What is machine learning? How can it help?

Praneeth Netrapalli

Microsoft Research India

What is machine learning?

Identifying patterns in data

- Very broad – data may be of different types, from different domains

- Interplay between
  1. Statistics: how much data do we need to see to be sure of patterns?
  2. Computer science: mainly algorithms & optimization
  3. Domain experts: Prior knowledge. Rough idea of what/where/how to look for patterns

Structure of the talk

1. Frameworks/problems studied in machine learning
2. Real world examples
3. Algorithms/methods to solve these problems
4. Some applications in Physics

I. Supervised learning

Training data



Cat



Cat



Cat



Cat



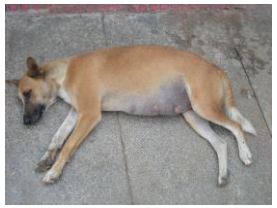
Cat



Cat



Dog



Dog



Dog



Dog



Dog



Dog

Test data



?



?



?

Challenge

100	123	20	19
79	67	45	30
33	48	82	104
16	20	74	93

Cat

43	56	26	24
16	35	11	46
46	74	64	34
24	35	67	65

Cat

37	54	74	61
64	35	56	26
35	27	87	54
36	46	54	39

Cat

38	15	35	36
22	48	21	43
6	53	55	72
26	75	41	43

Dog

65	84	81	45
76	43	23	45
67	23	76	43
21	48	43	32

Dog

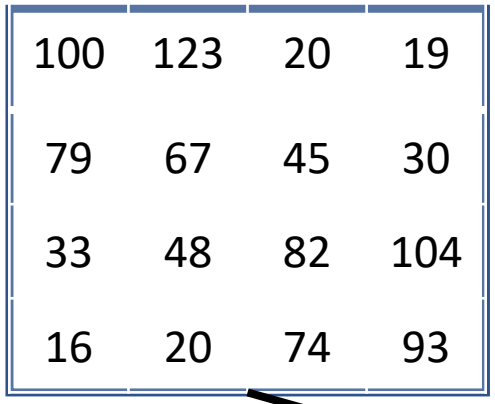
39	92	37	48
83	28	83	94
105	82	72	7
8	48	53	41

Dog

1	12	0	9
49	86	66	25
38	45	37	14
116	120	76	39

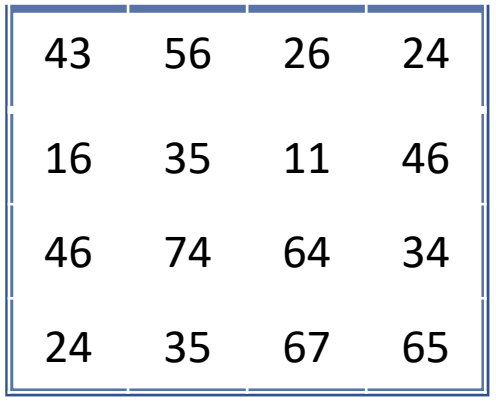
?

Problem formulation

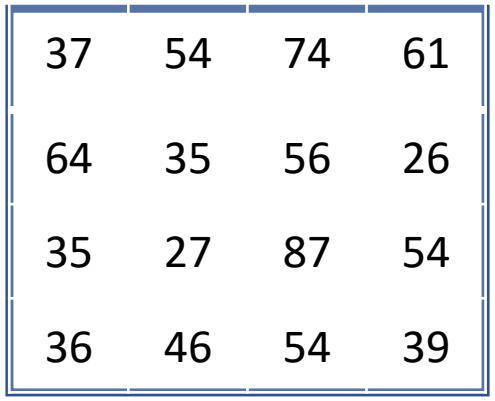


Cat $x_1 \in \mathbb{R}^{16}$

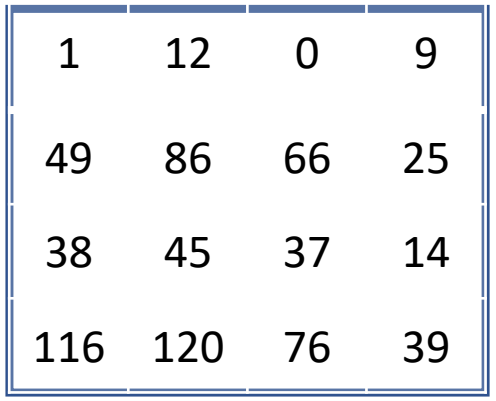
$y_1 \in \{-1,1\}$



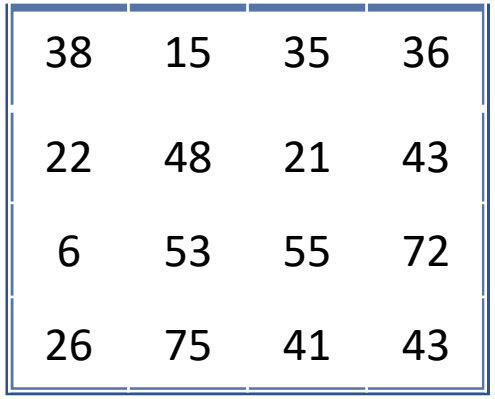
Cat (x_2, y_2)



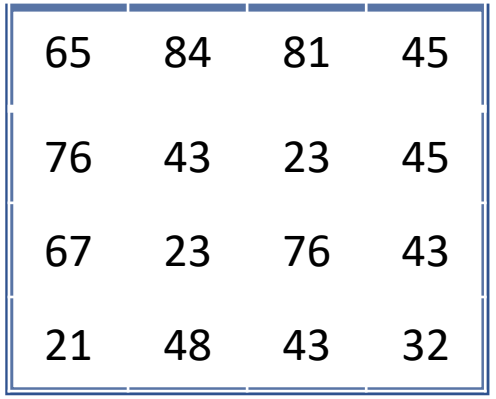
Cat (x_3, y_3)



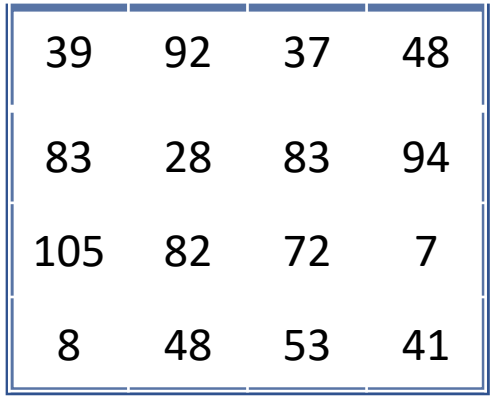
? $(x_7, ?)$



Dog (x_4, y_4)



Dog (x_5, y_5)

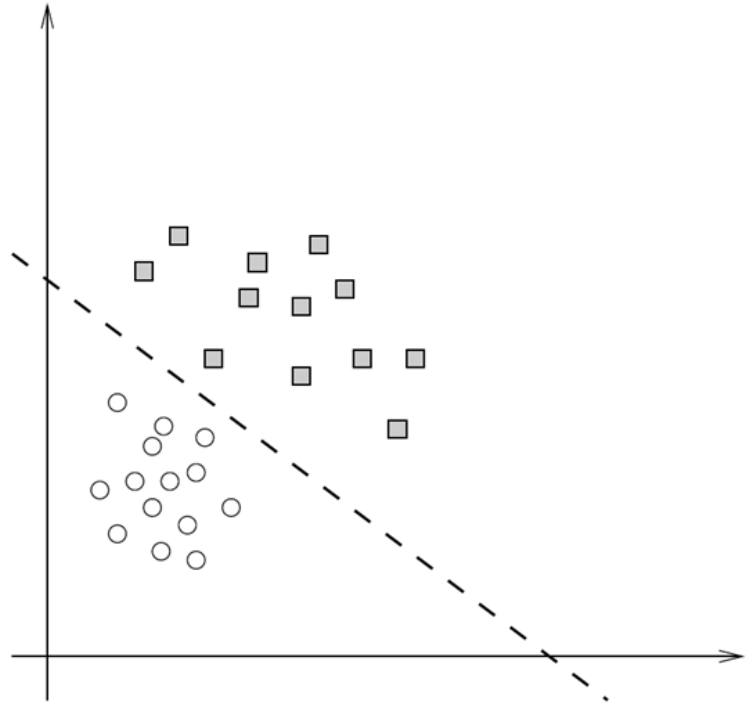


Dog (x_6, y_6)

Approach I – Linear classifiers

- **Problem**: Given $(x_1, y_1), \cdots, (x_n, y_n)$, predict y for a new x
- **Main idea**: approximate y using a **<u>linear</u>** function of x
- Perhaps $y \approx \text{sign}(\langle w, x \rangle + b)$ for some w, b
- **Formulation**: Find w, b such that for $i = 1, \cdots, n$

$$y_i = \text{sign}(\langle w, x_i \rangle + b)$$



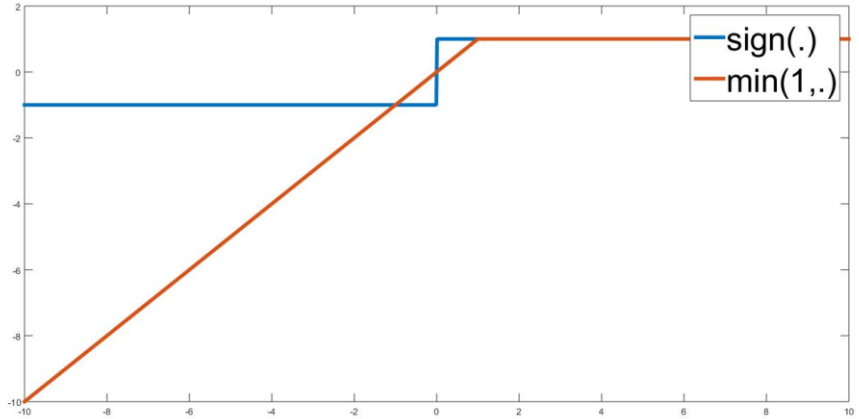
Approach I – Linear classifiers

- **Formulation**: Find w, b such that for $i = 1, \cdots, n$
$$y_i = \text{sign}(\langle w, x_i \rangle + b)$$
- (Computationally) easy to find such w, b if they exist
- Unlikely that y is such a simple function – definitely not the case in most practical applications
- **Reformulation**: $\max_{w,b} \sum_{i=1}^{n} y_i \cdot \text{sign}(\langle w, x_i \rangle + b)$
- Computationally very hard (NP hard)

