Disorder, phase transitions, and (lattice) quantum field-theoretic machine learning

Dimitrios Bachtis

Outline:

- Disorder

  What are the connections between machine learning and disordered systems?

  Prototypical types of disorder/inhomogeneity, Ising model, Ising spin glass, Ising random field, the φ^4 spin glass, replica theory and the overlap order parameter, intuitive connections with machine learning.

Outline:

- Disorder

  What are the connections between machine learning and disordered systems?

  Prototypical types of disorder/inhomogeneity, Ising model, Ising spin glass, Ising random field, the φ^4 spin glass, replica theory and the overlap order parameter, intuitive connections with machine learning.

- Quantum field-theoretic (and statistical-mechanical) machine learning

  How can we unify statistical mechanics, quantum field theory, and machine learning under a common mathematical framework?

  The (local) Markov property, Hammersley-Clifford theorem, Nelson construction of quantum field theories, φ^4 Markov fields, φ^4 multi-agent systems, reinforcement learning, φ^4 neural networks and their generalization of the Bernoulli-Bernoulli restricted Boltzmann machine [see Nobel Prize 2024 (Hinton)], applications.

Outline:

- **Disorder**

  What are the connections between machine learning and disordered systems?

  Prototypical types of disorder/inhomogeneity, Ising model, Ising spin glass, Ising random field, the φ^4 spin glass, replica theory and the overlap order parameter, intuitive connections with machine learning.

- **Quantum field-theoretic (and statistical-mechanical) machine learning**

  How can we unify statistical mechanics, quantum field theory, and machine learning under a common mathematical framework?

  The (local) Markov property, Hammersley-Clifford theorem, Nelson construction of quantum field theories, φ^4 Markov fields, φ^4 multi-agent systems, reinforcement learning, φ^4 neural networks and their generalization of the Bernoulli-Bernoulli restricted Boltzmann machine [see Nobel Prize 2024 (Hinton)], applications.

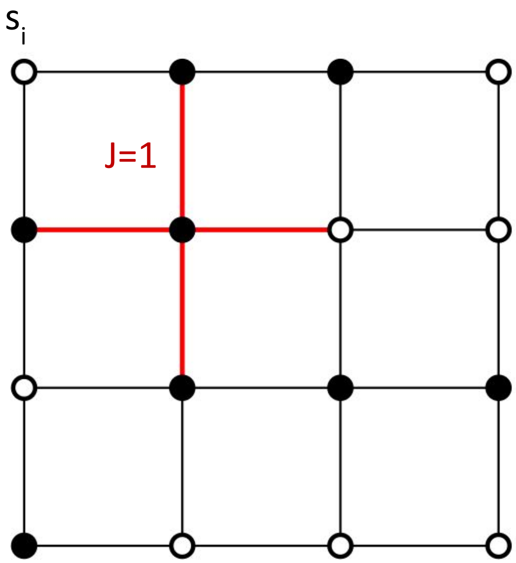
- **Phase transitions of machine learning algorithms**

  Are there ***genuine*** phase transitions in neural networks?

  Second-order phase transitions during the learning process of neural networks, order parameters, scaling and universality in probabilistic machine learning, a quest for unification of machine learning under universality classes, numerical examples.

The Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j$$

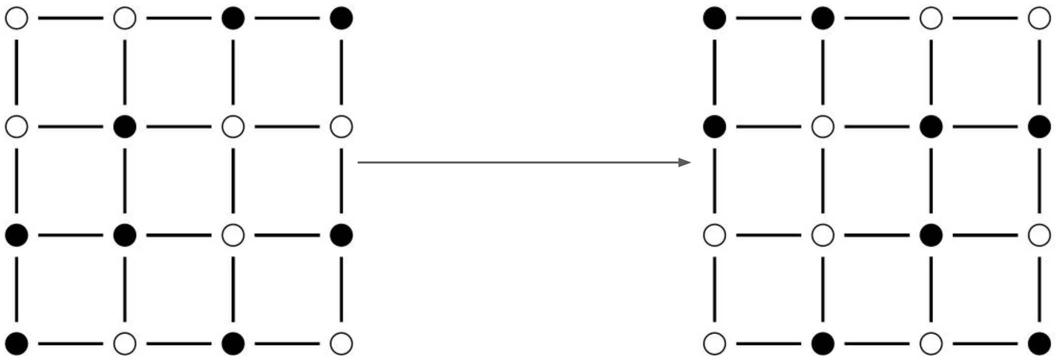


The Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j$$

Z_2 or Reflection symmetry

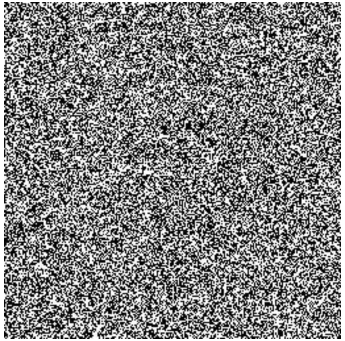
$s_i \to -s_i$



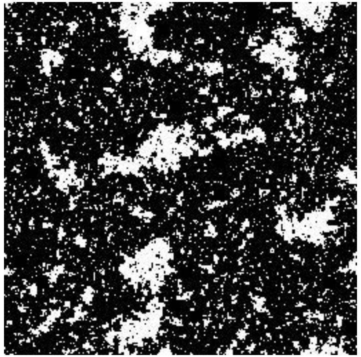
The Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j$$

Z_2 Symmetric phase



Critical region



Broken Z_2 symmetry phase

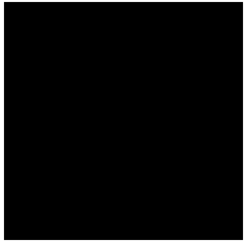
The Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j$$

Explicit breaking of the Z_2 symmetry via a magnetic field

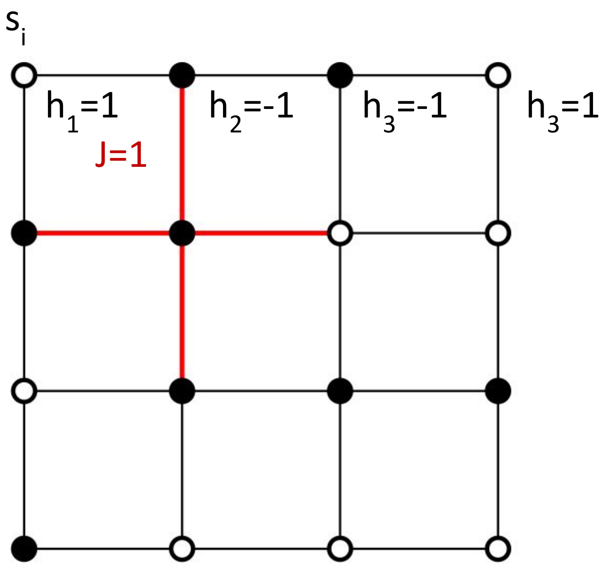
$$H = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i$$

Broken symmetry phase



The random field Ising model

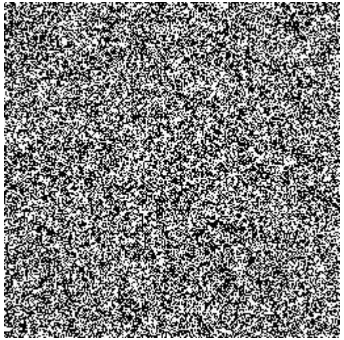
$$H = -J\sum_{\langle ij \rangle} s_i s_j - \sum_i h_i s_i$$



The random field Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i h_i s_i$$

Symmetric Phase



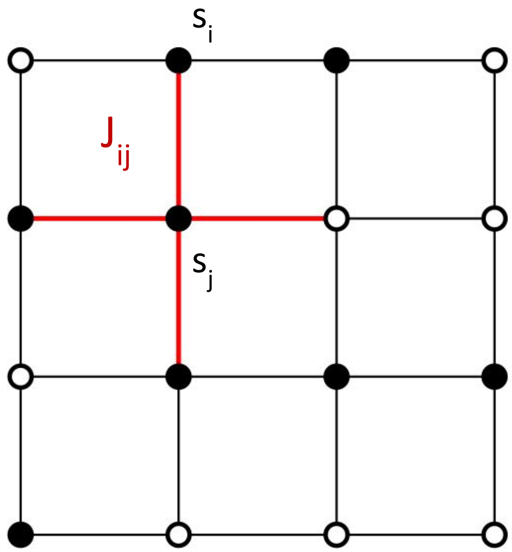
Broken symmetry phase

Random Magnetic Fields, Supersymmetry, and Negative Dimensions,G. Parisi, N. Sourlas, Phys. Rev. Lett. 43, 744, (1979).

