# Implementation of *Restricted Boltzman Machine* using *Contrastive Divergence*

**PROJECT PRESENTATION-Pattern Recognition and Machine Learnig-EE5610**

BY-**1.ANNU** (EE21RESCH01010)
**2.Shantanu Yadav**(EE20MTECH12001)
INDIAN INSTITIUTE OF TECHNOLOGY , HYDERABAD

*ee21resch01010@iith.ac.in ee20mtech12001@iith.ac.in*

July 23, 2021

# Overview

# Introduction to Restricted Boltzman Machine

## Restricted Boltzmann Machine (RBM):

1. Restricted Boltzmann machines (RBMs) are probabilistic graphical models that can be interpreted as stochastic neural networks.
2. RBMs are shallow neural nets that learn to reconstruct data by themselves in an unsupervised fashion.

# Why are RBMs important?

1. An RBM are a basic form of autoencoder.
2. It can automatically extract **meaningful** features from a given input.
3. The visible units constitute the first layer and correspond to the components of an observation (e.g., one visible unit for each pixel of a digital input image).
4. The hidden units model dependencies between the components of observations (e.g., dependencies between the pixels in the images) and can be viewed as non-linear feature detectors.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# RBM Architecture

- Boltzmann machines two types of units can be distinguished. They have visible neurons and potentially hidden neurons.
- Restricted Boltzmann machines always have both types of units, and these can be thought of as being arranged in two layers, see Fig. 1 for an illustration.

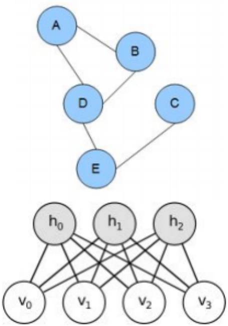



Figure 1: RBM Architecture

# RBM Learning

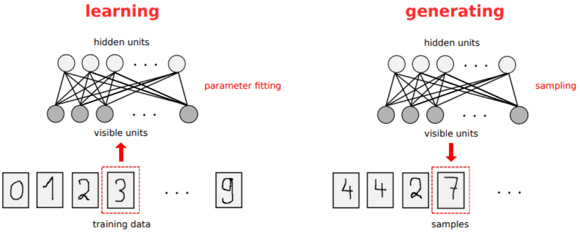


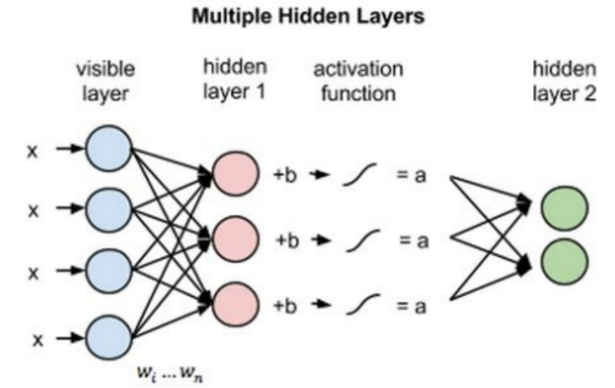

Figure 2: Left: Learning an RBM corresponds to fitting its parameters such that the distribution represented by the RBM models the distribution underlying the training data, here handwritten digits. Right: After learning, the trained RBM can be used to generate samples from the learned distribution.

# What is RBM,BM,EBM?

1. Energy Based Models
2. EBM with Hidden units
3. Boltzmann machine or a undirected graphical model, is a set of random variables having a Markov property.
4. **RBM**: restrict BMs to those without visible-visible and hidden-hidden connections.

# RBM vs Neural Networks

1. RBMs are shallow, two-layer neural nets that constitute the building blocks of deep-belief networks.
2. The 1st layer of the R.B.M is called the visible, or input, layer, and the second is the hidden layer.



# How Reconstruction is Possible in RBM?

1. How do they perform better without involving a deeper network?
2. Reconstruction estimates a continuous value based on many inputs and makes guesses about which discrete label to apply to a given input.

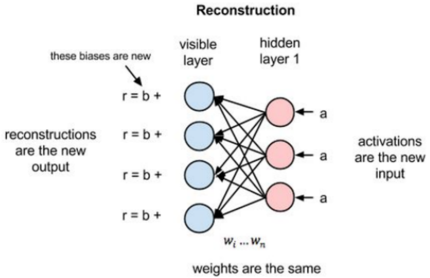



Figure 3: Reconstruction in RBM

**Undirected Graphical Models :**

An undirected graph is an ordered pair $G = (V, E)$, where $V$ is the finite set of nodes and $E$ is the set of undirected edges.