Approach I – Linear classifiers

- **Yet another reformulation**: $\max_{w,b} \sum_{i=1}^{n} \min\left(1, y_i \cdot (\langle w, x_i \rangle + b)\right)$

- Performance may be worse than that of original reformulation, but **can be solved efficiently**

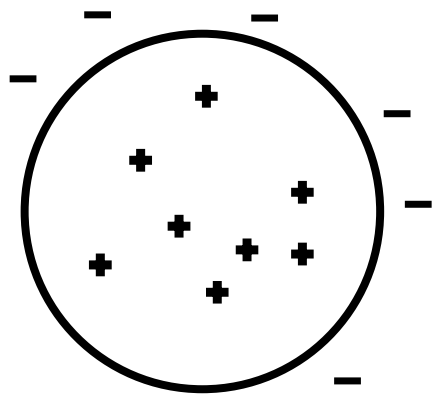


- This is the essential idea behind support vector machines (SVM)

- Efficiency vs performance: recurring theme in machine learning

Approach II – Nonlinear classifiers

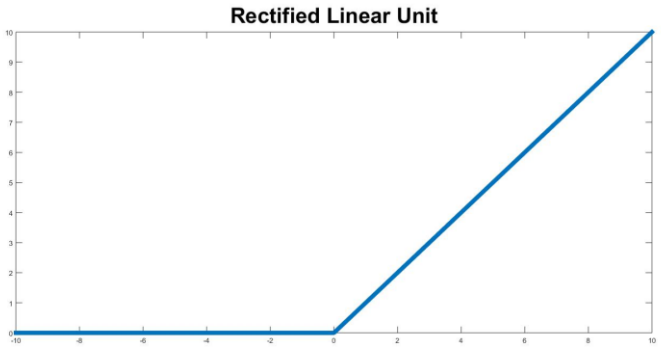
- Recall linear classifiers assume $y_i \approx \text{sign}(\langle w, x_i \rangle + b)$
- A better approximation would be $y_i \approx \text{sign}(\text{p}_d(x_i))$, where $\text{p}_d(\cdot)$ is a polynomial of degree d

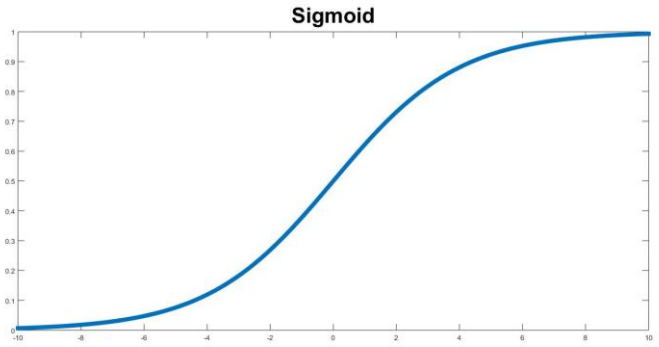


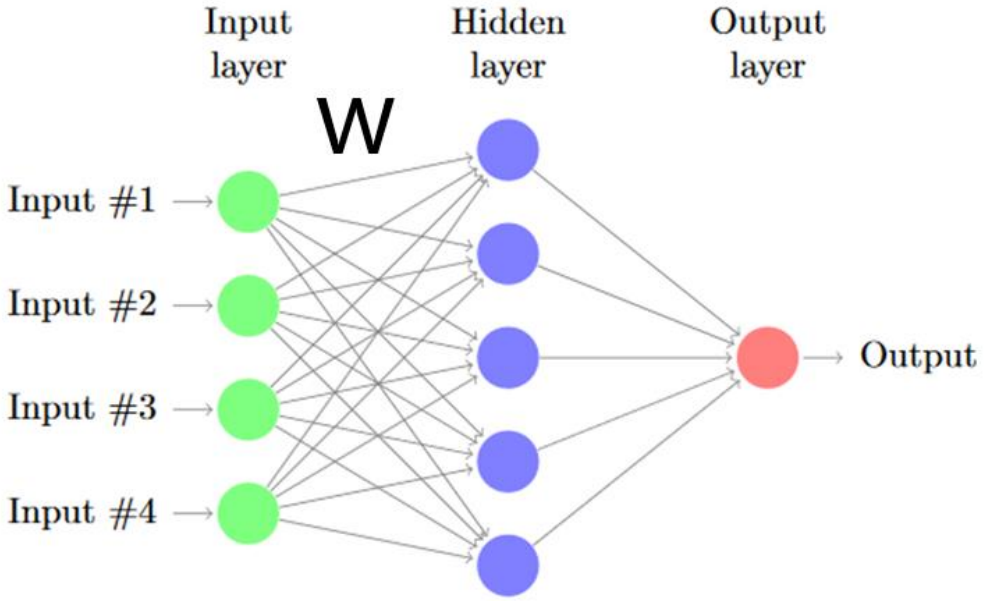
- **Formulation**: $\max_{\text{p}_d} \sum_{i=1}^{n} y_i \cdot \text{sign}\big(\text{p}_d(x_i)\big)$
- Computation time is **exponential** in d. People use what is known as "*kernel trick*". Works well in practice – kernel SVMs.

Approach III – Neural networks

- In general, can approximate $y_i \approx \text{sign}\left(f_w(x_i)\right)$ for some function class parameterized by w.
- For perceptual tasks such as vision, speech and language, function classes inspired by neuroscience make sense
- Each neuron computes a simple function



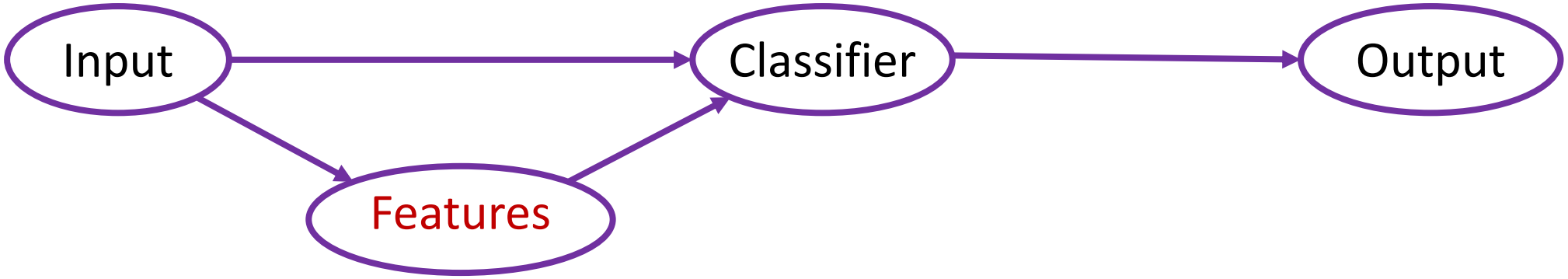




Are we done with supervised learning?

- **No!** If we apply this directly on pixel values, the results are not very impressive.
- Lots of symmetries in images – translation invariance, scale invariance, rotations/points of view etc.
- Other domains present their own set of symmetries.
- The functions we are using to approximate (SVMs, fully connected neural networks etc.) do not have these symmetries built into them.
- Major reason for poor performance on direct data

How do we fix it?

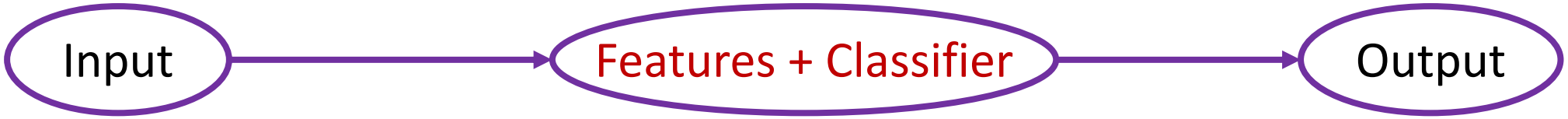


- E.g., magnitude of Fourier transform for translation invariance

$$\hat{f} \rightarrow \hat{f} e^{i2\pi t}$$

- Earlier approach was to use such hand-crafted features
- Very difficult and time consuming
- Can we minimize our effort in doing this?

How do we fix it?

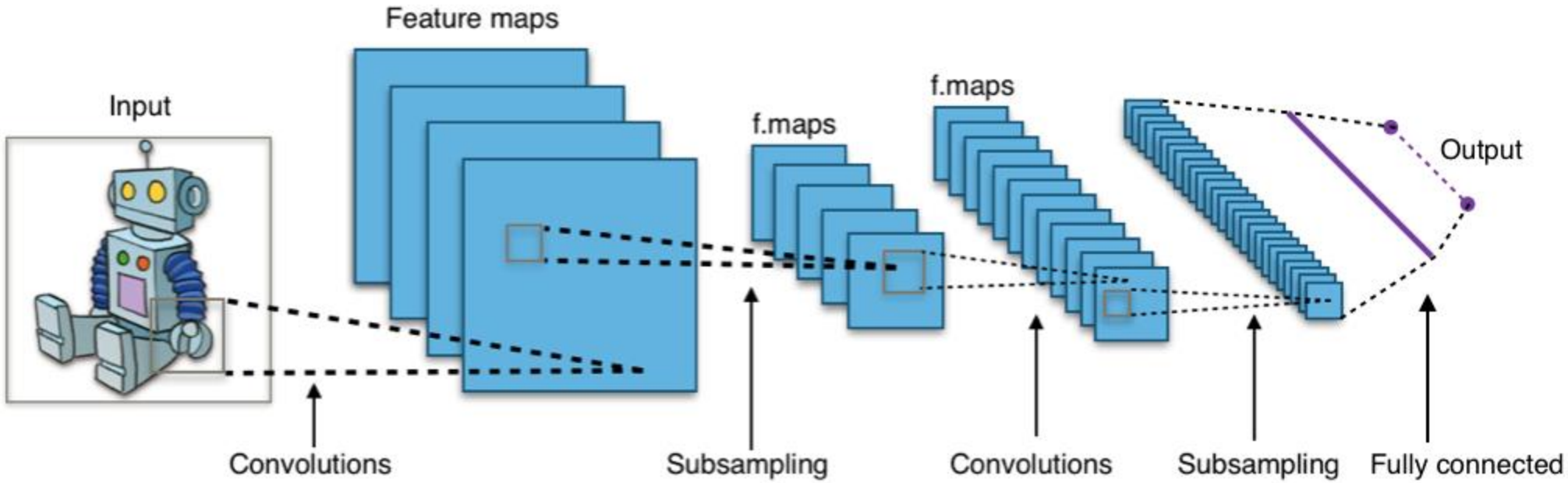


- E.g., magnitude of Fourier transform for translation invariance

$$\hat{f} \rightarrow \hat{f} e^{i2\pi t}$$

- Earlier approach was to use such hand-crafted features
- Very difficult and time consuming
- Can we minimize our effort in doing this?

