directions to pursue:

1. We can devise potential functions that satisfy the Hammersley-Clifford theorem to construct a Markov random field.
2. We can evaluate if known physical systems can be recast within this mathematical framework by verifying instead if they satisfy the theorem.

We will pursue the second direction.

Representation

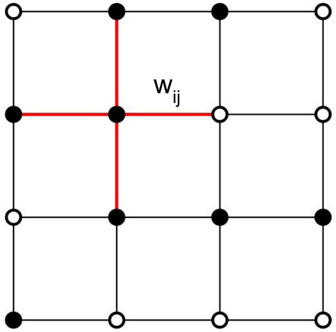
2d φ^4 theory:

$$\mathcal{L}_E = \frac{\kappa}{2}(\nabla\phi)^2 + \frac{\mu_0^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4,$$

$$S_E = -\kappa_L \sum_{\langle ij \rangle} \phi_i \phi_j + \frac{(\mu_L^2 + 4\kappa_L)}{2} \sum_i \phi_i^2 + \frac{\lambda_L}{4} \sum_i \phi_i^4.$$

k_L, μ_L, λ_L dimensionless parameters

$$w = \kappa_L, \; a = (\mu_L^2 + 4\kappa_L)/2, \; b = \lambda_L/4$$



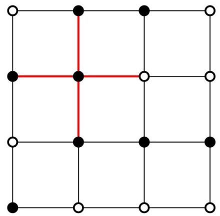
Inhomogeneous φ^4 theory:

$$S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i \phi_i^2 + \sum_i b_i \phi_i^4,$$

Representation

The φ^4 lattice field theory is, by definition, formulated on a square lattice which is equivalent to a graph $\mathcal{G}(\Lambda,e)$. A non-unique choice of potential function per each maximal clique is:

$$\psi_c = \exp\left[-w_{ij}\phi_i\phi_j + \frac{1}{4}(a_i\phi_i^2 + a_j\phi_j^2 + b_i\phi_i^4 + b_j\phi_j^4)\right],$$



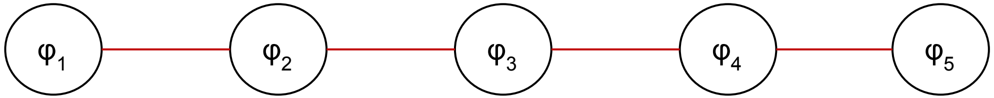
The probability distribution is expressed as a product of strictly positive potential functions ψ, over each maximal clique:

$$p(\phi;\theta) = \frac{\exp\left[\sum_{c\in C}\ln\psi_c(\phi)\right]}{\int_{\boldsymbol{\phi}}\exp\left[\sum_{c\in C}\ln\psi_c(\boldsymbol{\phi})\right]d\boldsymbol{\phi}} = \frac{1}{Z}\prod_{c\in C}\psi_c(\phi).$$

The φ^4 theory satisfies Markov properties and it is therefore a Markov random field.

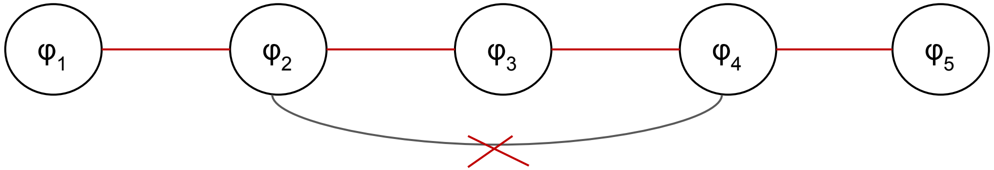
Representation

The Markov property in a Markov chain



Representation

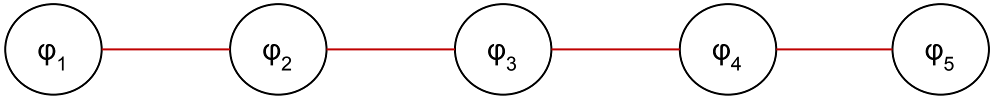
The Markov property in a Markov chain



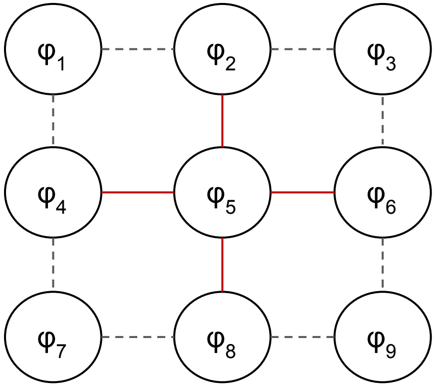
Not allowed!!

