Quantum fields and machine learning

Gert Aarts

with Dimitrios Bachtis and Biagio Lucini

GGI, April 2022

Introduction

- past four years or so has seen a rapid rise of applications of machine learning (ML) in lattice field theory and other areas of theoretical physics
- very much in exploratory phase: *anything goes*
- recent overview/forward look paper for SNOWMASS

arXiv > hep-lat > arXiv:2202.05838

Search...

Help | Advanced S

High Energy Physics – Lattice

[Submitted on 10 Feb 2022]

Applications of Machine Learning to Lattice Quantum Field Theory

Denis Boyda, Salvatore Calì, Sam Foreman, Lena Funcke, Daniel C. Hackett, Yin Lin, Gert Aarts, Andrei Alexandru, Xiao-Yong Jin, Biagio Lucini, Phiala E. Shanahan

There is great potential to apply machine learning in the area of numerical lattice quantum field theory, but full exploitation of that potential will require new strategies. In this white paper for the Snowmass community planning process, we discuss the unique requirements of machine learning for lattice quantum field theory research and outline what is needed to enable exploration and deployment of this approach in the future.

Outline

in this talk: quantum field-theoretical machine learning

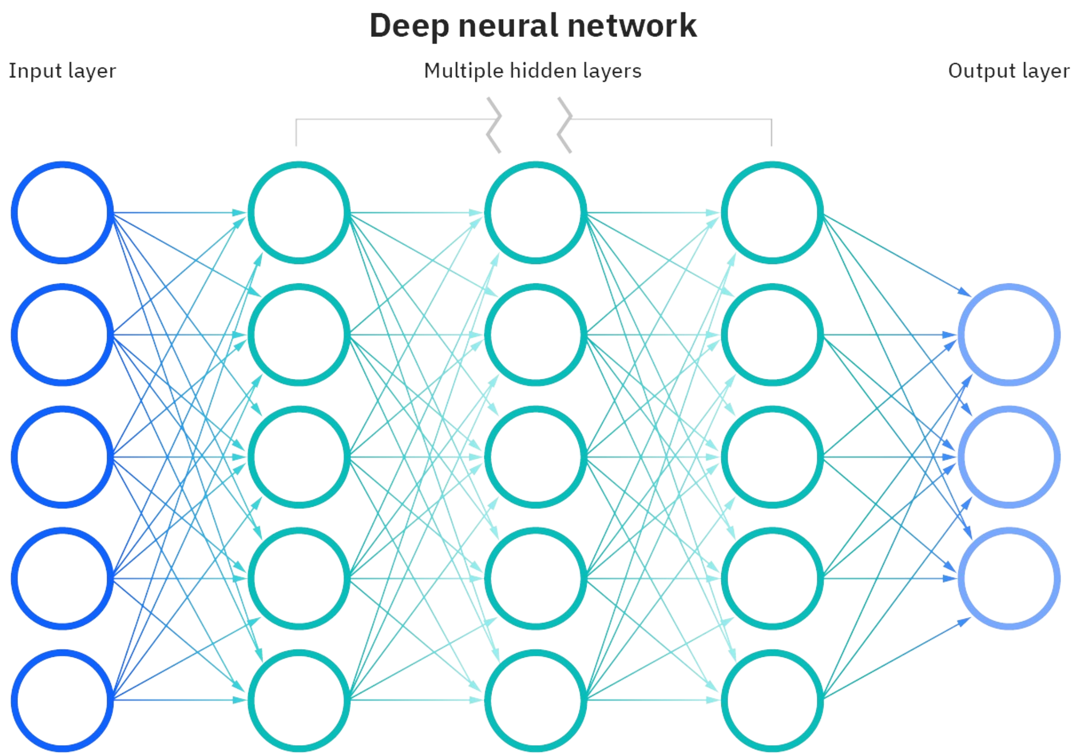
- new conceptual ideas to explore

- motivation
- scalar fields as Markov random fields
- some examples

✓ Quantum field-theoretic machine learning
Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]
Dimitrios Bachtis, GA and Biagio Lucini