**Properties:**

- An edge between two nodes $v$ and $w$ such that $v, w \in V$ is denoted by $\{v, w\}$
- The neighbourhood of $v \in V$ is the set $N_v = \{w : \{v, w\} \in E\}$
- The path from $v_1$ to $v_m$ is a sequence of nodes $v_i \in V$ for $i = 1, 2, \ldots m$ such that $\{v_j, v_j + 1\} \in E$ for $j = 1, 2, \ldots m - 1$.

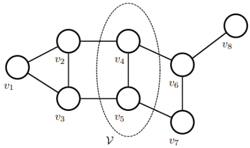



Figure 4: Undirected Graph

**Markov Random Fields**

Let each node $v \in V$ of the undirected graph $G$ be associated with a random variable $X_v$.

- The set of random variables $\mathbf{X} = (X_v)$ is called a Markov random field (MRF) if the joint probability distribution $p$ fulfills the (local) Markov property w.r.t. the graph. This property is fulfilled if for all $v \in V$ the random variable $X_v$ is conditionally independent of all other variables given its neighborhood $(X_w)_{w \in N\{v}$

# RBM as an MRF

- An RBM is an MRF associated with a bipartite undirected graph. It consists of m visible units $V = (V_1, ..., V_m)$ representing the observable data, and n hidden units $H = (H_1, ..., H_n)$ to capture the dependencies between the observed variables.
- For all $i \in 1, ..., n$ and $j \in 1, ..., m$, $w_{ij}$ is a real valued weight associated with the edge between the units $V_j$ and $H_i$ , and $b_j$ and $c_i$ are real valued bias terms associated with the jth visible and the ith hidden variable, respectively.

# The Inference Process-RBM

The Inference has 2 phases

1. Forward Pass
2. Backward Pass or Reconstruction

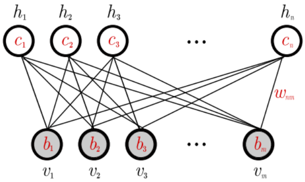



Figure 5: An RBM network

# Phase 1 -Forward Pass

1. $p(h_j|x) = \sigma(\sum_{i=1}^{V} W_{ij}x_i + h_{bias}$ ,where $\sigma(z) = \frac{1}{1+exp(-z)}$ is the logistic function.
2. We turn unit j with probability $p(h_j|v)$ , and turn it off with probability $1p(h_j|v)$ by generating a uniform random number vector $\epsilon$, and comparing it to the activation probability as
   If $\epsilon_j < p(h_j|v)$ , then $h_j = 1$ , else $h_j = 0$ .
3. The conditional probability of a configuration of **h** given **v** (for a training sample) is given by: $p(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{H} p(h_j|\mathbf{v})$ where $H$ is the number of hidden units.

# Phase 2-Backward Pass

1. RBM learns a probability distribution over the input, and then, after being trained, the RBM can generate new samples from the learned probability distribution.
2. The (conditional) probability distribution over the visible units v is given by-
   $p(v|h) = \prod_{i=1}^{V} p(v_i|h), where$
   $\mathrm{p}(\mathrm{v}_i|h) = \sigma(\sum_{j=1}^{H} W_{ji} h_j + v_{bias})$

# Is RBM a generative or Discriminative model?

## RBM

RBM is a Generative Model

- **Discriminative Model versus Generative Model**
  A Generative Model learns the joint probability distribution $p(x, y)$. It predicts the conditional probability with the help of Bayes Theorem.
- A Discriminative model learns the conditional probability distribution $p(y|x)$.

# What is the objective function in RBM?

1. **Goal:** Maximize the likelihood of our data being drawn from that distribution.
2. **Calculate error:** In each epoch, we compute the ”error” as a sum of the squared difference between step 1 and step n, e.g the error shows the difference between the data and its reconstruction.

# Training the Model

1. Gibbs Sampling
2. Contrastive Divergence

**Link to Jupyter Notebook**

# How Training happens?

- RBM is a 2 layer neural network. Simply, RBM takes the inputs and translates those into a set of binary values that represents them in the hidden layer.
- Then, these numbers can be translated back to reconstruct the inputs. Through several forward and backward passes, the RBM will be trained, and a trained RBM can reveal which features are the most important ones when detecting patterns.