Convolutional neural networks (CNN)



- Convolutional filters w; recall that we are optimizing over w
- Filters can be thought of as templates; high inner product ⇒ good match
- Lower level filters thought of as edge detectors, higher level filters thought of as object detectors

Current status

- CNNs and derivatives are best performing models for image data
- Require larger amounts of data compared to hand-crafted feature based methods
- Mostly empirical at the moment
- Both theory and applied work are major areas of research
- Several deep learning packages available in open source. Possible for new entrants to quickly learn these techniques and try on their problems.

Other settings and approaches

- Target values could be real valued – regression; easy extension
- Several other methods for classification studied in the literature. E.g., nearest neighbors, decision trees etc.
- Meta-techniques to improve performance: ensembles, bagging, distillation, boosting etc.

Phase transitions

nature physics

Letter | Published: 13 February 2017

Machine learning phases of matter

Juan Carrasquilla & Roger G. Melko

Nature Physics **13**, 431–434 (2017) | Download Citation

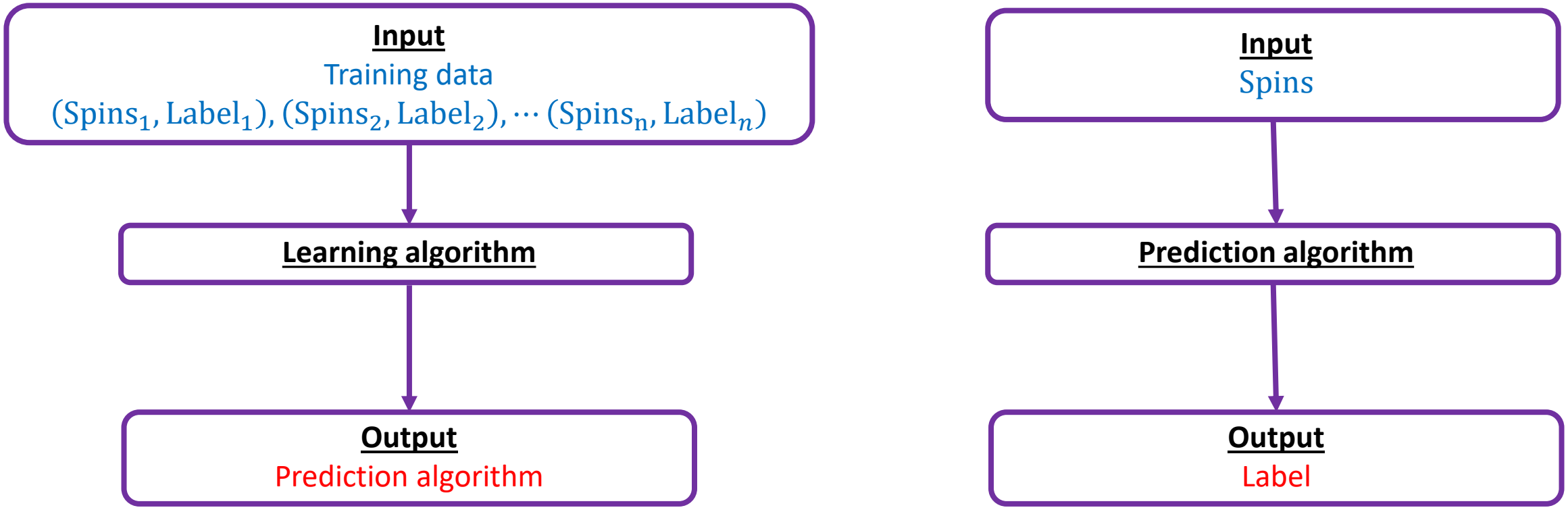
Abstract

Condensed-matter physics is the study of the collective behaviour of infinitely complex assemblies of electrons, nuclei, magnetic moments, atoms or qubits[1]. This complexity is reflected in the size of the state space, which grows exponentially with the number of particles, reminiscent of the 'curse of dimensionality' commonly encountered in machine learning[2]. Despite this curse, the machine learning community has developed techniques with remarkable abilities to recognize, classify, and characterize complex sets of data. Here, we show that modern machine learning architectures, such as fully connected and convolutional neural networks[3], can identify phases and phase transitions in a variety of condensed-matter Hamiltonians. Readily programmable through modern software libraries[4,5], neural networks can be trained to detect multiple types of order parameter, as well as

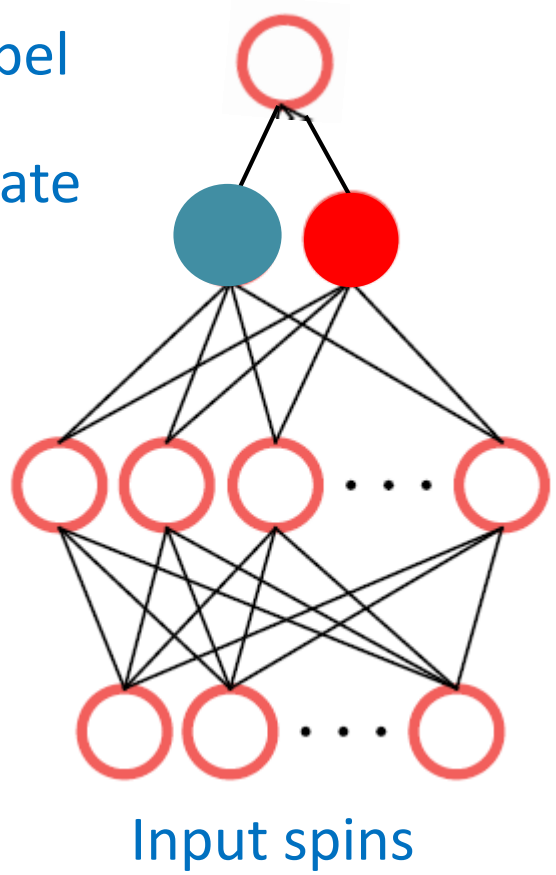
- Study of phases and their transitions important in Physics
- Physical quantities (order parameters) that illustrate phase transitions are motivated by Physics intuition/experiments
- Can we distinguish phases and learn order parameters from data?
- Ferromagnetic 2d-Ising model; $H = -\sum_{<i,j>} \sigma_i \sigma_j$
- Training data: (spin configurations, above/below critical temp)

$$\left(\{\pm 1\}^{\sqrt{n}\times\sqrt{n}}, \qquad \pm 1\right)$$

Phase transitions



Looking inside the prediction algorithm



Prediction algorithm
Fully connected neural network

Penultimate layer captures net magnetization
Order parameter

Predicting properties of molecules

Neural Message Passing for Quantum Chemistry

Justin Gilmer [1] Samuel S. Schoenholz [1] Patrick F. Riley [2] Oriol Vinyals [3] George E. Dahl [1]

Abstract

Supervised learning on molecules has incredible potential to be useful in chemistry, drug discovery, and materials science. Luckily, several promising and closely related neural network models invariant to molecular symmetries have already been described in the literature. These models learn a message passing algorithm and aggregation procedure to compute a function of their entire input graph. At this point, the next step is to find a particularly effective variant of this general approach and apply it to chemical prediction benchmarks until we either solve them or reach the limits of the approach. In this paper, we reformulate existing models into a single common framework we call Message Pass-

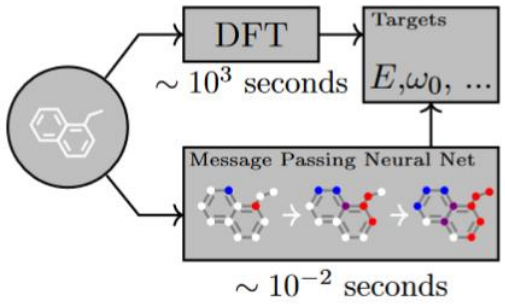


Figure 1. A Message Passing Neural Network predicts quantum properties of an organic molecule by modeling a computationally expensive DFT calculation.

