

---

# Restricted Boltzmann Machines

---

**Chaimaa Kadaoui**  
chaimaa.kadaoui\*

**Othman SBAI**  
othman.sbai

**Xi Shen**  
xi.shenxi

## 1 Abstract

For our project in Probabilistic Graphical Models course, we studied the use and the training of Restricted Boltzmann Machines as an instance of undirected graphical models. We mainly base our approach on the practical guide for training Restricted Boltzmann Machines (Hinton 2010 [1]). We start by an introduction to RBM and the interest that they have raised in the last years, then we define the theoretical aspects of main training methods of RBM. Finally we present our results by comparing our implementation of training RBM to an implementation from another deep learning library, on two popular image datasets MNIST and CIFAR-10. We show our convergence results and investigate the influence of different parameters of the training algorithm.

## 2 Introduction

### 2.1 Definition of an RBM

A Restricted Boltzmann Machine (RBM) is an undirected energy-based probabilistic graphical model which can be seen as a two layer neural network (visible and hidden units). RBMs can be used to model and learn important aspects of a probability distribution of a training data. They are called restricted, because we impose restrictions on the network topology, by allowing only connections between units of different layers. RBM are energy-based, since they define probability distribution through an energy function. Learning corresponds to shaping the energy so that desirable configurations have a low energy and thus maximize probability of training data under the model. Maximum likelihood learning is challenging for undirected graphical models because MLE parameters cannot be found analytically and the log likelihood gradient based optimization is not tractable. This optimization requires obtaining samples through Markov Chain Monte Carlo, which is computationally demanding.

### 2.2 Interest in RBM

A major interest have been dedicated to directed graphical models with one layer of observed variables and one or more layers of hidden variables such Mixture of Gaussians, probabilistic PCA, factor analysis, latent dirichlet analysis LDA... However, the undirected two layered analogue which was first introduced and called Harmonium by [5] and renamed as "Restricted Boltzmann Machines" by [2] have also seen advances through new training methods, which unlike directed graphical models, makes inference and learning very fast.

## 3 Applications of RBM

Restricted Boltzmann machines have successfully been applied to problems involving high dimensional data such as images, text [3], [6] as unsupervised feature extractors.

Recent developments have demonstrated the capacity of RBM to be powerful generative models, able to extract useful features from input data or construct deep artificial neural networks. In this setting

---

\*mail extension @eleves.enpc.fr for all authors

they can be seen as good preprocessing or initialization for some other models. They can also be used as self-contained classifiers as shown in [4].

- **RBM as a generative model:** learns and generates samples from data distribution
- **RBM as feature extractor:** unsupervised feature extraction from data, instead of handcrafting features
- **RBM for computer vision:** such as object recognition, image denoising and inpainting.
- **RBM for collaborative filtering:** given a set of  $N$  users and  $M$  movies, we would like to recommend movies to the users.

## 4 Theory of RBM

### 4.1 Energy and probability

### 4.2 Contrastive divergence theory ...

## 5 Training RBM on MNIST Data

### 5.1 Implementation of RBM on python

### 5.2 Implementation of RBM on tensorflow using a GPU

We use the implementation in tensorflow provided by Gabriele Angeletti as a library gathering many deep learning models and algorithms such as convolutional networks, RNN, RBM, Deep Belief Networks...

## References

- [1] Geoffrey Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.
- [2] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [3] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [4] Hugo Larochelle, Michael Mandel, Razvan Pascanu, and Yoshua Bengio. Learning algorithms for the classification restricted boltzmann machine. *Journal of Machine Learning Research*, 13(Mar):643–669, 2012.
- [5] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. Technical report, DTIC Document, 1986.
- [6] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.