

Renormalization and Boltzmann Machines

Lucas Burns

Dec 20, 2021

Abstract

We argue that the previously proposed correspondence between restricted Boltzmann machines (RBMs) and variational renormalization group (RG) methods [1] is almost but not quite exact.

The Boltzmann distribution arises in various fields of physics, statistics, and machine learning. The distribution is given by the exponential of energies

$$P(v) = e^{-H(v)} / \sum_{v \in \mathcal{V}} e^{-H(v)} \quad (1)$$

where $H : \mathcal{V} \rightarrow \mathbb{R}$ is a Hamiltonian over some configuration space \mathcal{V} .

We define an energy function $E_\lambda : \mathcal{V} \times \mathcal{H} \rightarrow \mathbb{R}$ with a vector parameters λ , a Hamiltonian

$$H_\lambda(v) = -\log \sum_{h \in \mathcal{H}} E_\lambda(v, h), \quad (2)$$

and an associated Boltzmann distribution

$$P_\lambda(v) = e^{-H_\lambda(v)} / \sum_{v \in \mathcal{V}} e^{-H_\lambda(v)}. \quad (3)$$

Additionally, we have the free energy

$$F = -\log \sum_{v \in \mathcal{V}} e^{-H(v)} \quad (4)$$

with F_λ defined analogously for H_λ .

[1] shows an almost exact correspondence between these quantities used in restricted Boltzmann machines in machine learning and in variational renormalization methods applied to physics problems, such as the Ising model [2, 3]. However, [1] does not discuss the relationship between the associated optimization problems in much depth.

Here we show that there is an important difference between the two variational problems in machine learning and physics, summarized in the table below, that complexifies the proposed correspondence.

	Restricted Boltzmann Machine (RBM)	Variational Renormalization (RG)
Optimization problem	$\min_\lambda \sum_{v \in \mathcal{V}} P(v) \log \left(\frac{P(v)}{P_\lambda(v)} \right)$	$\min_\lambda F - F_\lambda $
First order condition	$\langle \nabla_\lambda H_\lambda \rangle_{P_\lambda} = \langle \nabla_\lambda H_\lambda \rangle_P$	$\langle \nabla_\lambda H_\lambda \rangle_{P_\lambda} = 0$

where $\langle H \rangle_P = \sum_{v \in \mathcal{V}} H(v)P(v)$ is the expectation value of H with respect to the probability distribution P , and ∇_λ is the gradient with respect to a vector λ of parameters (weights).

The difference in first order extremization conditions for the two optimization problems tells us that these are inequivalent variational methods. Specifically, minimizing the difference in free energy is a special case of minimizing the Kullback-Leibler divergence, when $\langle \nabla_\lambda H_\lambda \rangle_P$ is constrained to 0.

That is, the correspondence isn't quite exact since the λ_{RBM} which minimizes the Kullback-Leibler divergence $D_{\text{KL}}(P||P_\lambda)$ is generally different from λ_{RG} which minimizes the free energy difference $|F - F_\lambda|$.

In particular, Boltzmann optimization allows for the loss of arbitrary amounts of free energy between the true and coarse grained systems. Since minimization of free energy plays an important role in renormalization, it is unclear whether the conclusions in [1] that deep neural networks implement renormalization still holds in general. That said, it's possible Boltzmann machines may simply allow for extractions of a subset of coarse grained features, allowing the loss of free energy that does not contribute significantly to probabilities for some reason. Further analysis is required.

Despite this, the difference between the optimization problems suggests a novel training method for restricted Boltzmann machine that minimizes the energy difference, $|F - F_\lambda|$ rather than the Kullback-Leibler divergence / maximum likelihood. We will leave this to future work.

Derivations

At a local or global minimum λ^* for a function f , we must have $|\nabla_\lambda f(\lambda^*)| = 0$. There are additional second order conditions that determine whether it is a minimum, and whether it is a local or global minimum, but since we simply aim to show that RBMs and RG methods are distinct optimization problems, it suffices to show that the first order extremization conditions differ.

We will use the following computations.

$$\nabla_\lambda F_\lambda = -\nabla_\lambda \log \sum_{v \in \mathcal{V}} e^{-H_\lambda(v)} \quad (5)$$

$$= \sum_{v \in \mathcal{V}} \nabla_\lambda H_\lambda(v) P_\lambda(v) \quad (6)$$

$$= \langle \nabla_\lambda H_\lambda \rangle_{P_\lambda} \quad (7)$$

$$\nabla_\lambda P_\lambda(v) = \nabla_\lambda e^{-H_\lambda(v)} / e^{-F_\lambda} + \nabla_\lambda e^{F_\lambda} e^{-H_\lambda(v)} \quad (8)$$

$$= -(\nabla_\lambda H_\lambda(v) - \nabla_\lambda F_\lambda) P_\lambda(v) \quad (9)$$

using $P_\lambda(v) = e^{-H_\lambda(v)} / e^{-F_\lambda}$.

Restricted Boltzmann Machines

The first order condition for RBMs is given by

$$\nabla_\lambda D_{\text{KL}}(P||P_\lambda) = \nabla_\lambda \sum_{v \in \mathcal{V}} P(v) \left(\log(P(v)) - \log(P_\lambda(v)) \right) \quad (10)$$

$$= - \sum_{v \in \mathcal{V}} P(v) \nabla_\lambda \log(P_\lambda(v)) \quad (11)$$

$$= - \sum_{v \in \mathcal{V}} P(v) \nabla_\lambda P_\lambda / P_\lambda(v) \quad (12)$$

$$= \sum_{v \in \mathcal{V}} P(v) (\nabla_\lambda H_\lambda(v) - \nabla_\lambda F_\lambda) \quad (13)$$

$$= \langle \nabla_\lambda H_\lambda \rangle_P - \langle \nabla_\lambda H_\lambda \rangle_{P_\lambda} = 0 \quad (14)$$

Variational Renormalization

The first order extremization condition for variational RG is given by

$$\nabla_\lambda(F - F_\lambda)^2 = -2\nabla_\lambda F_\lambda(F - F_\lambda) = 0 \quad (15)$$

which for inexact transformations $F \neq F_\lambda$, requires a λ which satisfies

$$\nabla_\lambda F_\lambda = \langle \nabla_\lambda H_\lambda \rangle_{P_\lambda} = 0. \quad (16)$$

References

- [1] Pankaj Mehta and David J. Schwab. An exact mapping between the Variational Renormalization Group and Deep Learning. *arXiv:1410.3831 [cond-mat, stat]*, October 2014. arXiv: 1410.3831.
- [2] K Wilson. The renormalization group and the ε expansion. *Physics Reports*, 12(2):75–199, August 1974.
- [3] Leo P Kadanoff. *Statistical Physics: Statics, Dynamics and Renormalization*. WORLD SCIENTIFIC, May 2000.