Input

Molecule i.e., atoms, bonds, perhaps geometry of the molecule

Prediction algorithm

Output

Various properties such as
Atomization energies
Electron energy gap
Fundamental vibrations

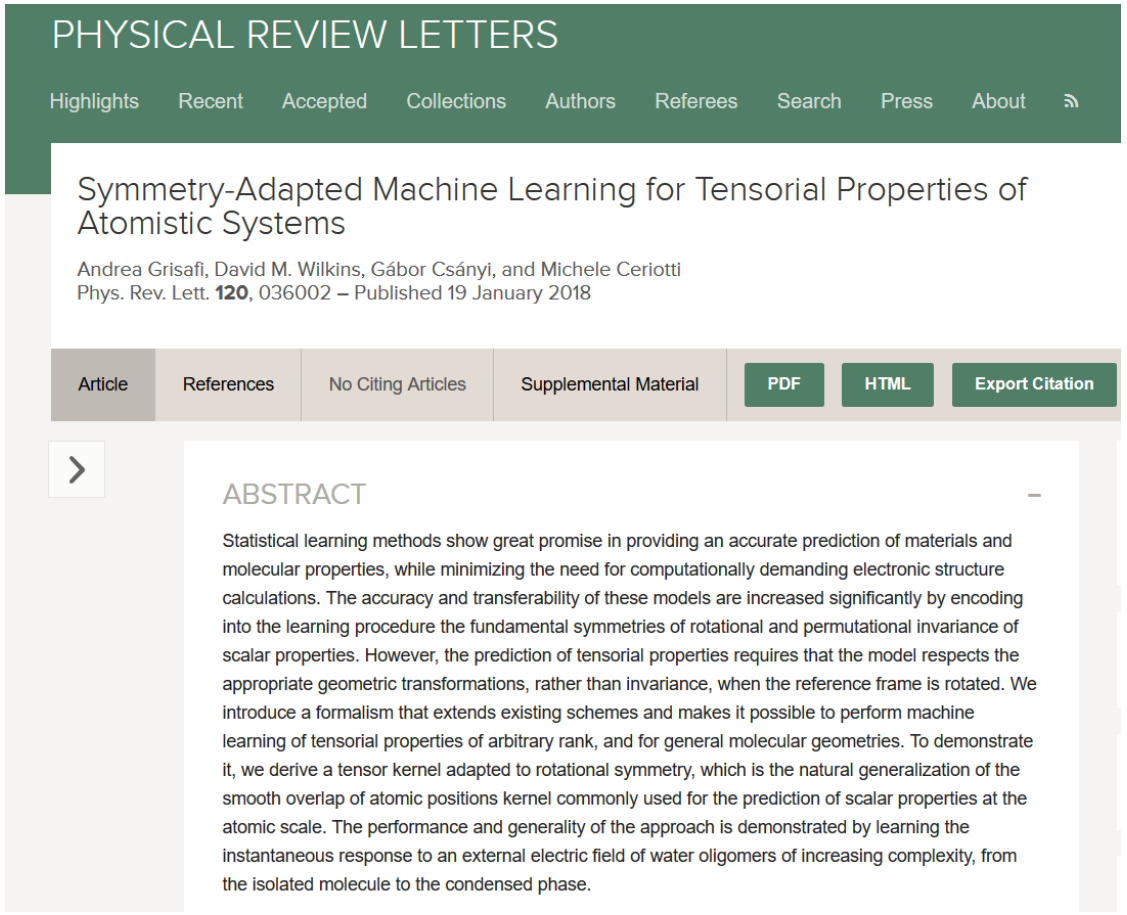
Why prediction algorithm?

- Gold standard: physical measurement – very expensive/time consuming
- Next best: simulation using density functional theory (DFT) – time consuming
- Proposed approach: much faster and inexpensive compared to both

Evaluation

- QM9 dataset – details for 134,000 molecules
- Target values computed using DFT
- Train a graph neural network for predicting the target values
- Predictions are very accurate (up to chemical accuracy)

Predicting *tensorial* properties



PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees Search Press About

Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems

Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti
Phys. Rev. Lett. **120**, 036002 – Published 19 January 2018

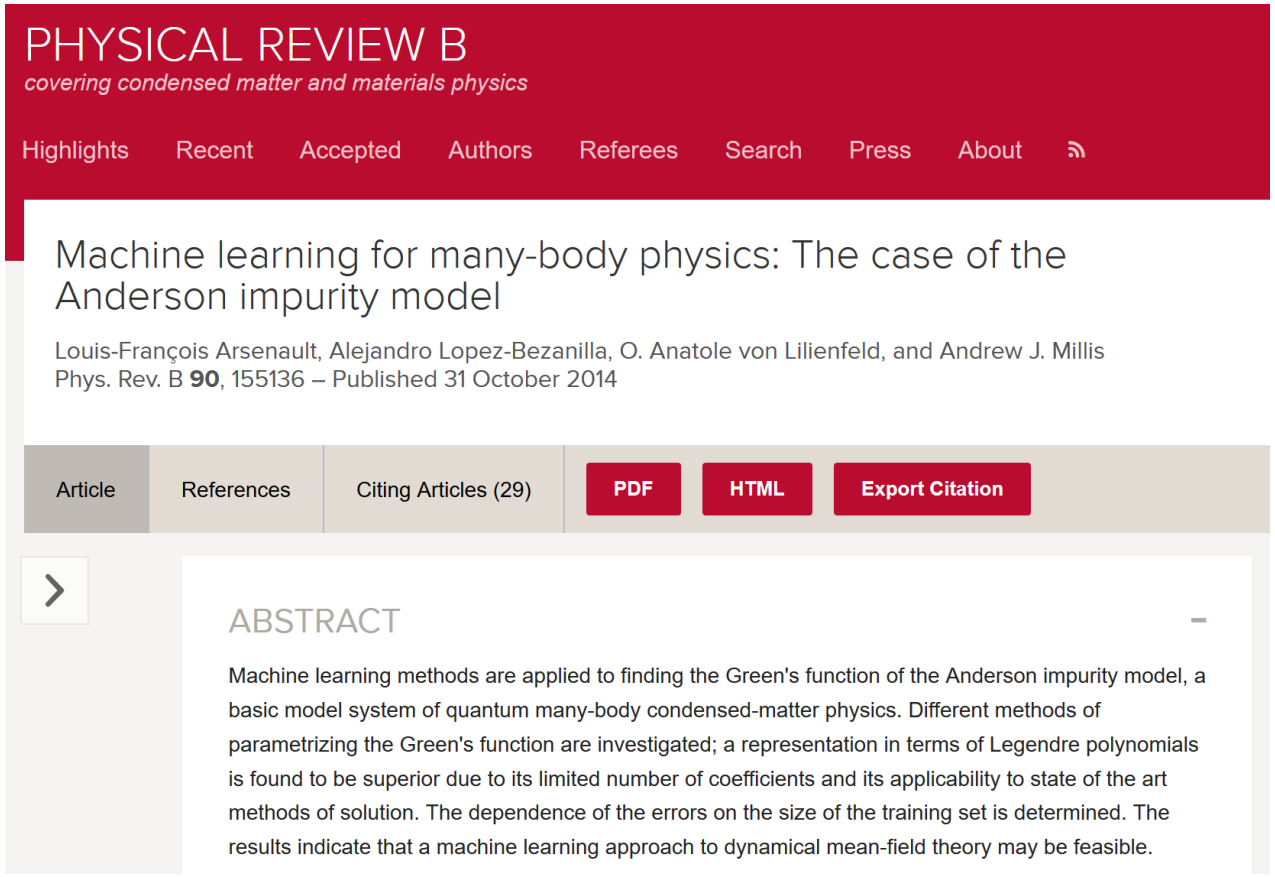
Article References No Citing Articles Supplemental Material PDF HTML Export Citation

ABSTRACT

Statistical learning methods show great promise in providing an accurate prediction of materials and molecular properties, while minimizing the need for computationally demanding electronic structure calculations. The accuracy and transferability of these models are increased significantly by encoding into the learning procedure the fundamental symmetries of rotational and permutational invariance of scalar properties. However, the prediction of tensorial properties requires that the model respects the appropriate geometric transformations, rather than invariance, when the reference frame is rotated. We introduce a formalism that extends existing schemes and makes it possible to perform machine learning of tensorial properties of arbitrary rank, and for general molecular geometries. To demonstrate it, we derive a tensor kernel adapted to rotational symmetry, which is the natural generalization of the smooth overlap of atomic positions kernel commonly used for the prediction of scalar properties at the atomic scale. The performance and generality of the approach is demonstrated by learning the instantaneous response to an external electric field of water oligomers of increasing complexity, from the isolated molecule to the condensed phase.

- Dipole moment, polarizability etc.
- Reformulate the problem to take advantage of rotational symmetries in tensors
- Improved results over naïve approach

Anderson impurity model



PHYSICAL REVIEW B
covering condensed matter and materials physics

Highlights Recent Accepted Authors Referees Search Press About

Machine learning for many-body physics: The case of the Anderson impurity model

Louis-François Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, and Andrew J. Millis
Phys. Rev. B **90**, 155136 – Published 31 October 2014

Article References Citing Articles (29) PDF HTML Export Citation

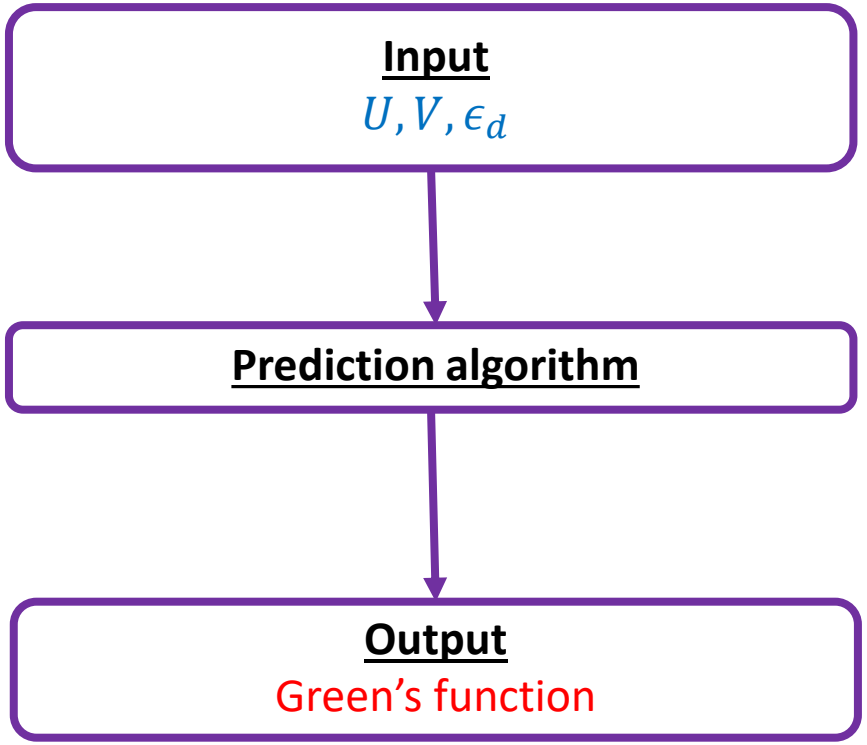
ABSTRACT

Machine learning methods are applied to finding the Green's function of the Anderson impurity model, a basic model system of quantum many-body condensed-matter physics. Different methods of parametrizing the Green's function are investigated; a representation in terms of Legendre polynomials is found to be superior due to its limited number of coefficients and its applicability to state of the art methods of solution. The dependence of the errors on the size of the training set is determined. The results indicate that a machine learning approach to dynamical mean-field theory may be feasible.