ML, neural networks, deep learning

- many degrees of freedom ("neurons'') associated with sites of a multi-layered network
- weights w_{ij} (connecting the neurons) and biases b_i (on-site) are tunable parameters
- learning: parameters are adjusted by minimising some cost or loss function (e.g. mean-squared error)
- neural network should then reproduce training task and generalise (predict for unseen data)



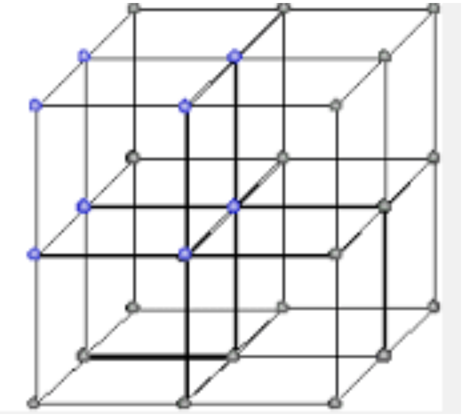
ML, neural networks, lattice field theory

- many degrees of freedom (“neurons’’) associated with sites of a multi-layered network
- weights w_{ij} (connecting the neurons) and biases b_i (on-site) are tunable parameters
- learning: parameters are adjusted by minimising some cost or loss function (e.g. mean-squared error)
- neural network should then reproduce training task and generalise (predict for unseen data)

- many degrees of freedom (“lattice fields’’) associated with sites of a *d*-dimensional lattice
- nearest neighbour (kinetic energy) and local couplings (potential) are external parameters
- configurations are generated by minimising the quantum action: Euclidean lattice field theory
- observables are computed using (independent, thermalised) configurations

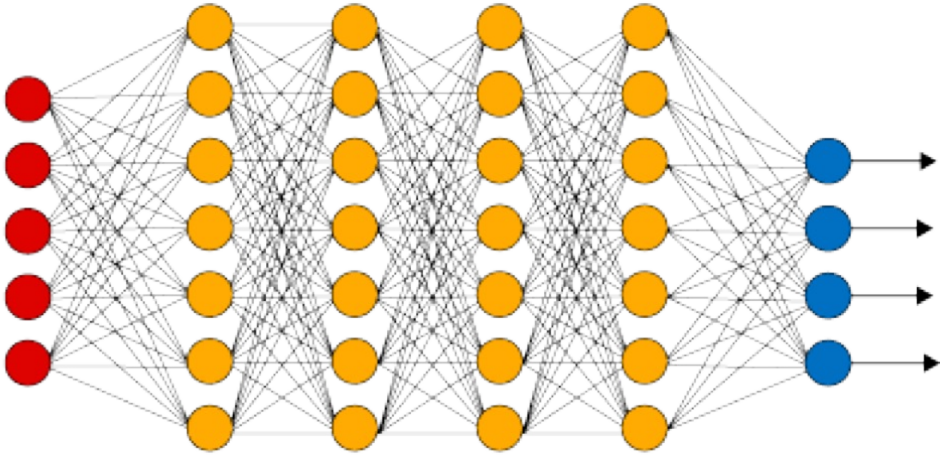
Quantum field-theoretical machine learning

- superficial similarities are obvious
- can this be made more precise?



use for common advantages:

- design ML algorithms
- develop synergies with lattice field theories
- apply ML to LFT, use LFT to develop ML



Main idea

derive machine learning algorithms from discretized Euclidean field theories

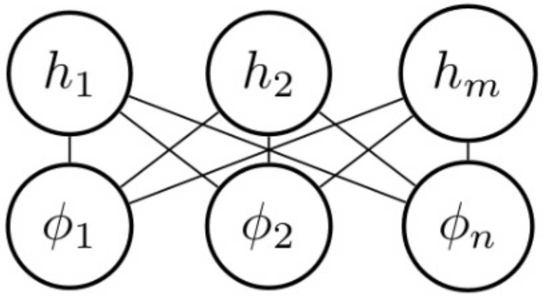
φ^4 scalar field theory $\longleftrightarrow$ Markov random field