Representation

The Markov property in a Markov chain



A Markov random field satisfies the Markov property in high-dimensions



Learning

Having established that certain physical systems are Markov random fields, how do we use them for machine learning?

Learning

Having established that certain physical systems are Markov random fields, how do we use them for machine learning?

Exactly in the same way as any other machine learning algorithm...

Learning

The φ^4 theory has a probability distribution $p(\varphi;\theta)$ with action $S(\varphi;\theta)$:

$$p(\phi;\theta) = \frac{\exp\big[-S(\phi;\theta)\big]}{\int_{\boldsymbol{\phi}} \exp[-S(\boldsymbol{\phi},\theta)]d\boldsymbol{\phi}}.$$

We now consider a quantum field theory with action A and a target probability distribution $q(\varphi)$:

$$q(\phi) = \exp[-\mathcal{A}]/Z_{\mathcal{A}}$$

Learning

We can then define an asymmetric distance between the probability distributions p(φ;θ) and q(φ), which is called the Kullback-Leibler divergence:

$$KL(p||q) = \int_{-\infty}^{\infty} p(\phi;\theta) \ln \frac{p(\phi;\theta)}{q(\phi)} d\phi \geq 0.$$

Learning

We can then define an asymmetric distance between the probability distributions p(φ;θ) and q(φ), which is called the Kullback-Leibler divergence:

$$KL(p||q) = \int_{-\infty}^{\infty} p(\phi;\theta) \ln \frac{p(\phi;\theta)}{q(\phi)} d\phi \geq 0.$$

We want to minimize the Kullback-Leibler divergence.

By minimizing it we would make the two probability distributions equal. We can then use the probability distribution p(φ;θ) of action S to draw samples from the target distribution q(φ) of action A.

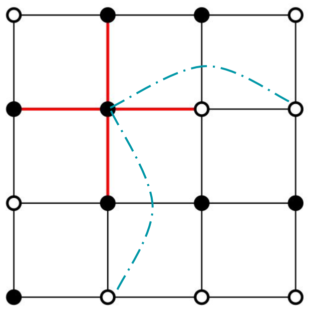
Learning

A proof-of-principle demonstration is to use the inhomogeneous action S:

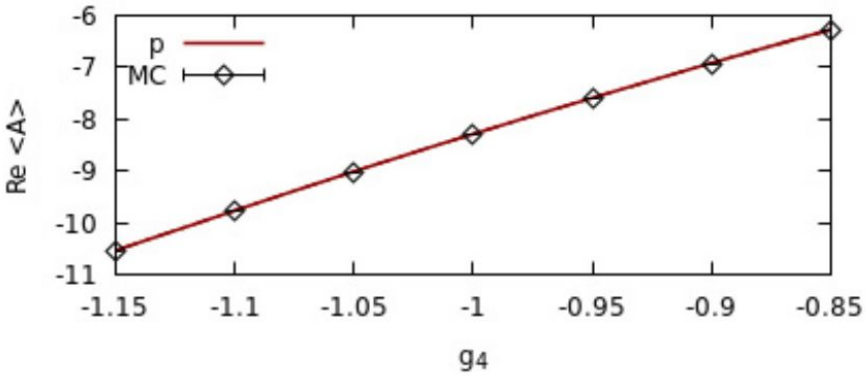
$$S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4,$$