- Quantum many body problem – properties of large number of strongly correlated particles
- Exact solutions impractical to compute
- Dynamical mean field theory (DMFT) computes approximate solutions
- DMFT subproblem – Anderson impurity model

Predicting *functions*

- Penalization for double occupancy – U
- Hybridization between bath and impurity – V
- Onsite energy of impurity – ϵ_d
- Want to predict Green's function



- Training data generated for various values of U, V, ϵ_d using *exact diagonalization*
- Green's function learned in a particular basis
- Legendre polynomials used as a basis

Open direction

Green's function at different points correlated but not enforced by the algorithm

(scope for improvement)

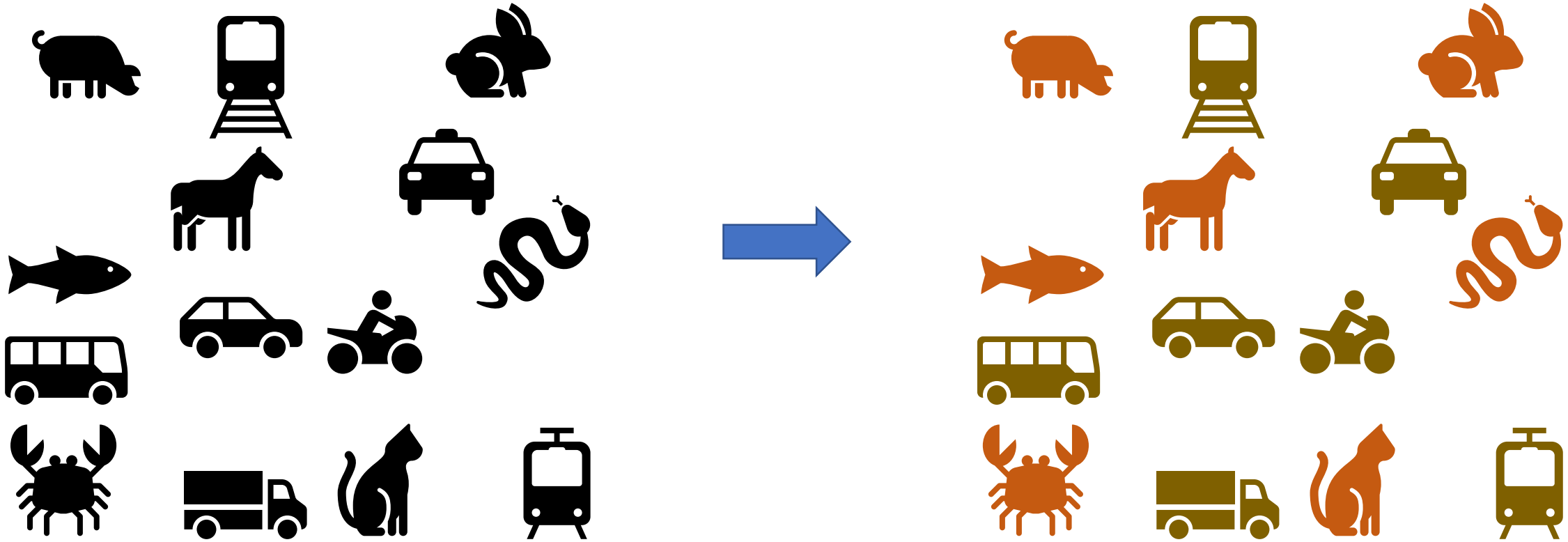
II. Unsupervised learning

Learn from unlabeled data

Given: Training data points, we would like to do

Clustering	**Group** data points into separate clusters
Density estimation	Given a new data point, **estimate** its **likelihood**
Generative models	**Generate new data** points from the same distribution

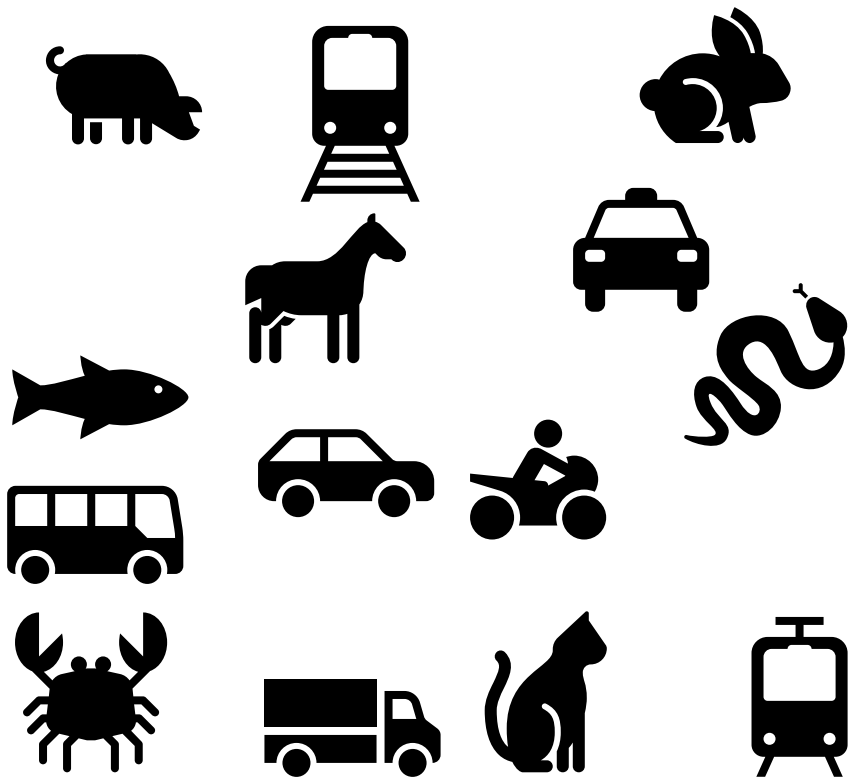
Clustering

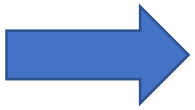


Algorithms/models for clustering

- Mixture of simple distributions e.g., mixture of Gaussians
  - K-means, Expectation-maximization (EM) etc.
- Subspace/manifold clustering
  - Principal components analysis (PCA)
- Graph clustering
- Hierarchical clustering

Density estimation





$$\mathbb{P}[\ \ ] = ?$$

$$\mathbb{P}[\ \ ] = ?$$

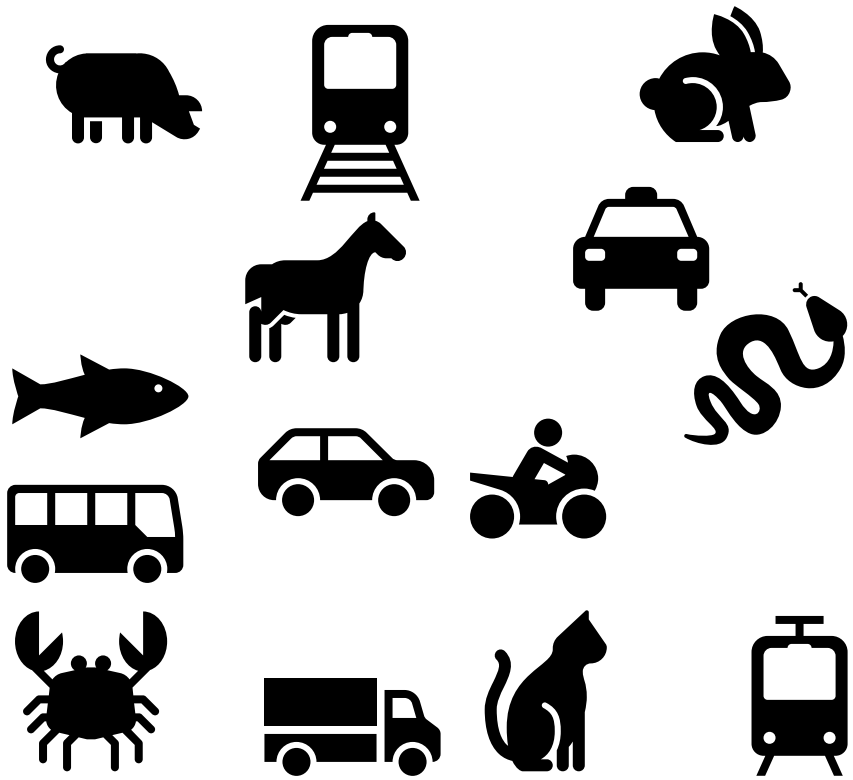
$$\mathbb{P}[\ \ ] = ?$$

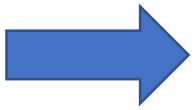
$$\mathbb{P}[\ \ ] = ?$$

Algorithms/models for density estimation

- Fit parametric/nonparametric distributions e.g., Gaussians etc.
- Use neural networks/other functions to approximate the density
- Subspace/manifold learning

Generative models







Algorithms/models for generating examples

- Variational autoencoders (VAE)
- Generative adversarial networks (GAN)
- Many more…



Picture from Radford et al. 2015

Challenges in unsupervised learning

- Much harder problem than supervised learning
- No ground truth ⇒ difficult to evaluate performance
- Much wider applicability than supervised learning
  - Lot more data available

Phase transitions – unsupervised learning



PHYSICAL REVIEW B
covering condensed matter and materials physics

Highlights Recent Accepted Authors Referees Search Press About

Discovering phase transitions with unsupervised learning

Lei Wang
Phys. Rev. B **94**, 195105 – Published 2 November 2016

Article References Citing Articles (28) PDF HTML Export Citation

ABSTRACT

Unsupervised learning is a discipline of machine learning which aims at discovering patterns in large data sets or classifying the data into several categories without being trained explicitly. We show that unsupervised learning techniques can be readily used to identify phases and phases transitions of many-body systems. Starting with raw spin configurations of a prototypical Ising model, we use principal component analysis to extract relevant low-dimensional representations of the original data and use clustering analysis to identify distinct phases in the feature space. This approach successfully finds physical concepts such as the order parameter and structure factor to be indicators of a phase transition. We discuss the future prospects of discovering more complex phases and phase transitions using unsupervised learning techniques.