(Hammersley-Clifford theorem)

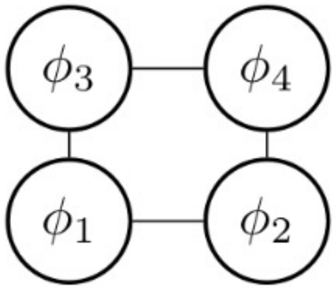
machine learning algorithms

Graphs, vertices, cliques

- graph has vertices and edges, e.g. the bipartite graph or square lattice below
- clique: a subset of points which are pairwise connected
- maximal clique *c*: no additional point can be included such that the resulting set is still a clique



for both graphs: each set of two vertices connected by a line is a maximal clique



Probability distribution $\mathrm{p}(\varphi)$

probability distribution $\mathrm{p}(\varphi)$ defined as product of nonnegative functions over maximal cliques:

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi)$$

then $\mathrm{p}(\varphi)$ satisfies local Markov property

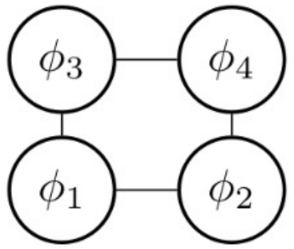
and set of random variables φ define a Markov random field

Theorem 1 (Hammersley-Clifford.) *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 nonnegative potential functions ψ_c over $\mathcal{G}$, one per maximal clique $c \in C$, i.e.,*

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi), \tag{2}$$

where $Z = \int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}$ is the partition function and $\boldsymbol{\phi}$ are all possible states of the system.

Partition function of local field theory



$$Z = \int d\varphi \, \exp(-S(\varphi)) \qquad p(\varphi) = \exp(-S(\varphi))/Z$$

- S depends on local (potential) and nearest neighbours (kinetic term)
- explicit example: 2d scalar field $\qquad \mathcal{L}_E = \frac{\kappa}{2}(\nabla\phi)^2 + \frac{\mu_0^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4$
- discretise and introduce local couplings: $\qquad S(\phi;\theta) = -\sum_{\langle ij\rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4$
- probability distribution: $\qquad p(\phi) = \frac{1}{Z}\prod_{c\in C}\psi_c(\phi)$
- satisfies Hammersley-Clifford theorem $\qquad \psi_c = \exp\left[-w_{ij}\phi_i\phi_j + \frac{1}{4}(a_i\phi_i^2 + a_j\phi_j^2 + b_i\phi_i^4 + b_j\phi_j^4)\right]$

➡ discretized φ^4 scalar field is a Markov random field

Lattice φ^4 theory as Markov random field

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

- set of coupling constants/variational parameters: $\theta = \{w_{ij}, a_i, b_i\}$
- search for an optimal set to complete ML tasks
- allowing them to be local/inhomogeneous increases expressivity
- from Euclidean QFT perspective: slightly strange theory
- but note: QFTs with random couplings or potentials

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

Quick recap

- variables φ define a Markov random field, our degrees of freedom
- tune the variational parameters $\theta = \{ w_{ij}, a_i, b_i\}$ in distribution $p(\varphi; \theta)$
- such that $p(\varphi; \theta)$ satisfies some conditions, i.e. completes a ML task