Parisi-Sourlas Supersymmetry in Random Field Models, A. Kaviraj, S. Rychkov, E. Trevisani, Phys. Rev. Lett. 129, 045701, (2022).

The Ising spin glass
(the Edwards-Anderson model)

$$H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j$$

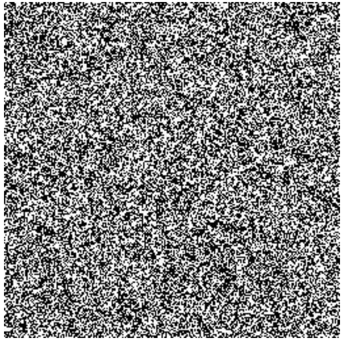


The Ising spin glass
(the Edwards-Anderson model)

$$H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j$$

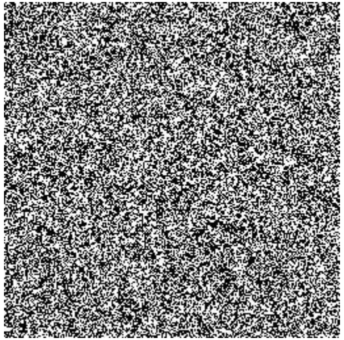
Symmetric phase

Evolving
randomly oriented configurations



Broken symmetry phase

Frozen
randomly oriented configuration



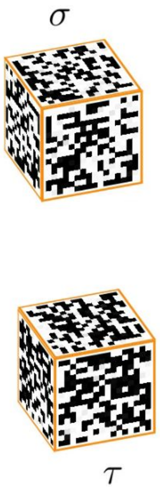
The Ising spin glass satisfies gauge invariance properties. For every nonzero function f:

$$s_i' \to f_i s_i \qquad J_{ij}' \to f_i^{-1} J_{ij} f_j^{-1}$$

The Ising spin glass

Real replicas

$$H_{\sigma,\tau} = H_\sigma + H_\tau = -\sum_{\langle ij \rangle} J_{ij}(s_i s_j + t_i t_j)$$



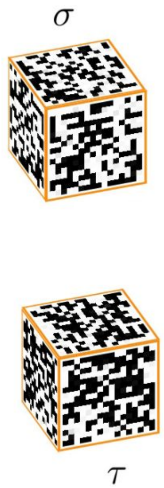
The Ising spin glass

Real replicas

$$H_{\sigma,\tau} = H_\sigma + H_\tau = -\sum_{\langle ij \rangle} J_{ij}(s_i s_j + t_i t_j)$$

Overlap variable

$$\rho_i = s_i t_i$$



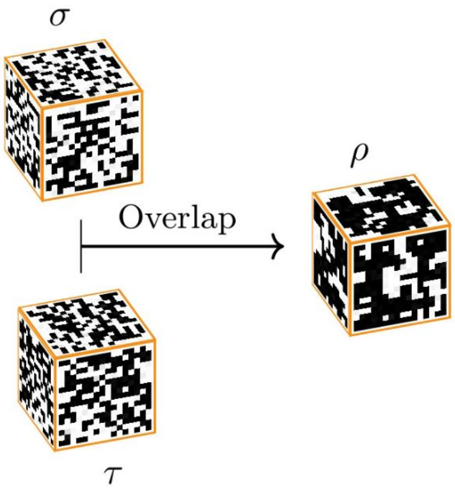
The Ising spin glass

Real replicas

$$H_{\sigma,\tau} = H_\sigma + H_\tau = -\sum_{\langle ij \rangle} J_{ij}(s_i s_j + t_i t_j)$$

Overlap variable

$$\rho_i = s_i t_i$$



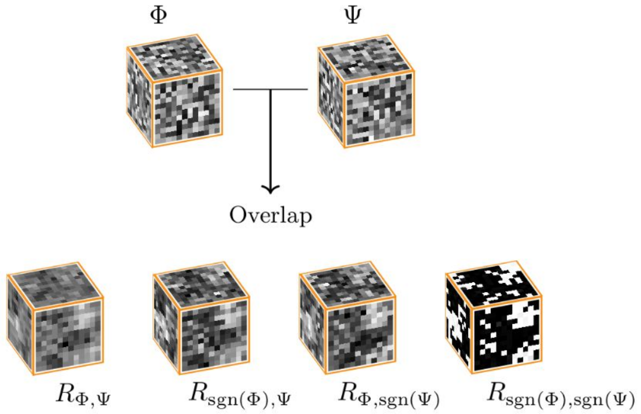
The φ^4 glass

$$S_\Phi = -\sum_{\langle ij \rangle} J_{ij}\phi_i\phi_j + \left(d + \frac{\mu^2}{2}\right)\sum_i \phi_i^2 + \frac{\lambda}{4}\sum_i \phi_i^4$$

Disordered Lattice Glass ϕ^4 Quantum Field Theory , D. Bachtis, arxiv2407.06569.

The φ^4 glass

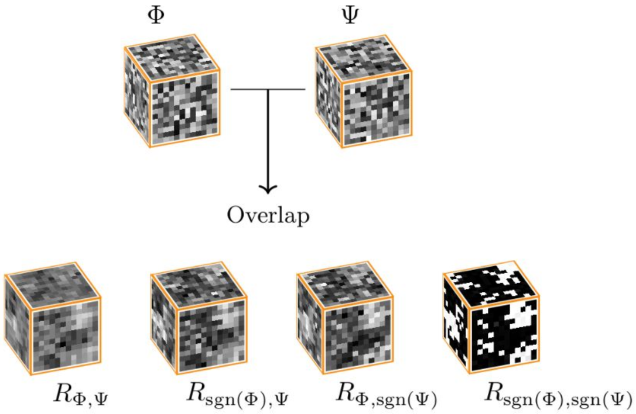
$$S_\Phi = -\sum_{\langle ij \rangle} J_{ij}\phi_i\phi_j + \left(d + \frac{\mu^2}{2}\right)\sum_i \phi_i^2 + \frac{\lambda}{4}\sum_i \phi_i^4$$



Disordered Lattice Glass ϕ^4 Quantum Field Theory , D. Bachtis, arxiv2407.06569.

The φ⁴ glass

$$S_\Phi = -\sum_{\langle ij\rangle} J_{ij}\phi_i\phi_j + \left(d + \frac{\mu^2}{2}\right)\sum_i \phi_i^2 + \frac{\lambda}{4}\sum_i \phi_i^4$$



Proof that the φ⁴ spin glass reduces to the Edwards-Anderson model

$$Z_\Phi < \left[\int_{-\infty}^{\infty} \exp\left[-\left(\frac{\mu^2}{2}\phi_i^2 + \frac{\lambda}{4}\phi_i^4\right)\right] d\phi_i\right]^V < \infty$$

$$\lim_{\lambda\to\infty} \frac{\sqrt{\lambda}}{\sqrt{\pi}} \exp[-\lambda(\phi^2 - 1)^2] = \delta(\phi^2 - 1)$$

Disordered Lattice Glass ϕ^4 Quantum Field Theory , D. Bachtis, arxiv2407.06569.