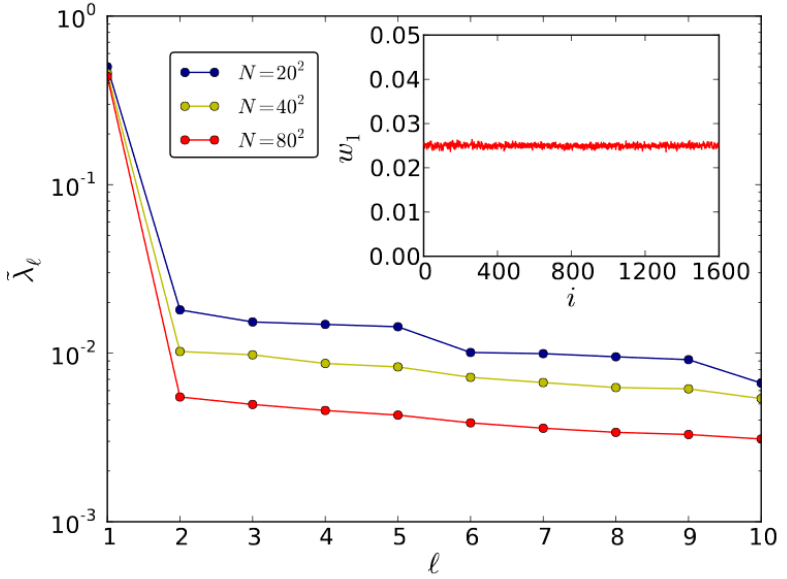
- Can we distinguish phases and learn order parameters from unlabeled data?
- Training data: Spin configurations at various temp.

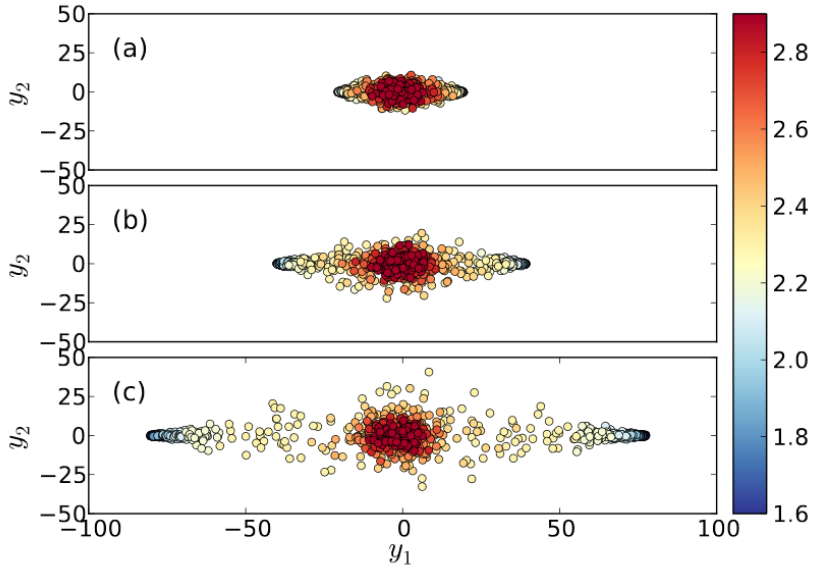
$$[\mathrm{Spins}_1|\mathrm{Spins}_2|\cdots|\mathrm{Spins}_n]$$

- Much more realistic setup
- Applicable to situations where there is less intuition about phases and order parameters

Technique: PCA + Clustering

- Compute the top- k singular directions of the matrix $[\mathrm{Spins}_1|\mathrm{Spins}_2|\cdots|\mathrm{Spins}_n]$. Denote them $w_1, \cdots, w_k$.
- Project each Spins_i onto $\mathrm{Span}(w_1, \cdots, w_k)$. Denote the projection $(y_1^i, \cdots, y_k^i)$.
- Cluster $\{(y_1^i, \cdots, y_k^i)|i = 1, \cdots, n\}$
- For a new test point, find out which cluster it belongs to





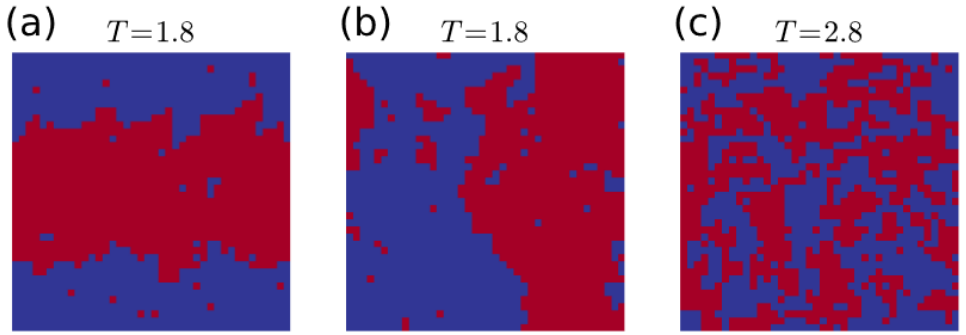
2d-Ising model

COP Ising model

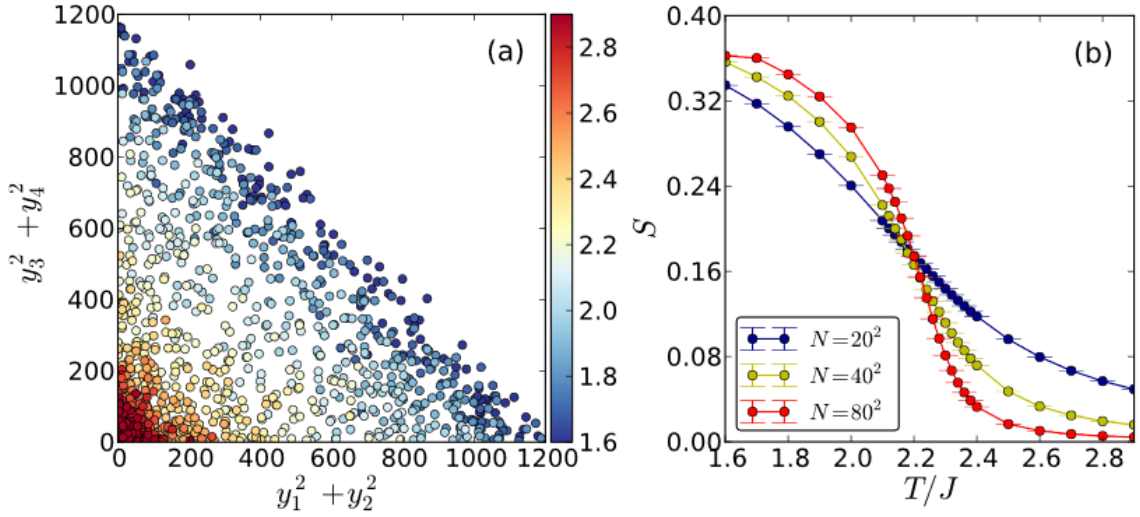
- Conserved order parameter (COP)