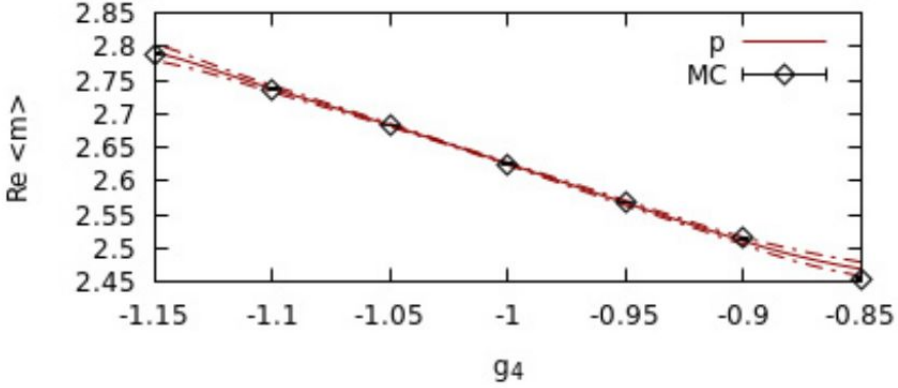
to learn an action that includes longer-range interactions:

$$\mathcal{A}_{\{4\}}(\phi) = -\sum_{\langle ij \rangle} \phi_i\phi_j + 1.52425\sum_i \phi_i^2 + 0.175\sum_i \phi_i^4 - \sum_{\langle ij \rangle_{nnn}} \phi_i\phi_j$$



Learning





The results include reweighting to a complex-valued coupling constant on the mass term and extrapolations in parameter space along the trajectory of the coupling constant g_4 in the longer-range interaction.

$$\mathcal{A}_{\{5\}}(\phi) = -\sum_{\langle ij \rangle} \phi_i \phi_j + 1.52425 \sum_i \phi_i^2 + 0.175 \sum_i \phi_i^4 - \sum_{\langle ij \rangle_{nnn}} g' \phi_i \phi_j + i0.15 \sum_i \phi_i^2$$

Learning

What if the target probability distribution q(φ) is unknown?

Learning

We are searching for the optimal values of the coupling constants in the φ^4 action that are able to reproduce the data as configurations in the equilibrium distribution.

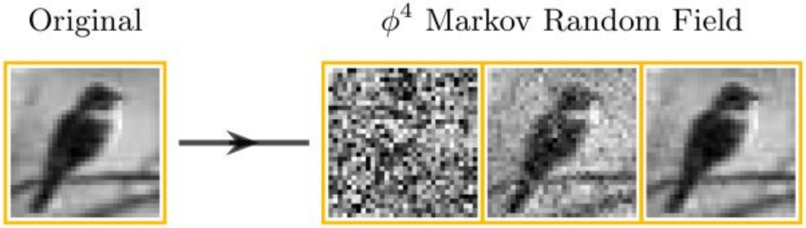
$$S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4,$$

Learning

We are searching for the optimal values of the coupling constants in the φ^4 action that are able to reproduce the data as configurations in the equilibrium distribution.

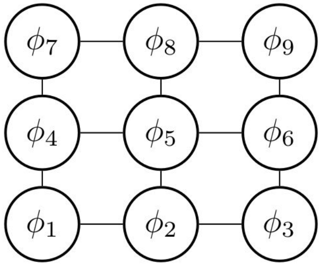
$$S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4,$$

Case of an image:



Neural Networks

φ⁴ Markov random field

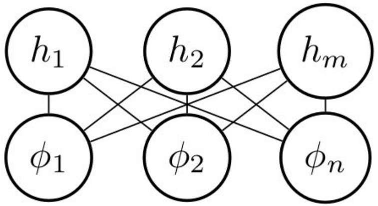


$$S(\phi;\theta) = -\sum_{\langle ij\rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4,$$

$$\theta = \{w_{ij}, a_i, b_i\}$$

$$p(\phi;\theta) = \frac{\exp\left[-S(\phi;\theta)\right]}{\int_{\boldsymbol{\phi}} \exp[-S(\boldsymbol{\phi},\theta)]d\boldsymbol{\phi}}.$$

φ⁴ neural network



$$S(\phi,h;\theta) = -\sum_{i,j} w_{ij}\phi_i h_j + \sum_i r_i\phi_i + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4 + \sum_j s_j h_j + \sum_j m_j h_j^2 + \sum_j n_j h_j^4,$$

$$\theta = \{w_{ij}, r_i, a_i, b_i, s_j, m_j, n_j\}$$

$$p(\phi,h;\theta) = \frac{\exp[-S(\phi,h;\theta)]}{\int_{\boldsymbol{\phi},\boldsymbol{h}} \exp[-S(\boldsymbol{\phi},\boldsymbol{h};\theta)]d\boldsymbol{\phi}d\boldsymbol{h}}.$$

