Restricted Boltzmann Machines and the Renormalization Group: Learning Relevant Information in Statistical Physics

Jesse Hoogland

Amsterdam University College

26-06-2019

Introduction

This talk will revolve around the intersection of:

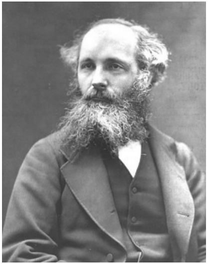
- Statistical Physics — The Renormalization Group (RG)
- Machine Learning (ML) — Restricted Boltzmann Machines (RBMs)
- Information Theory

Introduction

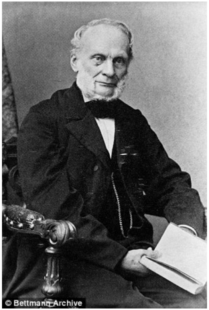
- We derive a more *exact* correspondence between RG and RBMs than in previous works.
- We provide a new implementation of an existing algorithm that learns *optimal* RG transformations, and we use this to calculate the Ising model's correlation length critical exponent.
- We describe a generalization of this algorithm to *arbitrary* lattice systems.

Introduction to Statistical Physics

(a) Ludwig Boltzmann [1]



(b) James Clerk Maxwell [2]



(c) Rudolf Clausius [3]

The Kinetic Theory of Gases

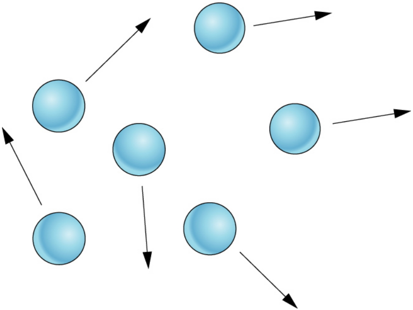


Figure 2: Statistical physics began with the kinetic theory of gases, positing gases as collections of tiny particles.

The Fundamental Responsibility of Statistical Physicists

Microscopic $\rightarrow$ Macroscopic

Ferromagnetism

Figure 3: Ferromagnetism is the process by which materials like iron form permanent magnets [4].

The Ising Model

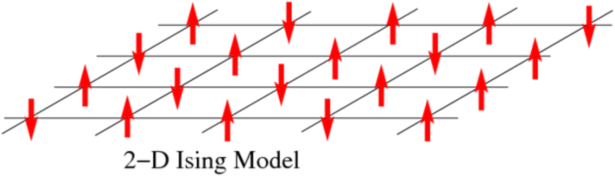
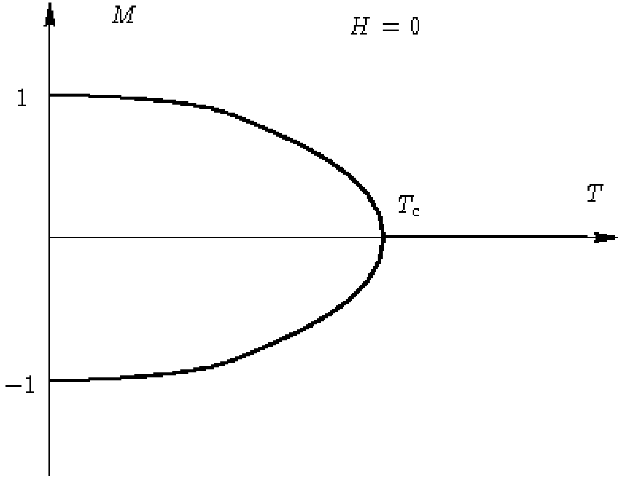


Figure 4: The Ising model [5].

Critical Phenomena and Phase Transitions



The Renormalization Group

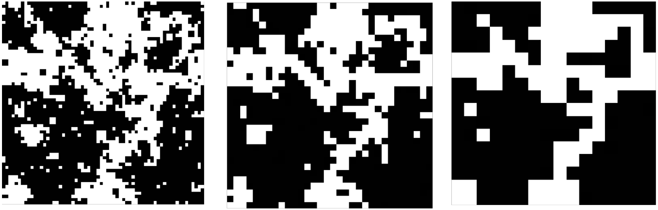


Figure 5: Three steps of majority-rule block-spin renormalization, preceding left to right (block size $b = 2$).