  $H = -\sum_{<i,j>} \sigma_i \sigma_j$ subject to $\sum_i \sigma_i = 0$

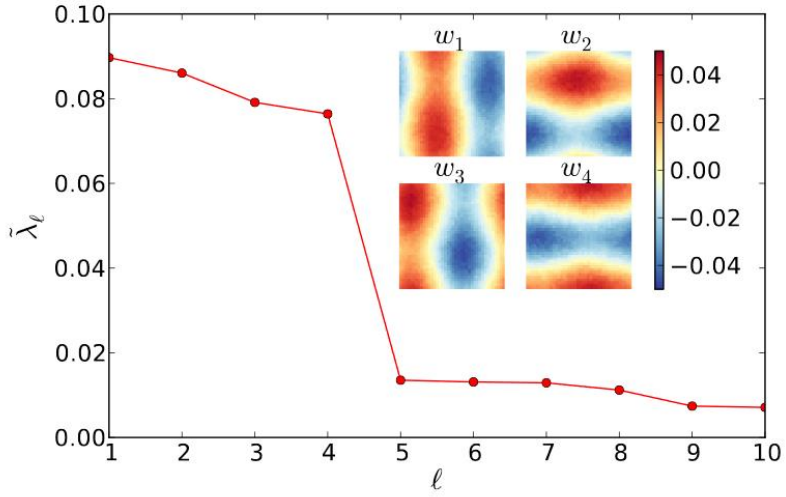
- Describes classical lattice gasses



Typical configurations below (a) & (b); and above (c) critical temperature



Phase transition and order parameter

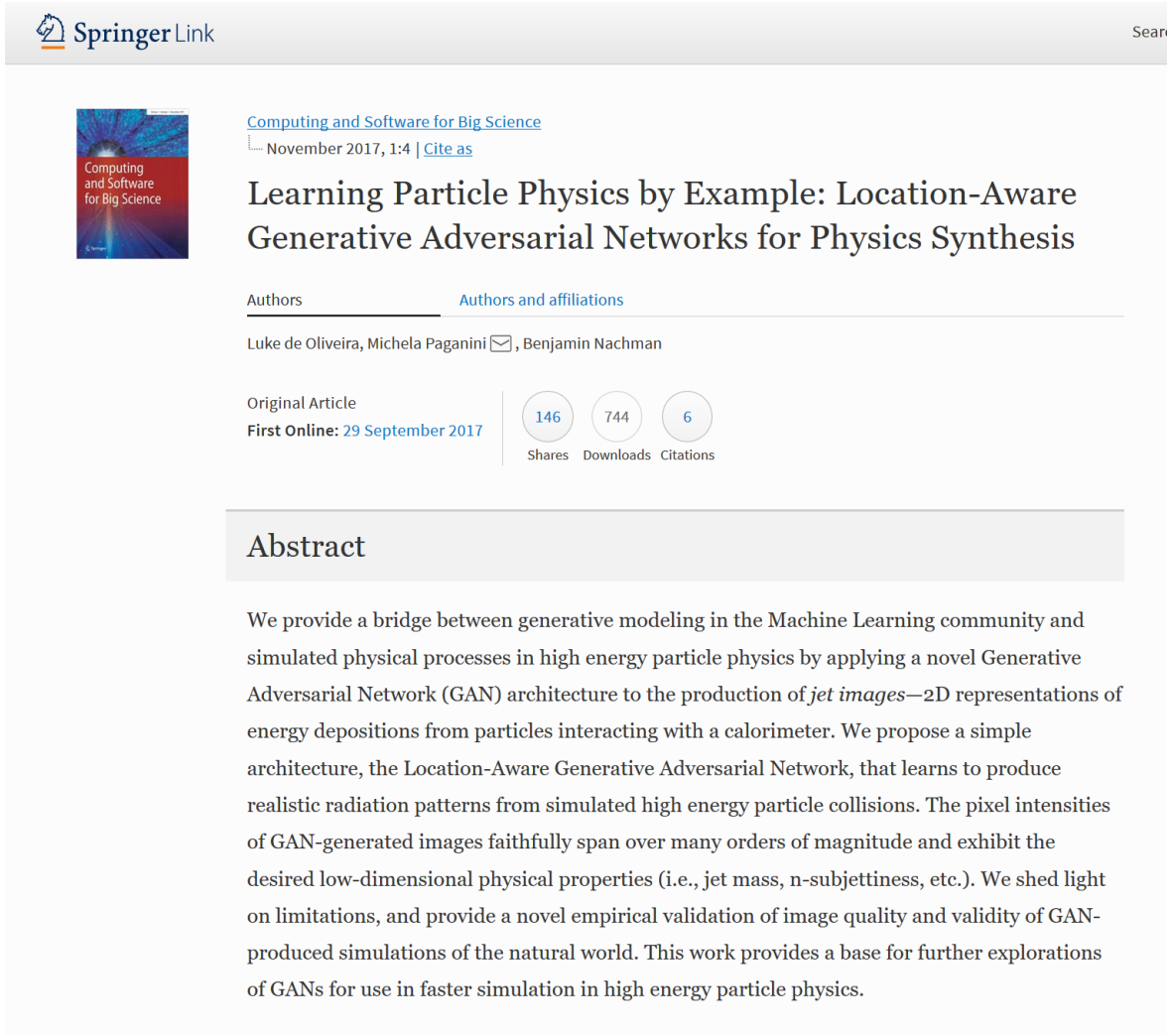


Singular value spectrum and singular vectors

Flurry of recent work…

1. Sebastian Johann Wetzel, “Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders,” Physical Review E 96, 022140 (2017), arXiv:1703.02435.
2. Kelvin Ch’ng, Nick Vazquez, and Ehsan Khatami, “Un-supervised machine learning account of magnetic transitions in the Hubbard model,” (2017), arXiv:1708.0335.
3. Evert P.L. van Nieuwenburg, Ye-Hua Liu, and Sebastian D. Huber, “Learning phase transitions by confusion,” Nature Physics , 4037 (2017), arXiv:1610.0204.
4. Ce Wang and Hui Zhai, “Machine learning of frustrated classical spin models. i. principal component analysis,” Phys. Rev. B 96, 144432 (2017).
5. Wenjian Hu, Rajiv R. P. Singh and Richard T. Scalettar, “Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination,” Phys. Rev. E 95, 062122 (2017).
6. Patrick Huembeli, Alexandre Dauphin, and Peter Wittek, “Identifying quantum phase transitions with adversarial neural networks,” Phys. Rev. B 97, 134109 (2018).

Generating jet images



Springer Link

Sear

Computing and Software for Big Science
November 2017, 1:4 | Cite as

Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis

Authors | Authors and affiliations

Luke de Oliveira, Michela Paganini ✉, Benjamin Nachman

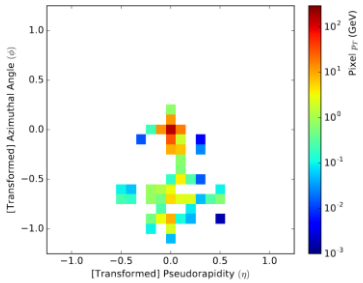
Original Article
First Online: 29 September 2017

146 Shares | 744 Downloads | 6 Citations

Abstract

We provide a bridge between generative modeling in the Machine Learning community and simulated physical processes in high energy particle physics by applying a novel Generative Adversarial Network (GAN) architecture to the production of *jet images*—2D representations of energy depositions from particles interacting with a calorimeter. We propose a simple architecture, the Location-Aware Generative Adversarial Network, that learns to produce realistic radiation patterns from simulated high energy particle collisions. The pixel intensities of GAN-generated images faithfully span over many orders of magnitude and exhibit the desired low-dimensional physical properties (i.e., jet mass, n-subjettiness, etc.). We shed light on limitations, and provide a novel empirical validation of image quality and validity of GAN-produced simulations of the natural world. This work provides a base for further explorations of GANs for use in faster simulation in high energy particle physics.

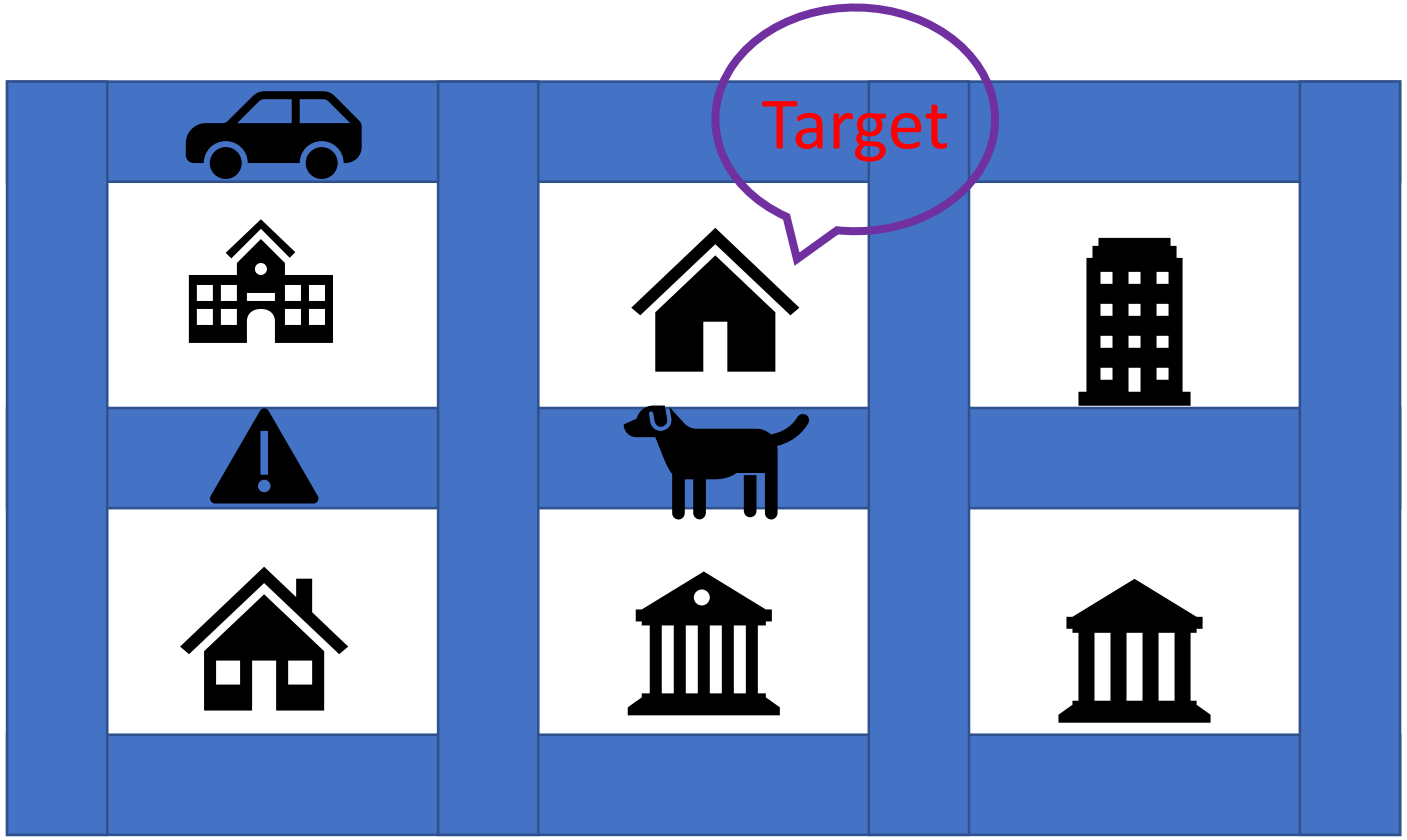
- Jets are streams of protons and hadrons formed due to scattering quarks/gluons
- Jet image – 2-dim radiation pattern



- Very sparse and NOT translation invariant
- Physicists use lots of simulated data which is very (computationally) expensive to generate
- Develops a framework to learn jet images

III. Reinforcement learning – example



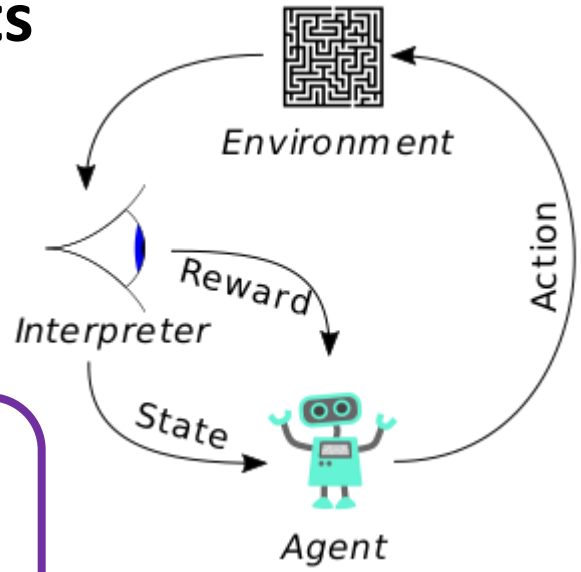


Formal setup

Exploration-exploitation trade off

Learn in dynamic & unknown environments

Components: agent & environment



Repeat

1. Agent performs an action
2. Action changes the state of the environment
3. Action also produces a reward which the agent observes