Why neural networks resemble disordered systems?

The random field Ising model

$$H = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i h_i s_i$$

The Ising spin glass

$$H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j$$

The random field Ising model

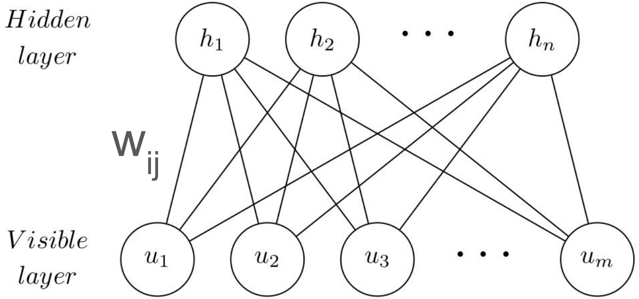
$$H = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i h_i s_i$$

The Ising spin glass

$$H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j$$

Neural networks, e.g. the restricted Boltzmann machine, have a Hamiltonian that looks like this:

$$H = -\sum_{i,j} w_{ij} u_i h_j - \sum_i a_i u_i - \sum_j b_j h_j$$



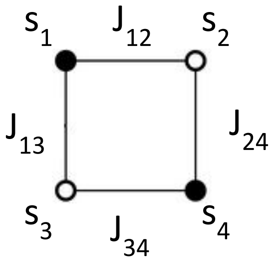
Quantum field-theoretic machine learning

How can we unify statistical mechanics, quantum field theory, and machine learning under a common mathematical framework?

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

Quantum field-theoretic machine learning

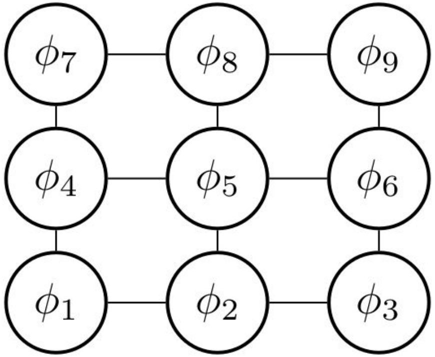
We require some form of representation to construct a probability distribution. We are going to use a finite set Λ that we express as a graph $G(\Lambda,e)$ where e is the set of edges in G.



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.

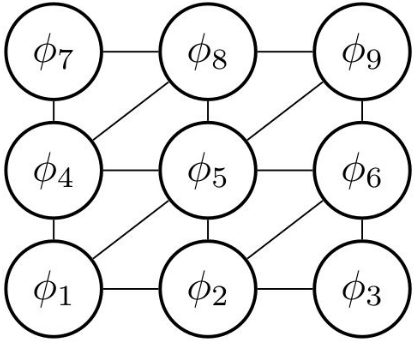
Quantum field-theoretic machine learning

On the square lattice a maximal clique is a two-site edge.



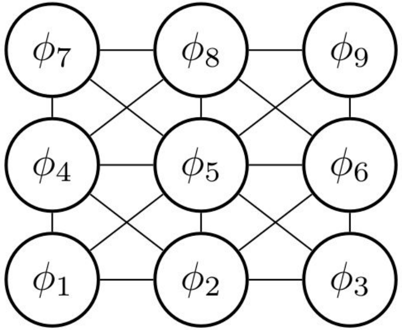
(a)

On a triangular lattice a maximal clique is a triangle.

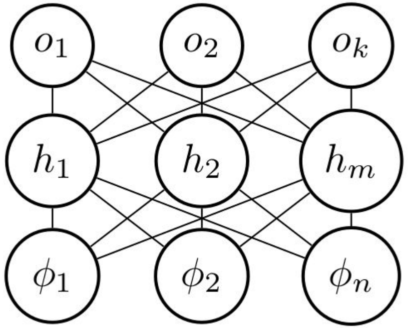


(b)

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 defined by an edge connection.

Quantum field-theoretic machine learning

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

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

Quantum field-theoretic machine learning

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.

Quantum field-theoretic machine learning

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

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

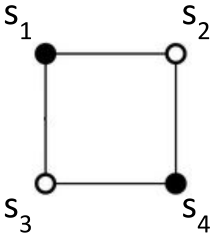
1. Factors are the fundamental building blocks of probability distributions.

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

Quantum field-theoretic machine learning

Factorization of probability distributions

$$H = -\sum_{\langle ij \rangle} s_i s_j$$

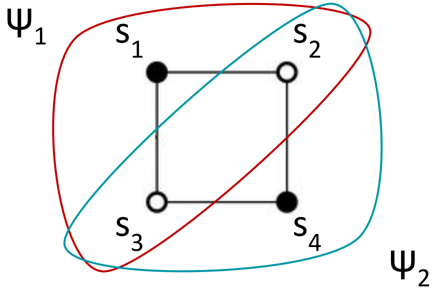


Quantum field-theoretic machine learning

Factorization of probability distributions

$$H = - \sum_{\langle ij \rangle} s_i s_j$$

$$\psi_1 = \exp[s_1 s_2 + s_1 s_3]$$



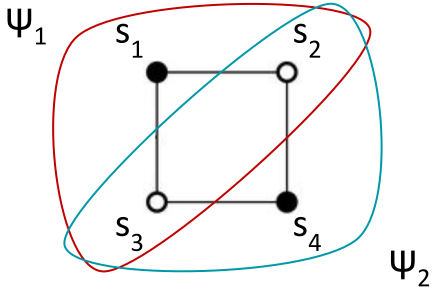
$$\psi_2 = \exp[s_2 s_4 + s_3 s_4]$$

Quantum field-theoretic machine learning

Factorization of probability distributions

$$H = -\sum_{\langle ij \rangle} s_i s_j$$

$$\psi_1 = \exp[s_1 s_2 + s_1 s_3]$$



$$\psi_2 = \exp[s_2 s_4 + s_3 s_4]$$

$$p(s_1, s_2, s_3, s_4) = \frac{\psi_1 \psi_2}{\sum \psi_1 \psi_2}$$

Hammersley-Clifford theorem

A strictly positive distribution p satisfies the local Markov property of an undirected graph G:

$$p(\phi_i|(\phi_j)_{j\in\Lambda-i}) = p(\phi_i|(\phi_j)_{j\in\mathcal{N}_i})$$

if and only if p can be represented as a product of strictly positive potential functions ψ_c over G, one per maximal clique c, i.e.

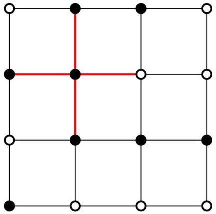
$$p(\phi) = \frac{1}{Z}\prod_{c\in C}\psi_c(\phi), \quad Z = \int_{\boldsymbol{\phi}}\prod_{c\in C}\psi_c(\boldsymbol{\phi})d\boldsymbol{\phi}$$

where Z is the partition function and **φ** are all possible states of the system.