Neural Networks

The φ^4 neural network:

$$S(\phi, h; \theta) = -\sum_{i,j} w_{ij}\phi_i h_j + \sum_i r_i \phi_i + \sum_i a_i \phi_i^2 + \sum_i b_i \phi_i^4 + \sum_j s_j h_j + \sum_j m_j h_j^2 + \sum_j n_j h_j^4,$$

is a generalization of other neural network architectures:

Gaussian-Gaussian restricted Boltzmann machine:

b_i=n_j=0

Gaussian-Bernoulli restricted Boltzmann machine:

b_i=n_j=m_j=0
h_j binary

Bernoulli-Bernoulli restricted Boltzmann machine:

b_i=n_j=m_j=a_i=0
$\varphi_{I,}$ h_j binary

φ^4-Bernoulli restricted Boltzmann machine:

m_j=n_j=0
h_j binary

φ^4 equivalence with the Ising model (under an appropriate limit)

Conclusions

1. What one needs to do machine learning is simply a probability distribution. Lattice field theories therefore emerge as natural machine learning algorithms. We can investigate machine learning as a physical concept within quantum field theory: e.g. what are the phase transitions of quantum field-theoretic machine learning algorithms? How do they behave when they interact with external fields?

Conclusions

1. What one needs to do machine learning is simply a probability distribution. Lattice field theories therefore emerge as natural machine learning algorithms. We can investigate machine learning as a physical concept within quantum field theory: e.g. what are the phase transitions of quantum field-theoretic machine learning algorithms? How do they behave when they interact with external fields?
2. There is an overlap with work in mathematical foundations of quantum field theory. Lattice field theory is also a way to rigorously define quantum fields in a mathematical manner (and now their machine learning counterparts). (Construction of quantum fields from Markoff fields, E. Nelson, J. Funct. Anal. 12, 97 (1973))

Conclusions

1. What one needs to do machine learning is simply a probability distribution. Lattice field theories therefore emerge as natural machine learning algorithms. We can investigate machine learning as a physical concept within quantum field theory: e.g. what are the phase transitions of quantum field-theoretic machine learning algorithms? How do they behave when they interact with external fields?
2. There is an overlap with work in mathematical foundations of quantum field theory. Lattice field theory is also a way to rigorously define quantum fields in a mathematical manner (and now their machine learning counterparts). (Construction of quantum fields from Markoff fields, E. Nelson, J. Funct. Anal. 12, 97 (1973))
3. Lattice field theory is inherently a computational research field. Easy to implement quantum field-theoretic machine learning algorithms and to study their physics computationally. (Renormalization Group in Field Theories with Quantum Quenched Disorder, V. Narovlansky and O. Aharony, Phys. Rev. Lett. 121, 071601 (2018))

Conclusions

1. What one needs to do machine learning is simply a probability distribution. Lattice field theories therefore emerge as natural machine learning algorithms. We can investigate machine learning as a physical concept within quantum field theory: e.g. what are the phase transitions of quantum field-theoretic machine learning algorithms? How do they behave when they interact with external fields?
2. There is an overlap with work in mathematical foundations of quantum field theory. Lattice field theory is also a way to rigorously define quantum fields in a mathematical manner (and now their machine learning counterparts). (Construction of quantum fields from Markoff fields, E. Nelson, J. Funct. Anal. 12, 97 (1973))
3. Lattice field theory is inherently a computational research field. Easy to implement quantum field-theoretic machine learning algorithms and to study their physics computationally. (Renormalization Group in Field Theories with Quantum Quenched Disorder, V. Narovlansky and O. Aharony, Phys. Rev. Lett. 121, 071601 (2018))
4. Experimental implementations of machine learning based on quantum field theory? An interesting read: The Hintons in your Neural Network: a Quantum Field Theory View of Deep Learning, R. Bondesan, M. Welling, arXiv:2103.04913 (2021).

Thank you for your attention!

Quantum field-theoretic machine learning, D. Bachtis, G. Aarts and B. Lucini, Phys. Rev. D 103, 074510, (arXiv:2102.09449).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813942

EuroPLEx
European network for Particle physics, Lattice field theory and Extreme computing