"Learn" target distributions $q(\varphi)$

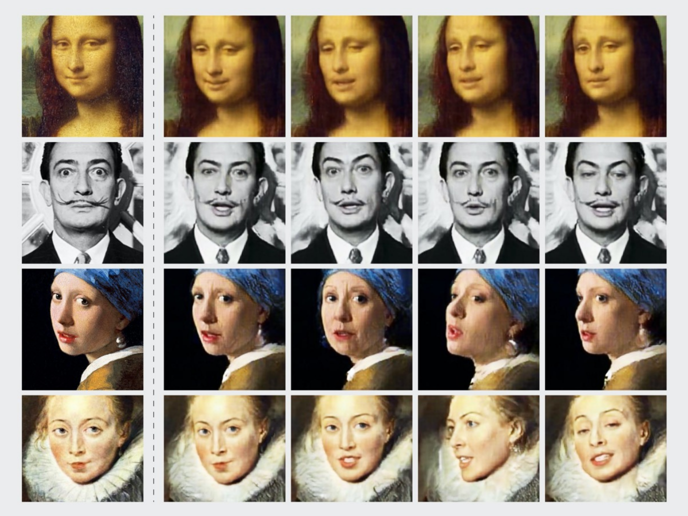
two possibilities:

- target distribution $q(\varphi)$ is known: construct $p(\varphi;\theta) \sim q(\varphi)$ by tuning θ
- why? might be difficult to generate data from $q(\varphi)$ but easier when using $p(\varphi;\theta)$

- target distribution $q(\varphi)$ is not known, but a lot of data is available
- construct $p(\varphi;\theta)$ by tuning θ
- why? use the distribution $p(\varphi;\theta)$ to create "fake" images,

  e.g. faces, paintings, music

deepfake



https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/

Approximate target distribution

- consider target probability distribution $q(\varphi)$
- goal: approximate this distribution with $p(\varphi;\theta)$ by tuning θ
- use asymmetric measure of distance between two probability distributions

Kullback-Leibler (KL) divergence:

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

- minimize KL divergence with respect to θ, by sampling configurations from $p(\varphi;\theta)$ and adapting θ using gradient descent

Variational principle

- write $q(\varphi) = \exp(-A(\varphi))/Z_A$ with $Z_A = \exp(-F_A)$
- and $p(\varphi;\theta) = \exp(-S(\varphi;\theta))/Z$ with $Z = \exp(-F)$
- then $F_A \leq \langle A - S\rangle_p + F \equiv \mathcal{F}$ (follows from KL divergence)
- upper bound on free energy of target system, RHS depends only on $p(\varphi;\theta)$ distribution of Markov random field

→ map an arbitrary system to a φ^4 scalar field theory by minimizing the KL distance

Variational principle, minimisation

- $F_A \leq \langle A - S \rangle_p + F \equiv \mathcal{F}$

- gradient-based approach to minimize the variational free energy wrt parameters θ:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{\partial \mathcal{F}}{\partial \theta}$$

$$\frac{\partial \mathcal{F}}{\partial \theta_i} = \langle \mathcal{A} \rangle \left\langle \frac{\partial S}{\partial \theta_i} \right\rangle - \left\langle \mathcal{A} \frac{\partial S}{\partial \theta_i} \right\rangle + \left\langle S \frac{\partial S}{\partial \theta_i} \right\rangle - \langle S \rangle \left\langle \frac{\partial S}{\partial \theta_i} \right\rangle$$

η learning rate

Target lattice field theory

- target action:

$$A(\phi) = -\sum_{\langle ij\rangle_{nn}} \phi_i\phi_j + g_4 \sum_{\langle ij\rangle_{nnn}} \phi_i\phi_j + \sum_i \left[(g_2 + ig_5)\phi_i^2 + g_3\phi_i^4\right]$$