Quantum field-theoretic machine learning

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

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



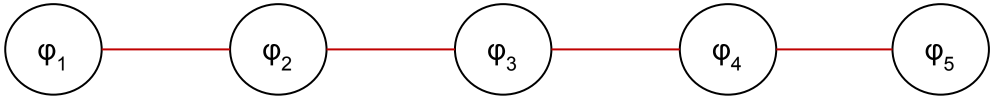
The probability distribution is expressed as a product of strictly positive potential functions ψ, over each maximal clique:

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

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

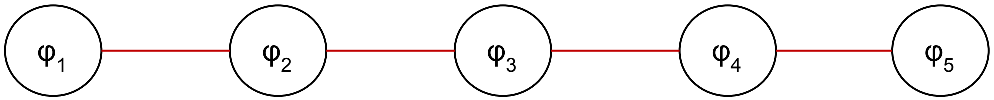
Quantum field-theoretic machine learning

The Markov property in a Markov chain

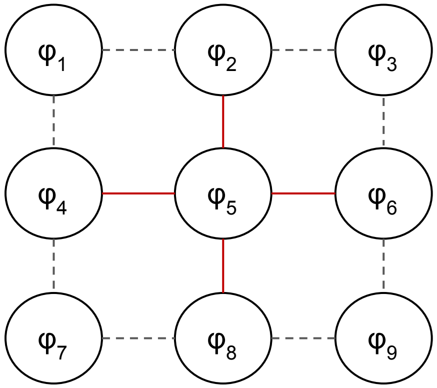


Quantum field-theoretic machine learning

The Markov property in a Markov chain



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



Representation

Hammersley-Clifford theorem

Markov fields on finite graphs and lattices, J. M. Hammersley, P. Clifford, (1971).

Proofs

1) A theorem about random fields, G. R. Grimmett, Bulletin of the London Mathematical Society, 5 (1): 81–84 (1973).
2) Generalized Gibbs states and Markov random fields, C. J. Preston, Advances in Applied Probability, 5 (2): 242–261, (1973).
3) Markov random fields and Gibbs random fields, S. Sherman, Israel Journal of Mathematics, 14 (1): 92–103, (1973).
4) Spatial interaction and the statistical analysis of lattice systems, J. Besag, Journal of the Royal Statistical Society, Series B, 36 (2): 192–236,, (1974).

Representation

Constructive field theory (see also J. Halverson talks!)

based on Garding-Wightman

Markov Property

Construction of quantum fields from Markoff fields, E. Nelson, J. Funct. Anal. 12, 97 (1973)

Reflection Positivity

Reflection Positivity Then and Now, A. Jaffe, arxiv:1802.07880 (2018)

Algebraic Quantum Field Theory

An Algebraic Approach to Quantum Field Theory, R. Haag, D. Kastler, Journal of Mathematical Physics, 5 (7), (1964)

Quantum field-theoretic machine learning

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

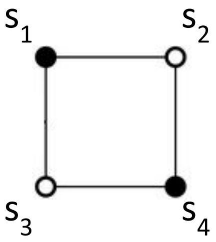
Quantum field-theoretic machine learning

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...

Quantum field-theoretic machine learning

We will now implement the lattice φ^4 field theory/Ising model to represent interactions of agents who aim to buy or sell a stock in the context of financial markets.

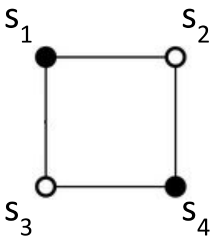


Lattice ϕ^4 field theory as a multi-agent system of financial markets, D. Bachtis, arxiv:2411.15813.

D. Bachtis, D. Berman, in preparation.

Quantum field-theoretic machine learning

We will now implement the lattice φ^4 field theory/Ising model to represent interactions of agents who aim to buy or sell a stock in the context of financial markets.

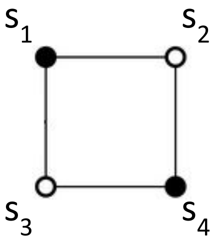


We want to model two interactions:

1) People tend to imitate the behaviour of their neighbours

Quantum field-theoretic machine learning

We will now implement the lattice φ^4 field theory/Ising model to represent interactions of agents who aim to buy or sell a stock in the context of financial markets.



We want to model two interactions:

1) People tend to imitate the behaviour of their neighbours
2) People are influenced by the opinion of the majority or the minority

Quantum field-theoretic machine learning

Our model will then represent two types of agents

1) Fundamentalists, agents who invest based on the fundamental value of an asset. For example, a fundamentalist might decide to buy an asset when it is undervalued or sell an asset when it resides above the fundamental value.

Quantum field-theoretic machine learning

Our model will then represent two types of agents

1) Fundamentalists, agents who invest based on the fundamental value of an asset. For example, a fundamentalist might decide to buy an asset when it is undervalued or sell an asset when it resides above the fundamental value.
2) Chartists or noise traders, agents who invest based on emerging trends (i.e. by relying on "charts") rather than the fundamental value of an asset. A chartist who is invested in an undervalued asset might decide to spontaneously sell the asset due to a momentary and emerging trend.

Quantum field-theoretic machine learning

We want to model two interactions:

1) People tend to imitate the behaviour of their neighbours

This information already exists in the Hamiltonian of the Ising model or the lattice action of the φ^4 theory due to the Z_2 symmetry breaking phase transition!

Quantum field-theoretic machine learning

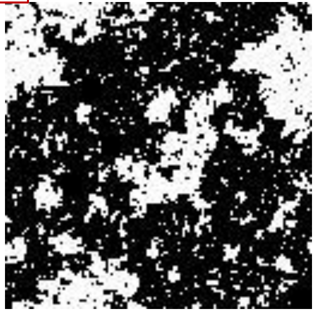
We want to model two interactions:

1) People tend to imitate the behaviour of their neighbours

This information already exists in the Hamiltonian of the Ising model or the lattice action of the φ^4 theory due to the Z_2 symmetry breaking phase transition!

$$H = -J \sum_{\langle ij \rangle} s_i s_j$$

$$S = -\sum_{\langle ij \rangle} \phi_i \phi_j + \left(d + \frac{\mu^2}{2}\right) \sum_i \phi_i^2 + \frac{\lambda}{4} \sum_i \phi_i^4$$



Quantum field-theoretic machine learning

We want to model two interactions:

2) People are influenced by the opinion of the majority or the minority

Quantum field-theoretic machine learning

We want to model two interactions:

2) People are influenced by the opinion of the majority or the minority

- This information does not exist in the Hamiltonian or lattice action and we need to explicitly introduce it.

Quantum field-theoretic machine learning

We want to model two interactions:

2) People are influenced by the opinion of the majority or the minority

- This information does not exist in the Hamiltonian or lattice action and we need to explicitly introduce it.
- We observe that the intensive magnetization, i.e. the sum over all the degrees of freedom in the lattice, represents the opinion of the majority.

Quantum field-theoretic machine learning

We want to model two interactions:

2) People are influenced by the opinion of the majority or the minority

- This information does not exist in the Hamiltonian or lattice action and we need to explicitly introduce it.
- We observe that the intensive magnetization, i.e. the sum over all the degrees of freedom in the lattice, represents the opinion of the majority.
- We can then introduce a term in the Hamiltonian that forces an agent to take into consideration the opinion of the majority or minority via the magnetization.