# Why Contrastive Divergence?

1. Computing the likelihood of an undirected graphical model or its gradient is in general computationally intensive, and this also holds for RBMs.
2. Thus, sampling-based methods are employed to approximate the likelihood gradient.

# Contrastive Divergence Illustration

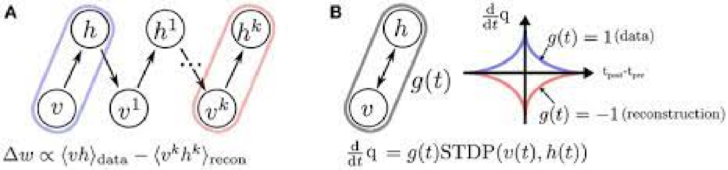



Figure 6: Contrastive Divergence

# What are the applications of an RBM?

1. **Collaborative Filtering:**
   Collaborative filtering (CF) is a technique used by recommender systems.Collaborative filtering has two senses, a narrow one and a more general one.
   In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating).
2. Dimensionality Reduction
3. Classifications
4. Regression
5. Feature learning
6. Deep Belief Network

# Restricted Boltzman Machine for Collaborative Filtering

## RBM for collaborative filterig

1. Low-rank approximations based on minimizing the sum-squared distance can be found using SVD(singular value Decomposition).
2. But in collaborative filtering domain, most of the data sets are sparse and this creates a complex non-convex problem. Thus we use RBM.

# Movie Rating Prediction Using RBM

1. If we have M movies, N users and integral ratings from 1-K. If all N users rated the same set of M movies then we just use a single training case for an RBM which has M softmax visible units symmetrically connected to a set of binary hidden units.
2. But in case of lot of ratings missing, we model a different RBM for each user.
3. Every RBM has the same number of hidden units, but an RBM only has visible softmax units for the movies rated by that user, so an RBM has few connections if that user rated few movies.

# Movie Rating Prediction Model

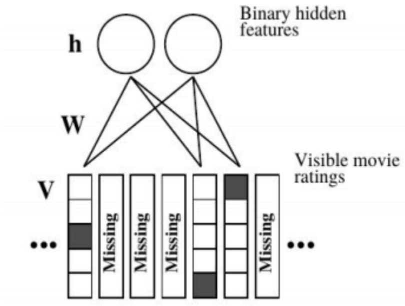



Figure 7: Movie Rating Prediction

# RBM in Deep belief Networks

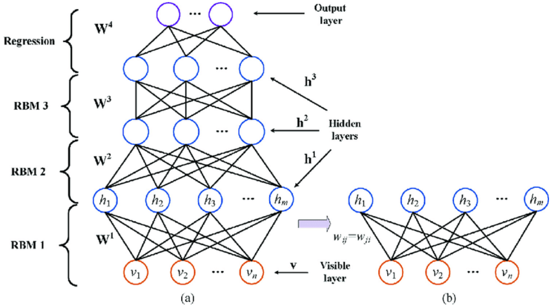



Figure 8: RBM used in DBN

# RBM Implementation on MNIST data

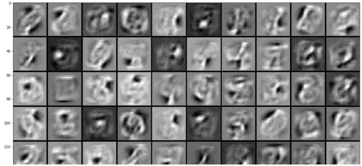



Figure 9: Learned Features On MNIST data

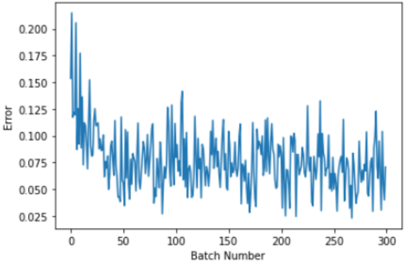



Figure 10: Error versus Epoch Curve

# Conclusion

1. RBMs learn from the data. They actually encode their own structure. This is why they're grouped into a larger family of models known as autoencoders
2. Restricted Boltzmann machines are useful in many applications, like dimensionality reduction, feature extraction, and collaborative filtering just to name a few. On top of that, RBMs are used as the main block of another type of deep neural network which is called deep belief networks

# Reference

Training Restricted Boltzmann Machines: An Introduction Asja Fischer and Christian Igel Institut f¨ur Neuroinformatik, Ruhr-Universit¨at Bochum Universit¨atsstraße 150, 44780 Bochum, Germany 2 Department of Computer Science, University of Copenhagen Universitetsparken 5, 2100 Copenhagen Ø, Denmark**Link to Jupyter Notebook**

# THANK YOU.......