nearest and next-to-nearest neighbours, complex mass parameter

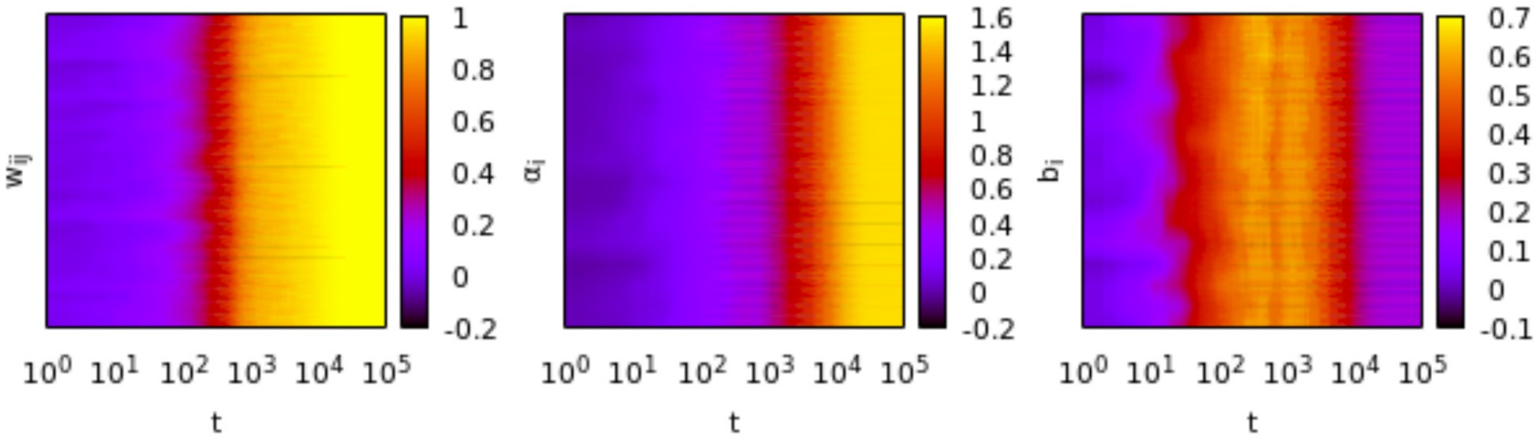
- ML action:

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

- generate configurations with $S(\varphi;\theta)$, use to further approximate $\mathrm{A}(\varphi)$

Example 1: φ^4 theory with constant couplings

- target action: standard φ^4 action with given couplings
- ML action: initialise $S(\varphi;\theta)$ with random couplings θ



- evolution of (inhomogeneous) couplings to expected (homogeneous) values

Example 2: φ^4 theory with nnn terms

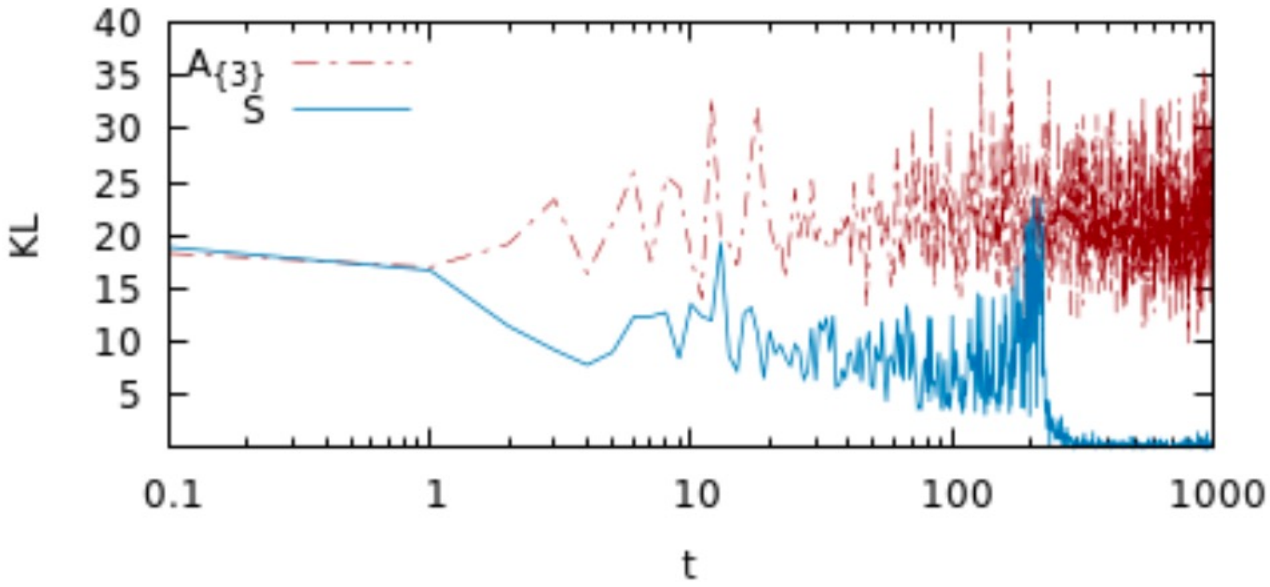
- target action: next-to-nearest neighbour terms, homogeneous couplings
- ML action (S) with no nnn terms but inhomogeneous couplings