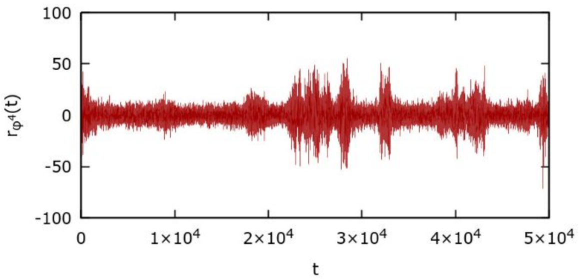
$$-s_i \left| \frac{1}{N} \sum_{j, j \in \Lambda} s_j \right|$$

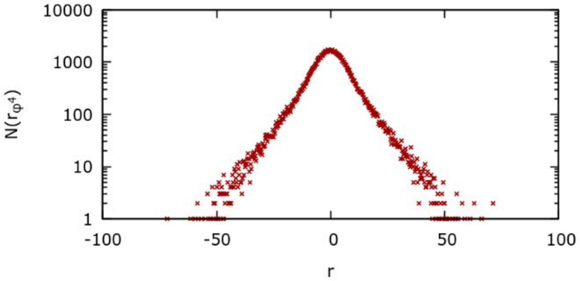
Quantum field-theoretic machine learning

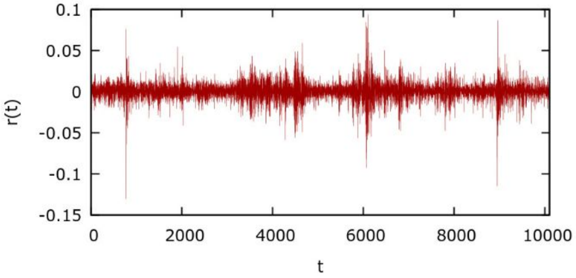
The physical behaviour that we obtain from the multi-agent φ^4 model of financial markets:

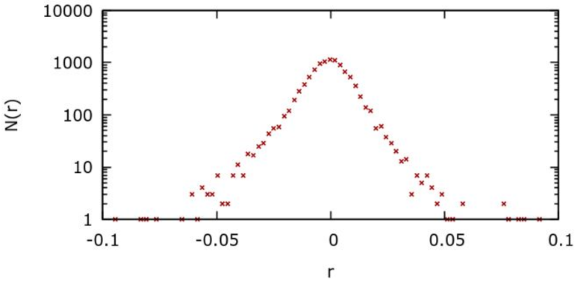
Quantum field-theoretic machine learning

The φ^4 multi-agent model can reproduce nontrivial aspects of financial markets, called stylized facts, such as volatility clustering and fat-tailed probability distribution of returns.









Quantum field-theoretic machine learning

To employ quantum field-theoretic or statistical mechanical algorithms for other machine learning tasks, we need to define a form of asymmetric distance between two probability distributions, called the Kullback-Leibler divergence:

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

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

Quantum field-theoretic machine learning

1.

We can use the first definition of the Kullback-Leibler divergence:

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

to minimize the asymmetric distance between our model probability distribution p and an unknown empirical probability distribution q for which we have data available.

Quantum field-theoretic machine learning

1.

We can use the first definition of the Kullback-Leibler divergence:

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

to minimize the asymmetric distance between our model probability distribution p and an unknown empirical probability distribution q for which we have data available.

2.

We can use the second definition of the Kullback-Leibler divergence:

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

to minimize the asymmetric distance between our model probability distribution p and a known empirical probability distribution q for which we do not have data available.

Quantum field-theoretic machine learning

Let's first investigate the case of a known empirical probability distribution q for which we do not have data available.

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

We want to minimize the Kullback-Leibler divergence.

We consider that q(φ) is the probability distribution of a system described by a new Hamiltonian or action A.

By minimizing the KL divergence we will make the two probability distributions equal. We can then use the probability distribution p(φ;θ) of the φ^4 theory to draw samples from the target distribution q(φ).

Quantum field-theoretic machine learning

We substitute the two probability distributions in the Kullback-Leibler divergence to obtain:

Gibbs-Bogoliubov-Feynman Inequality

$$F_{\mathcal{A}} \leq \langle \mathcal{A} - S \rangle_{p(\phi;\theta)} + F \equiv \mathcal{F},$$

<> denotes expectation value

There are two important observations on the above equation:

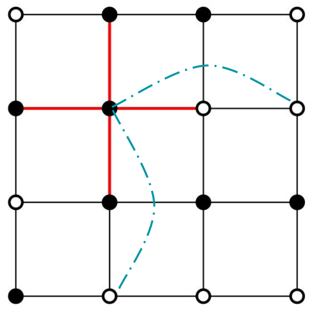
1. It sets a rigorous upper bound to the calculation of the free energy of the system with action A.
2. The bound is dependent entirely on samples drawn from the distribution p(φ;θ) of the φ^4 theory.

Quantum field-theoretic machine learning

As a very simple proof-of-principle demonstration, we can consider a disordered lattice action of the φ^4 theory:

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

that is able to represent more intricate actions, such as actions that include longer range interactions



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

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

We can use the first definition of the Kullback-Leibler divergence:

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

to minimize the asymmetric distance between our model probability distribution p and an unknown empirical probability distribution q for which we have data available.

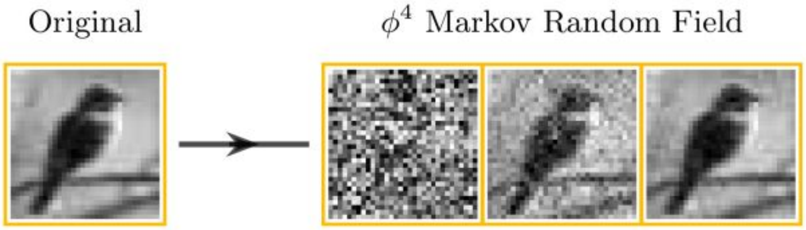
Quantum field-theoretic machine learning

We consider as an example a problem of image segmentation:

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.

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

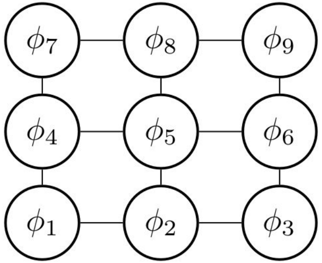
Quantum field-theoretic machine learning



Quantum field-theoretic machine learning

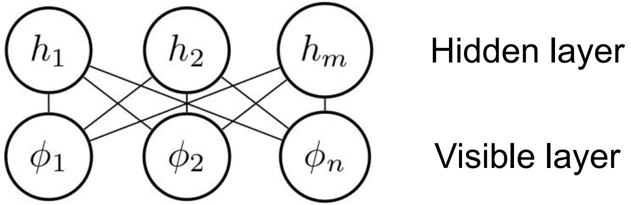
So far we have investigated the behaviour of a

φ^4 Markov random field



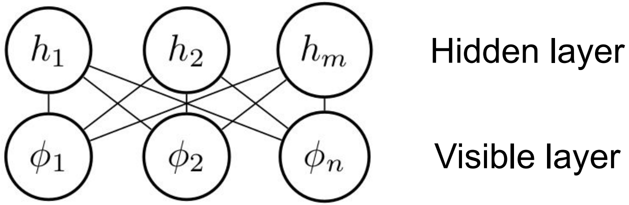
We are now interested in investigating the behaviour of a

φ^4 neural network



Quantum field-theoretic machine learning

φ⁴ neural network



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

The Boltzmann probability distribution of the φ^4 neural network is now a joint probability distribution of the visible φ and the hidden h variables:

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

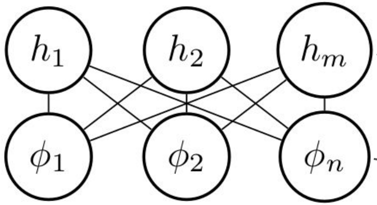
Quantum field-theoretic machine learning

From the joint probability distribution of the φ^4 neural network

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

We are able to marginalize out variables and derive marginal probability distributions $p(\varphi;\theta)$ and $p(h;\theta)$:

Hidden layer



Visible layer

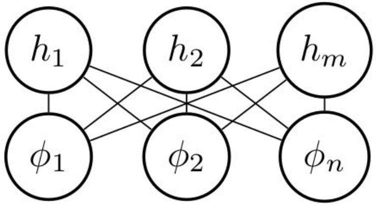
$$p(h; \theta) = \int_{\boldsymbol{\phi}} p(\boldsymbol{\phi}, h; \theta) d\boldsymbol{\phi} = \frac{\int_{\boldsymbol{\phi}} \exp[-S(\boldsymbol{\phi}, h; \theta)] d\boldsymbol{\phi}}{\int_{\boldsymbol{\phi}, \boldsymbol{h}} \exp[-S(\boldsymbol{\phi}, \boldsymbol{h}; \theta)] d\boldsymbol{\phi} d\boldsymbol{h}},$$

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

Quantum field-theoretic machine learning

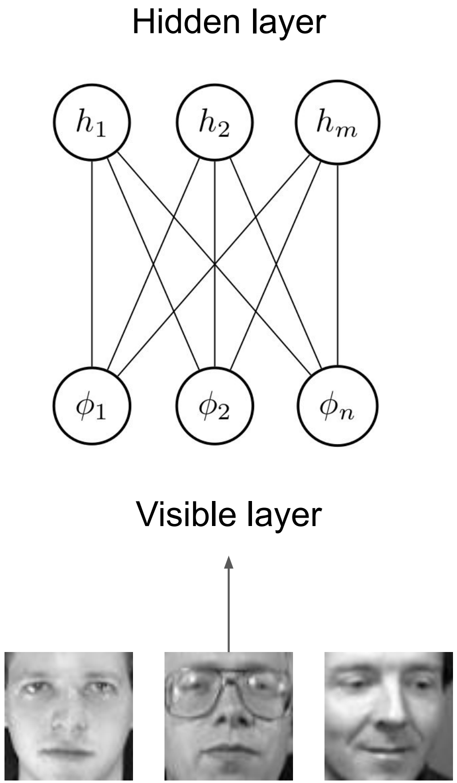
We now want to minimize the asymmetric distance between the empirical probability distribution q(φ) and the marginal probability distribution p(φ;θ):

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



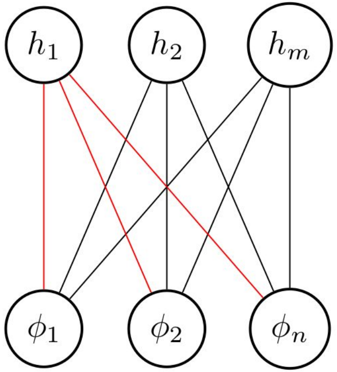
In other words, we want to reproduce the dataset in the visible layer. The hidden layer will then uncover dependencies on the data.

Quantum field-theoretic machine learning



Quantum field-theoretic machine learning

Hidden layer



Examples of the coupling constants w_{ij} with j fixed

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

Visible layer

Quantum field-theoretic machine learning

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)

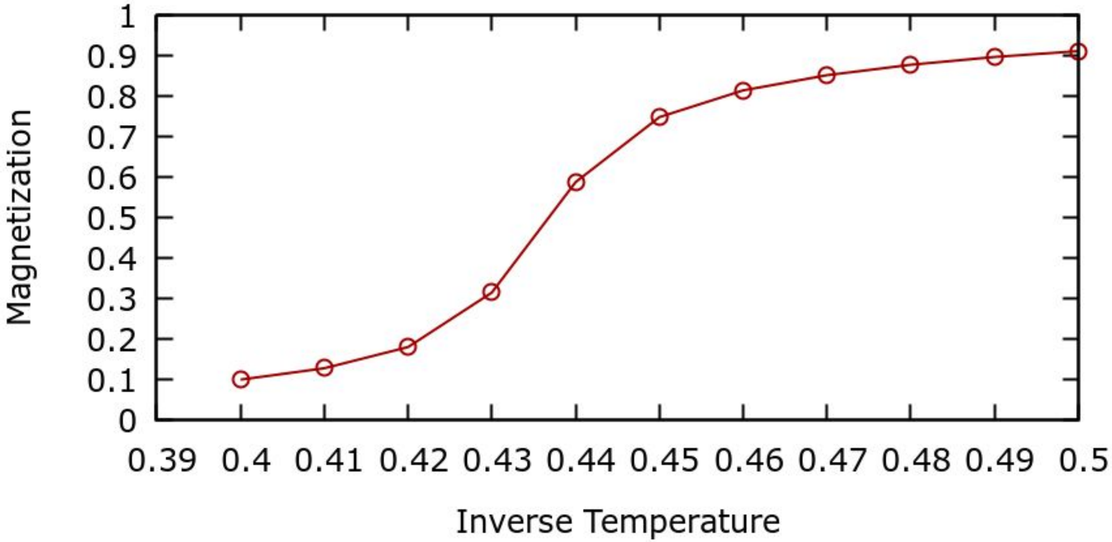
Quantum field-theoretic machine learning, D. Bachtis, G. Aarts and B. Lucini, Phys. Rev. D 103, 074510, (arXiv:2102.09449).

Are there "genuine" phase transitions in the learning process of neural networks?

Cascade of phase transitions in the training of energy-based models ,D. Bachtis, G. Biroli, A. Decelle, B. Seoane, Advances in Neural Information Processing Systems 37 (NeurIPS 2024).

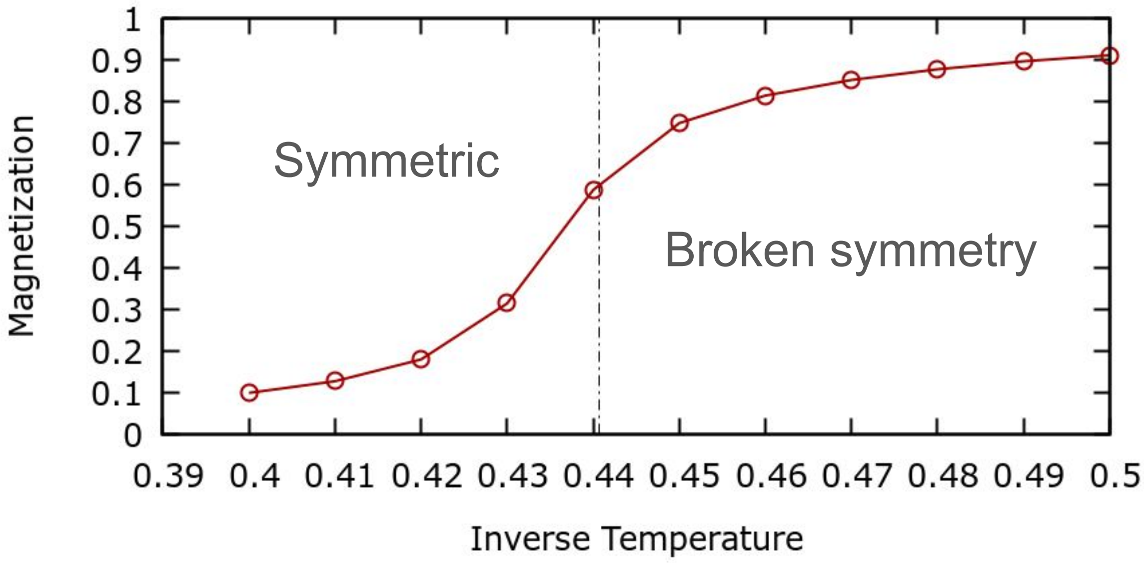
Phase transitions of machine learning algorithms

We saw that an order parameter is a quantity which can be used to distinguish between two phases of a system. For example in the case of the Ising model, the order parameter was the magnetization.



Phase transitions of machine learning algorithms

We saw that an order parameter is a quantity which can be used to distinguish between two phases of a system. For example in the case of the Ising model, the order parameter was the magnetization.