Goal: Maximize cumulative reward

Applications

- Robotics
- Planning & execution (manufacturing etc.)
- Human interaction systems e.g., conversational agents etc.
- Health and medicine e.g., design of clinical trials

Approaches

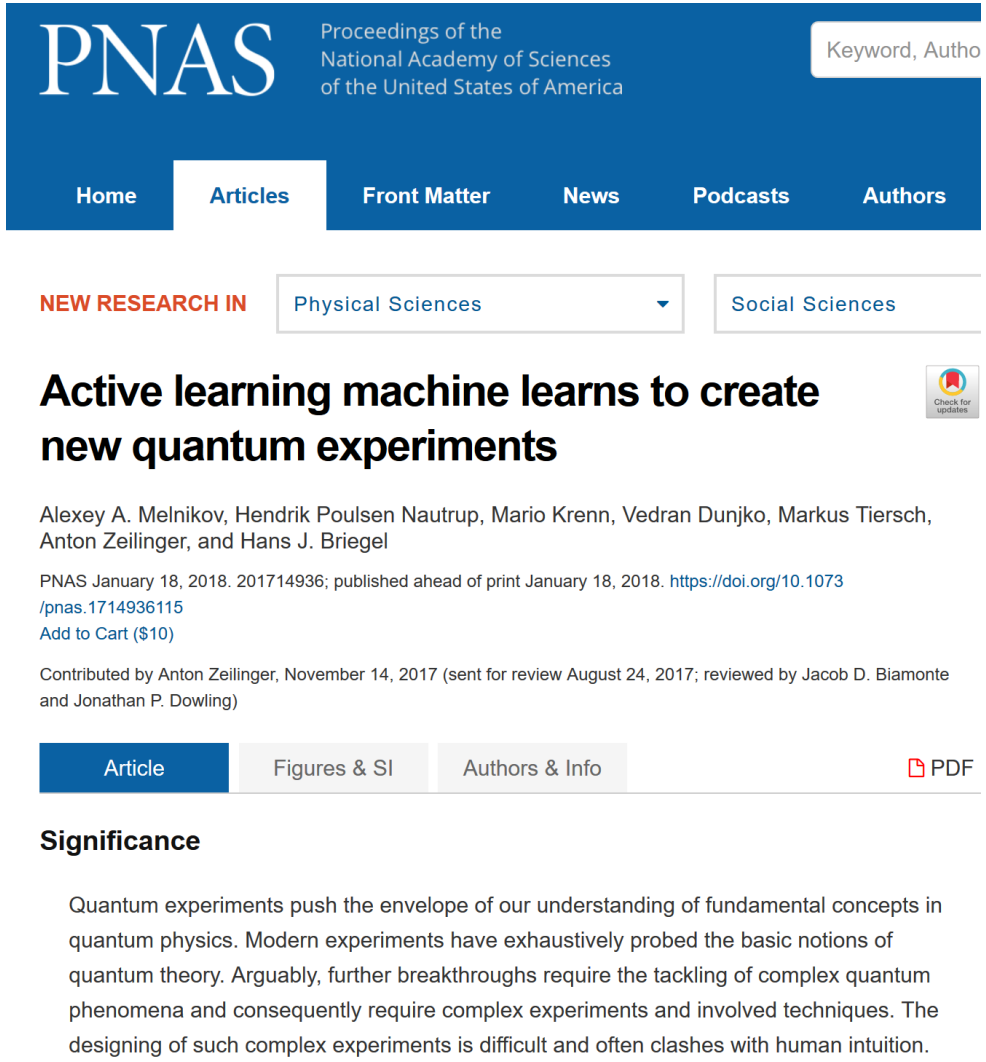
- Wide variety depending on the context
  - Do we have a model of the environment?
  - Can rewards be observed instantaneously?

- Learn the state transitions of the environment

- Learn the value function

- Learn the best policy for the agent

New quantum experiments



PNAS

Proceedings of the National Academy of Sciences of the United States of America

Keyword, Autho

Home | Articles | Front Matter | News | Podcasts | Authors

NEW RESEARCH IN Physical Sciences | Social Sciences

Active learning machine learns to create new quantum experiments

Alexey A. Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J. Briegel

PNAS January 18, 2018. 201714936; published ahead of print January 18, 2018. https://doi.org/10.1073/pnas.1714936115

Add to Cart ($10)

Contributed by Anton Zeilinger, November 14, 2017 (sent for review August 24, 2017; reviewed by Jacob D. Biamonte and Jonathan P. Dowling)

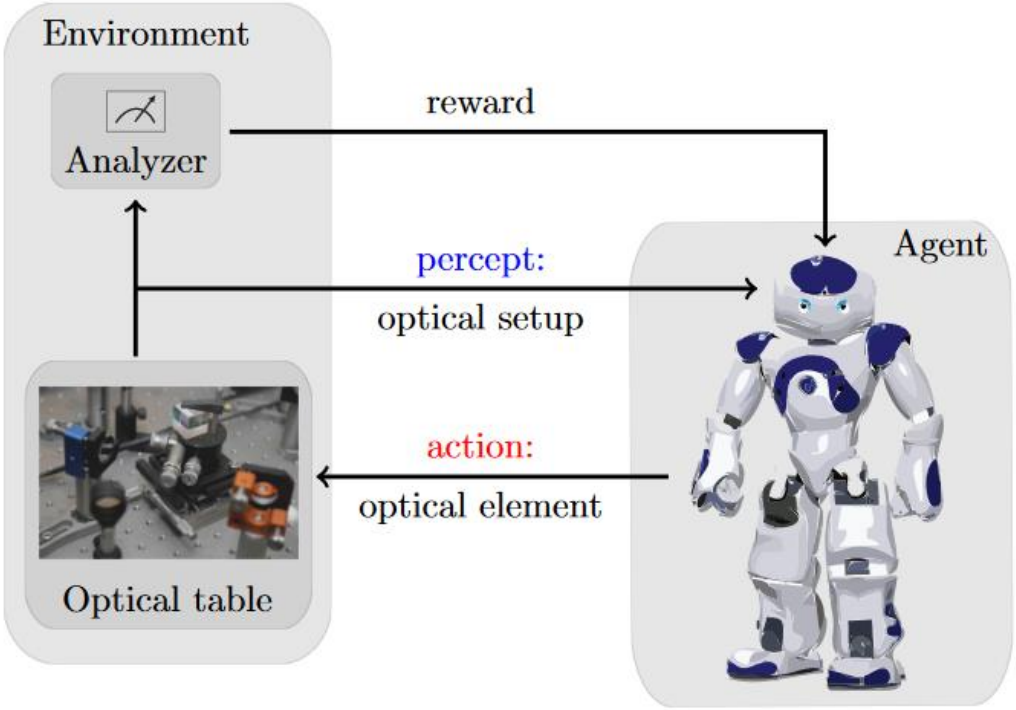
Article | Figures & SI | Authors & Info | PDF

Significance

Quantum experiments push the envelope of our understanding of fundamental concepts in quantum physics. Modern experiments have exhaustively probed the basic notions of quantum theory. Arguably, further breakthroughs require the tackling of complex quantum phenomena and consequently require complex experiments and involved techniques. The designing of such complex experiments is difficult and often clashes with human intuition.

- Given some pieces of equipment, how to design experiments that let us create interesting phenomena?
- Toolbox: Beam splitters, mirrors, holograms and Dove prisms
- Goal: experimental setup that produces high-dimensional multipartite entanglement

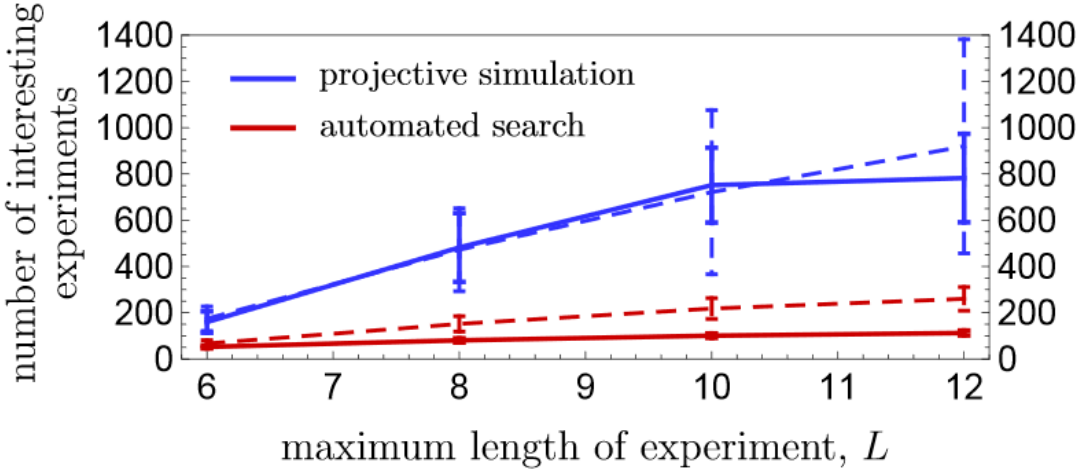
Use reinforcement learning



- Learning proceeds in episodes
- Each episode ends when either
  1. Maximum number of components used or
  2. Reward obtained
- Agent chooses actions based on previous experience
- Reward given for producing desired state
- Finds orbital angular momentum (3,3,3) states

More results

Next task: Find as many experimental setups as possible for observing desired state



Agent *independently discovered* interesting techniques such as

1. nonlocal parity sorter and
2. a setup to increase dimensionality of photons

Both are modern quantum optical gadgets that have only been recently discovered by Physicists!

Variations

- Learning from sequences (time series data)
- Semi-supervised learning
- Active learning
- Transfer learning
- Learning under constraints
  - E.g., time constraints, fairness constraints etc.
- Interpretable learning
- And many more

Conclusion

- Use of machine learning for solving Physics problems has just begun
- Shows promise but ways to go
- Requires coming together of Physicists and ML researchers
- Potential to create new directions in both fields