In RG, we reexpress our system in terms of a simpler set of variables, iterating to remove information about the shortest-distances.

Challenges in RG.

- Devising RG procedure depends on system at hand.
- No obvious way to come up with appropriate transformations.

Challenges in RG.

- Devising RG procedure depends on system at hand.
- No obvious way to come up with appropriate transformations.

Solution: Machine Learning

Neural Networks

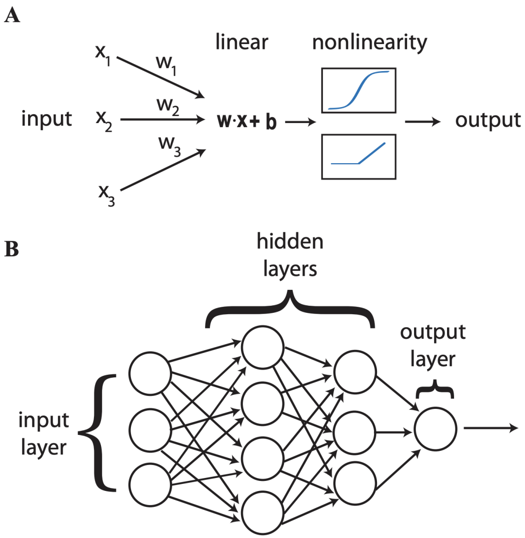


Figure 6: A neural network consists of alternating linear and non-linear transformations [7].

Qualitative similarities between RG and RBMs

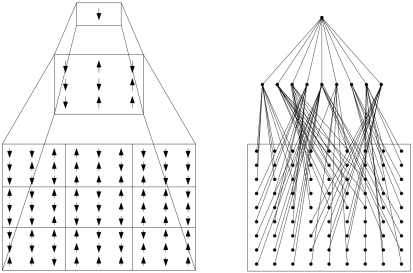


Figure 7: Two iterations of block renormalization and a deep Boltzmann machine of three layers.

Qualitative similarities between RG and RBMs

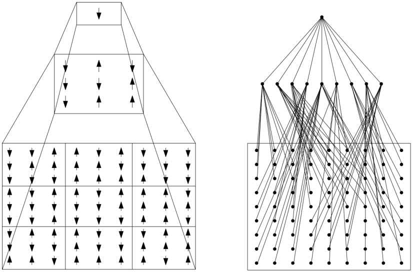


Figure 8: Two iterations of block renormalization and a deep Boltzmann machine of three layers.

Can we make this more precise?

An Exact Correspondence between Kadanoff's Variational RG and RBMs

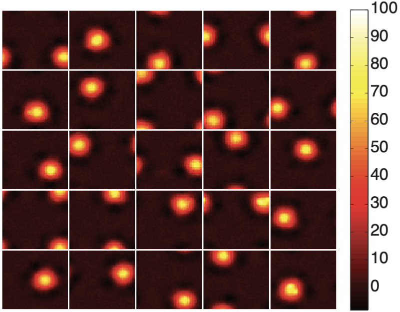


Figure 9: The receptive fields of hidden units in DBMs [10]

.

Surprisingly, this local block spin structure emerges from the training process, suggesting the [deep neural network] is self-organizing to implement block spin renormalization [10].

Do RBMs Learn Block Spin RG?

General block spin tranformations are **NOT** all appropriate RG procedures.

Extra conditions on RG Transformations

The Renormalization Group satisfies additional criteria that neural networks do not necessarily:

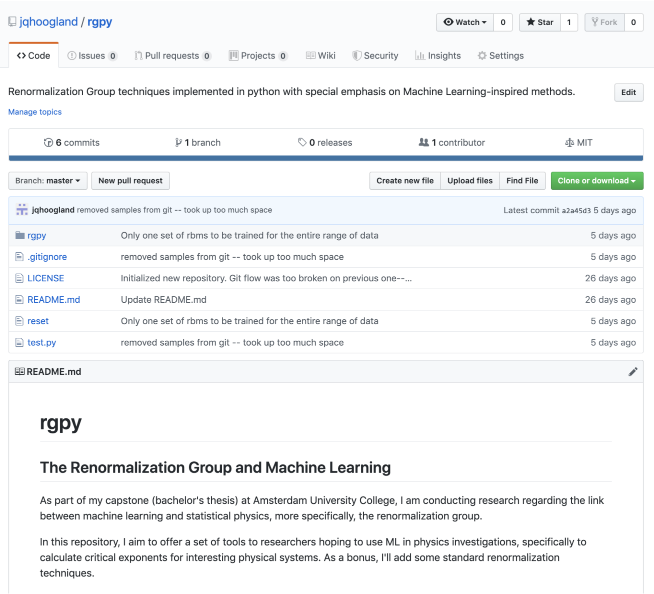
- RG transformations (should) respect the symmetry of the system under investigation.
- RG transformations (should) extract only the long-distance information.

Information Theory

Information theory gives us an exact way to calculate:

- The amount of information contained in a signal.
- The amount of *relevant* information contained in a signal.

‘rgpy’ — A python library for RG techniques.



A Recalculation of the Correlation-Length Critical Exponent

$$\boxed{\nu \approx 0.79 \pm 0.39}$$

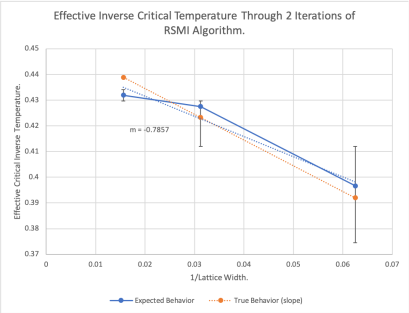


Figure 10: Finite-size collapse curve for our implementation of the RSMI algorithm.

Generalization to n-Spin and $O(n)$ Models

Models	Symmetry of order parameter	α	β	γ	δ	ν	η
$2-d$ Ising	2-component scalar	0 (log)	1/8	7/4	15	1	1/4
$3-d$ Ising	2-component scalar	0.10	0.33	1.24	4.8	0.63	0.04
$2-d$ Potts, $q=3$	q-component scalar	1/3	1/9	13/9	14	5/6	4/15
$3-d$ X-Y	$2-d$ vector	0.01	0.34	1.30	4.8	0.66	0.04
$3-d$ Heisenberg	$3-d$ vector	-0.12	0.36	1.39	4.8	0.71	0.04
Mean field		0 (dis)	1/2	1	3	1/2	0

Figure 11: Universality classes for different numbers of dimensions d and spin components n [13].

Discussion and Conclusions

- Information theory as conceptual framework for comparing ML and RG and devising *optimal* procedures: e.g. the RSMI algorithm.
- Symmetries of our systems as restricting allowed RG and ML transformations and enabling understanding of "black box" neural networks: e.g. convolutional architectures.
- (Unsupervised) Machine learning as guiding the "physical reasoning process," going beyond data analysis alone: e.g. calculating critical exponents [11].

References

[1] JT. Top 10 scientists who committed suicide, oct 2007.

[2] Who2 Biographies. James clerk maxwell biography.

[3] Sarah Griffiths. From child prodigies to playwrights, the world's 40 smartest people of all time revealed, Aug 2016.

[4] Eurico Zimbres FGEL/UERJ. Magnetez, 2005.

[5] JGTechSol. Optical computing, Feb 2017.

[6] John Cardy. *Scaling and Renormaliztion in Statistical Physics*. Cambridge University Press, Cambridge, United Kingdom, 1996.

[7] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G. R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *arXiv e-prints*, page arXiv:1803.08823, Mar 2018.

[8] Marc-Alexandre Côté and Hugo Larochelle. An infinite restricted boltzmann machine. *Neural Computation*, 28:1265–1288, 2016.

[9] Theano Development Team. Restricted boltzmann machines, 2013.

[10] Pankaj Mehta and David J. Schwab. An exact mapping between the variational renormalization group and deep learning. *arXiv e-prints*, page arXiv:1410.3831, Oct 2014.

[11] Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14:578–582, Jun 2018.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[13] Sitangshu B. Santra. Advanced statistical mechanics - models and universality, November 2013.