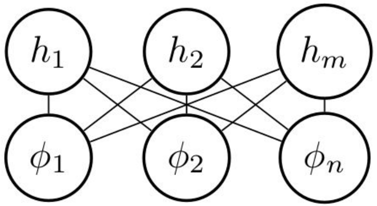
Phase transitions of machine learning algorithms

We consider again the case of the Bernoulli-Bernoulli restricted Boltzmann machine:

$$H = -\sum_{i,j} w_{ij} u_i h_j - \sum_i a_i u_i - \sum_j b_j h_j$$

And for simplicity we assume that $\alpha_i = b_j = 0$, and that the visible and hidden variables take values of {-1,1}

Hidden layer



Visible layer

Phase transitions of machine learning algorithms

To study mathematically the behaviour of the restricted Boltzmann machine:

1) We consider that important information about the learning process of a neural network is evidently encoded within the weight matrix. We are going to extract this information via a singular value decomposition.
2) We rely on the implementation of techniques from statistical mechanics, such as the replica method to infer the phase diagram of the neural network and, equivalently, extract information about the order parameters which characterize the phases of the system.

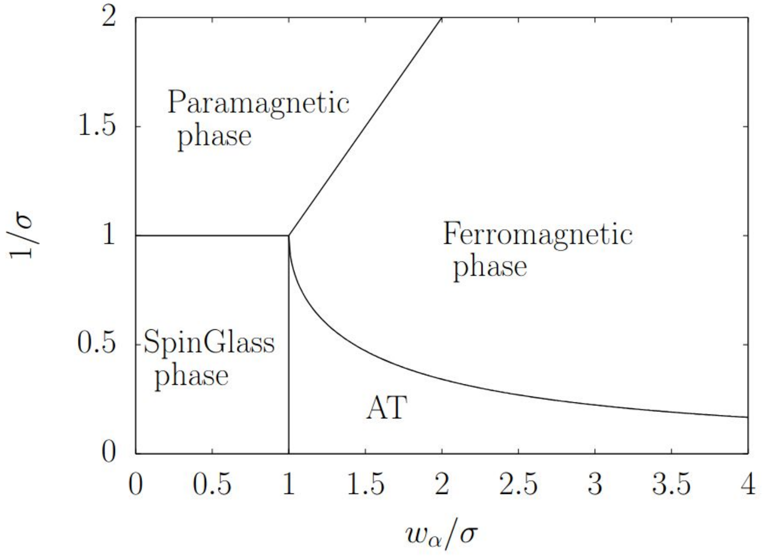
Thermodynamics of Restricted Boltzmann Machines and related learning dynamics, A. Decelle, G. Fissore, C. Furtlehner, J Stat Phys 172, 1576–1608 (2018).

Phase transitions of machine learning algorithms

Phase diagram for the very first phase transition of an RBM:

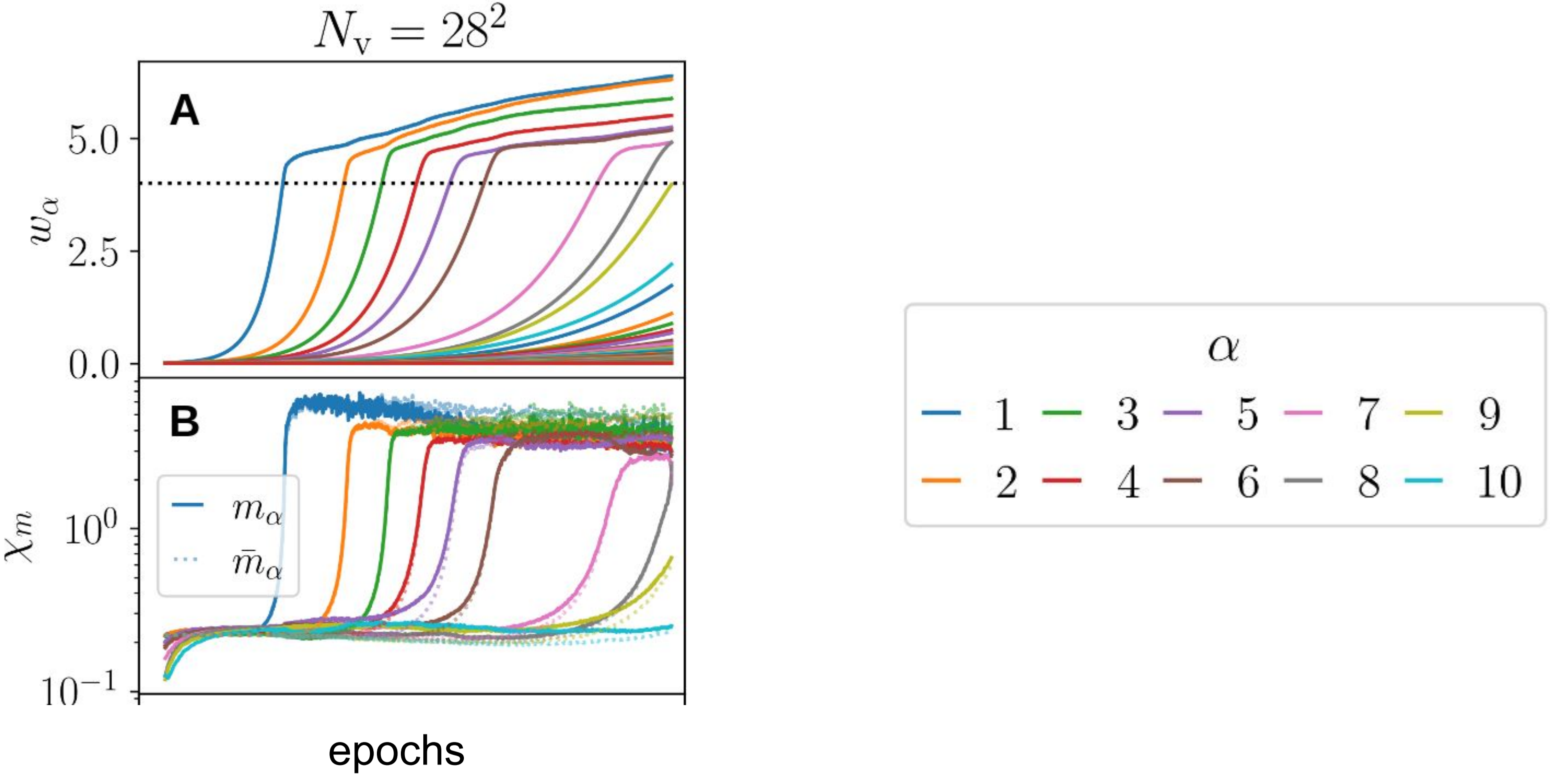
σ is the variance of the weight matrix

w_a is the largest eigenvalue as obtained from a SVD



Thermodynamics of Restricted Boltzmann Machines and related learning dynamics, A. Decelle, G. Fissore, C. Furtlehner, J Stat Phys 172, 1576–1608 (2018).

Phase transitions of machine learning algorithms



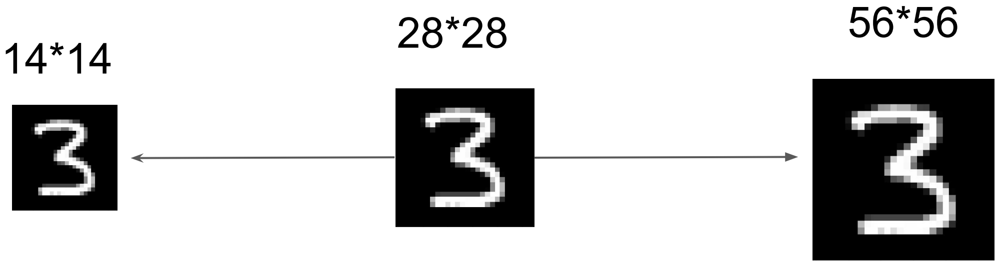
Cascade of phase transitions in the training of energy-based models ,D. Bachtis, G. Biroli, A. Decelle, B. Seoane, Advances in Neural Information Processing Systems 37 (NeurIPS 2024).