  AND
- ML action (A_3) with no nnn terms and only homogeneous couplings

evolution of KL divergence

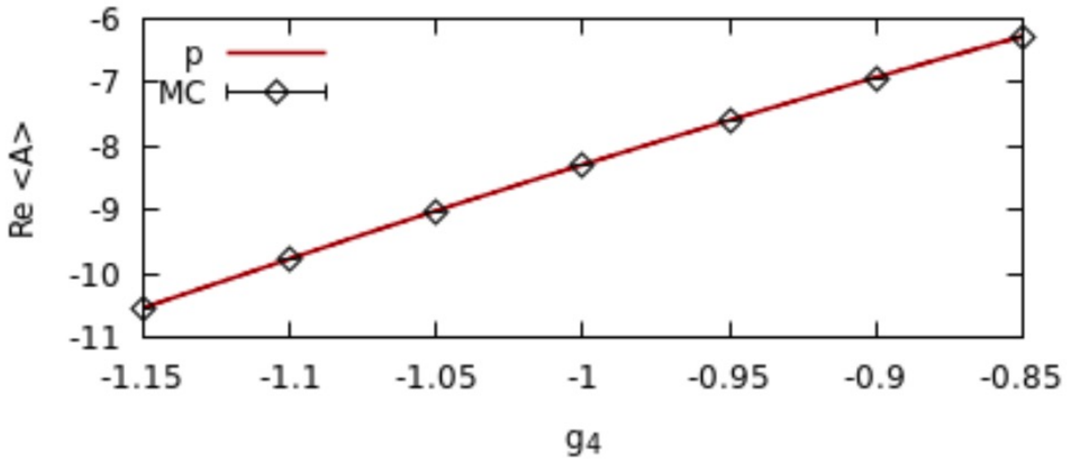
towards zero for action S

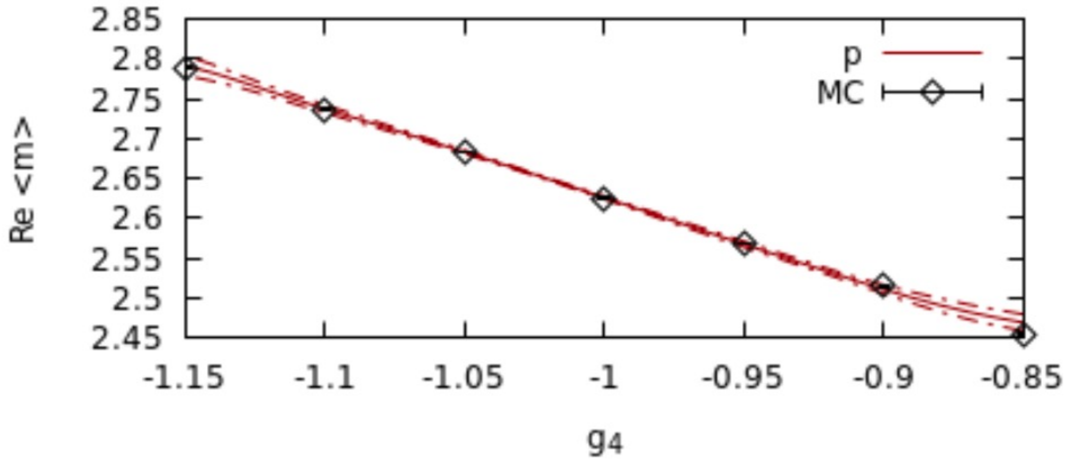
(but not for A_3)