Phase transitions of machine learning algorithms

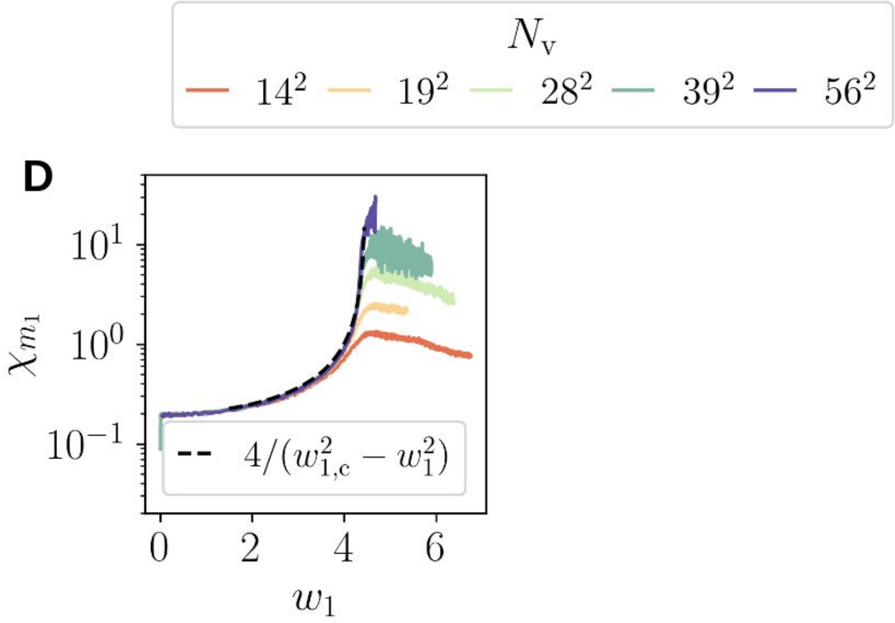
If a phase transition is “genuine” it should persist also to the thermodynamic limit

Phase transitions of machine learning algorithms

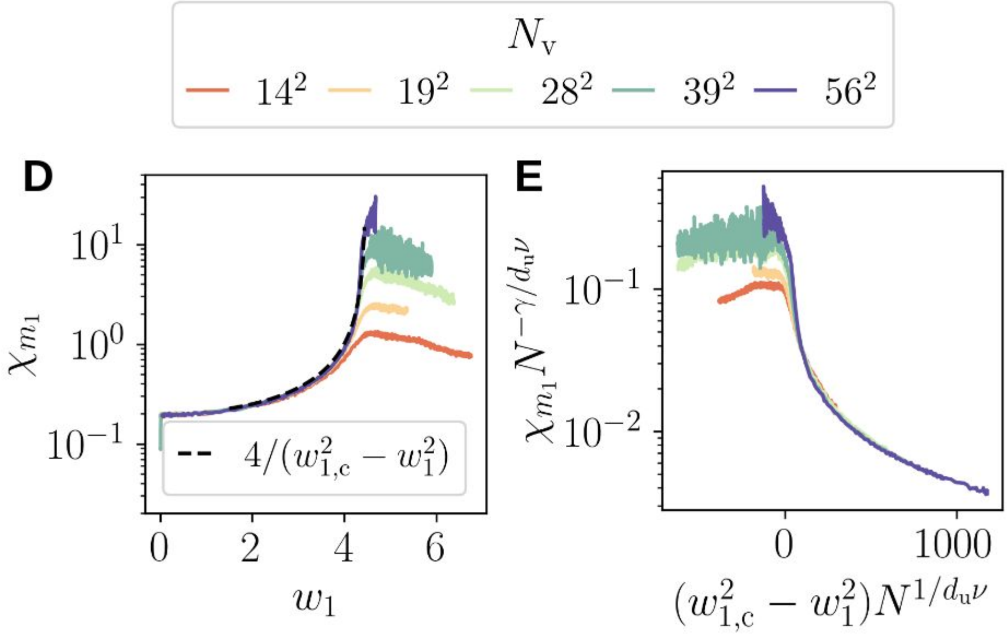
To investigate the scaling of the phase transitions one can progressively rescale the dataset:



Phase transitions of machine learning algorithms



Phase transitions of machine learning algorithms



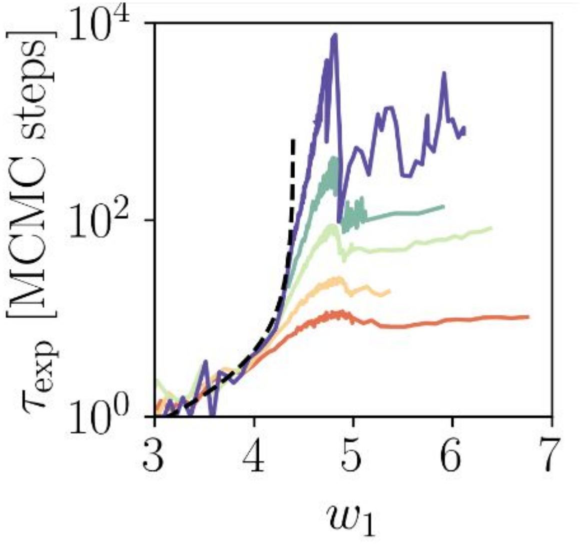
Numerical data are consistent with a mean-field universality class

Phase transitions of machine learning algorithms

Phase transitions might help us understand how machine learning algorithms function, but can we extract any other useful information from them?

Phase transitions of machine learning algorithms

Additional useful information about the learning process relates to the emergence of a critical slowing down effect:



Phase transitions of machine learning algorithms

A long-term ambition of studying phase transitions of neural networks is a potential quest for unification of machine learning algorithms based on universality classes:

Ising universality class:
Ising model, critical opalescence, liquid to gas phase transition, and many more…

Phase transitions of machine learning algorithms

A long-term ambition of studying phase transitions of neural networks is a potential quest for unification of machine learning algorithms based on universality classes:

Ising universality class:
Ising model, critical opalescence, liquid to gas phase transition, and many more…

Can we simplify the study of machine learning algorithms through universality?

Summary:

- **Disorder**

  What are the connections between machine learning and disordered systems?

  Prototypical types of disorder/inhomogeneity, Ising model, Ising spin glass, Ising random field, the φ^4 spin glass, replica theory and the overlap order parameter, intuitive connections with machine learning.

- **Quantum field-theoretic (and statistical-mechanical) machine learning**

  How can we unify statistical mechanics, quantum field theory, and machine learning under a common mathematical framework?

  The (local) Markov property, Hammersley-Clifford theorem, Nelson construction of quantum field theories, φ^4 Markov fields, φ^4 multi-agent systems, reinforcement learning, φ^4 neural networks and their generalization of the Bernoulli-Bernoulli restricted Boltzmann machine [see Nobel Prize 2024 (Hinton)], applications.

- **Phase transitions of machine learning algorithms**

  Are there ***genuine*** phase transitions in neural networks?

  Second-order phase transitions during the learning process of neural networks, order parameters, scaling and universality in probabilistic machine learning, a quest for unification of machine learning under universality classes, numerical examples.

Thank you for your attention!