Example 3: from $\mathbf{q}(\varphi)$ to $p(\varphi;\theta)$ and back

- after training: generate configs with ML action S, not original action A
- reweight by including corrective step $\sim \exp(S - A)$
- reweight in (complex) parameter space, starting from $g_4 = -1$



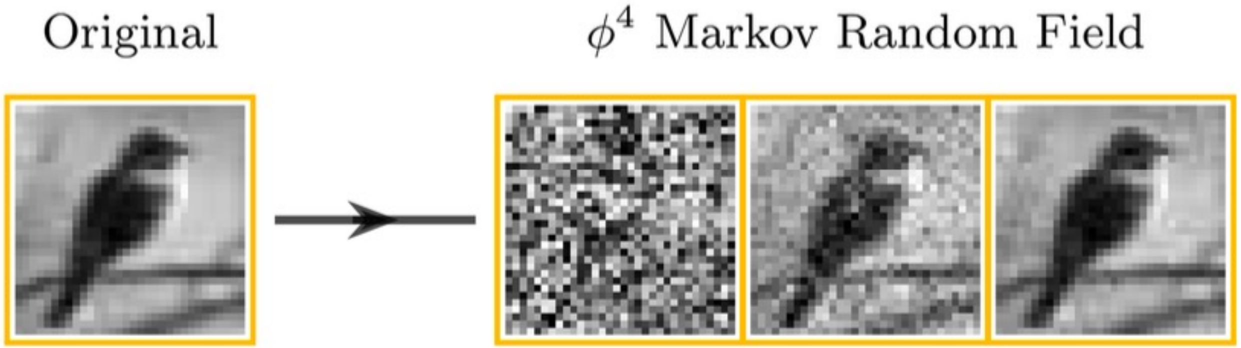


From data to target distribution

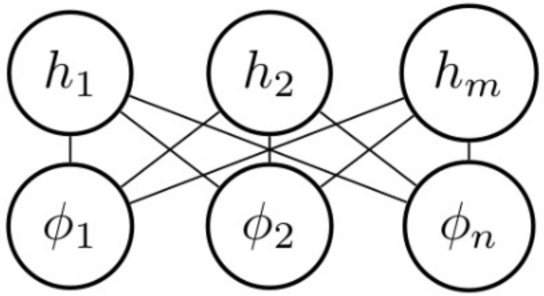
- what if we only have access to data, but not the underlying distribution?
- minimize opposite KL divergence:

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

- apply to problem in image analysis: given an image (i.e. the “data”), search for the optimal values of the local coupling constants in the φ^4 theory to reproduce the image as a configuration in the equilibrium distribution



φ^4 neural networks



- multiple layers in the neural network architecture
- bipartite graph: φ^4 as a variant of a restricted Boltzmann machine (RBM)
- joint probability distribution $p(\varphi, h; \theta)$ with action

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

- includes several cases:

  $b_i = n_j = 0$: Gaussian-Gaussian RBM

  $b_i = n_j = m_j = 0$ and $h_j \in \{-1,1\}$: Gaussian-Bernoulli RBM

  $m_j = n_j = 0$ and $h_j \in \{-1, 1\}$: φ^4-Bernoulli RBM (not yet studied, afawk)

Example: Olivetti faces dataset

- 64^2 visible units and 32 hidden units
- are there learned features? coupling constants w_{ij} for a fixed j

- neural network has learned hidden features: abstract face shapes and characteristics
- hidden units can serve as input to a new φ^4 neural network to progressively extract more abstract features in data

Summary and outlook

- ✓ machine learning algorithms from discretized Euclidean field theories
- ✓ in principle any system can be mapped to a φ^4 scalar field theory with inhomogeneous couplings by minimizing KL divergence
- ✓ inhomogeneous couplings: perspective of spin glasses? additional averaging over the space of couplings?
- ✓ develop a computational and mathematical framework of machine learning within quantum field theory