

# Restricted Boltzmann Machine - RBM

Eduardo M. Rubik e Lucas Santana

**LAMFO**

**Laboratório de Aprendizado de Máquina em Finanças e Organizações**

June 20, 2020

# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

## Netflix Prize

Home Rules Leaderboard Register Update Submit Download

### Leaderboard

10.05% Display top 20 leaders.

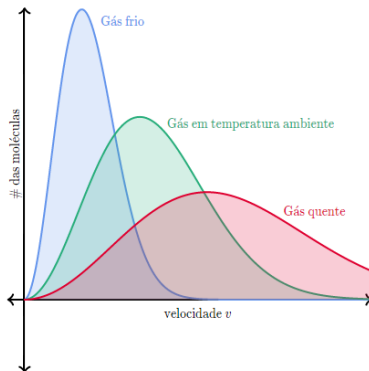
Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	<a href="#">BellKor's Pragmatic Chaos</a>	0.8558	10.05	2009-06-26 18:42:37
<b>Grand Prize - RMSE &lt;= 0.8563</b>				
2	<a href="#">PragmaticTheory</a>	0.8582	9.80	2009-06-25 22:15:51
3	<a href="#">BellKor in BigChaos</a>	0.8590	9.71	2009-05-13 08:14:09
4	<a href="#">Grand Prize Team</a>	0.8593	9.68	2009-06-12 08:20:24
5	<a href="#">Dace</a>	0.8604	9.56	2009-04-22 05:57:03
6	<a href="#">BigChaos</a>	0.8613	9.47	2009-06-23 23:06:52

# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

- Ludwig Eduard **Boltzmann** (1844-1906) foi um físico austríaco que estabeleceu os principais conceitos da física clássica estatística.
- A **Distribuição de Boltzmann** descreve a probabilidade de se encontrar moléculas de um gás em um determinado estado energético. Essa distribuição a menos de constantes é dada por (1):

$$\frac{N_i}{N} = \frac{g_i e^{\frac{E_i}{kT}}}{Z(T)} \quad (1)$$

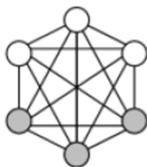


- A **Máquina de Boltzmann** leva esse nome, pois a distribuição de probabilidade utilizada é a **Distribuição de Boltzmann**.

# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann**
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

- **Máquinas de Boltzmann** são modelos probabilísticos (ou geradores) não supervisionados, baseados em energia, representados por um sistema totalmente conectado.

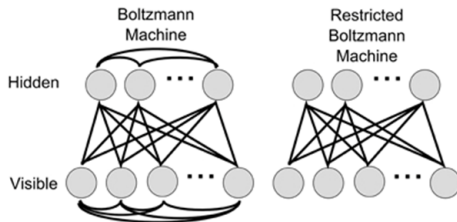


- Para cada configuração do sistema, atribui-se um valor de energia e uma probabilidade associada a essa energia.
- **Objetivo:** ajustar o modelo para que pouca energia represente alta probabilidade e muita energia baixa probabilidade.



# Máquina de Boltzmann Restrita

- **Smolensky (1986)** propôs alterações práticas na máquina de Boltzmann, a partir de restrições na sua topologia.
  - Inexistência de conexões intra-camadas.

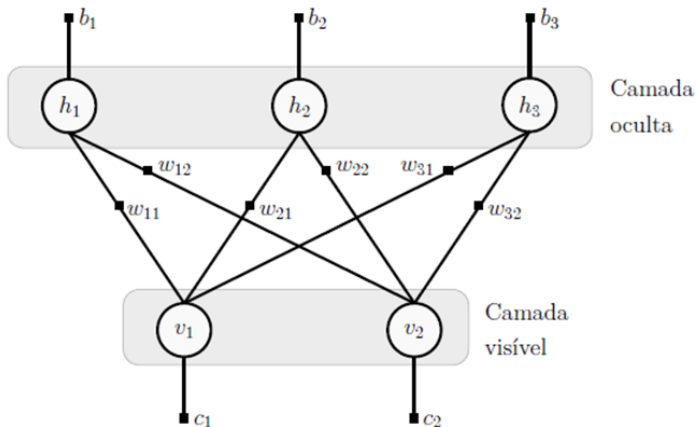


- **Hinton (2002)** propôs um modelo de aprendizado eficiente e que vem sendo utilizado amplamente na academia e na indústria.

# Sumário

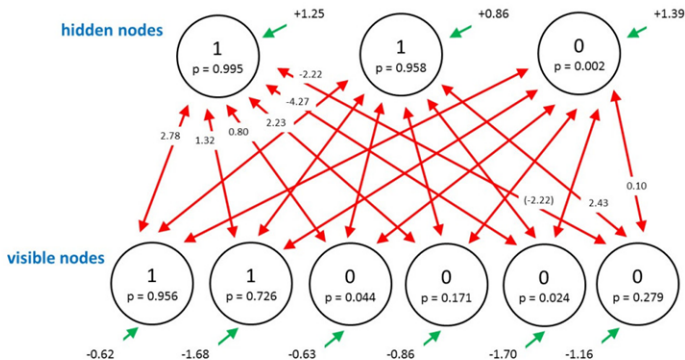
- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura**
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

Representação da estrutura de uma RBM.



# Exemplo Numérico

A figura abaixo apresenta o resultado de uma aplicação em um Sistema de Recomendação de Filmes.



# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática**
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

## Definição

De acordo com Borin (2016), as RBM fazem parte de uma classe de modelos probabilísticos que são baseados em energia.

- Energia Total

$$E(v, h) = - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} c_i v_i - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} h_j w_{ji} v_i \quad (2)$$

- Distribuição de Probabilidade

$$P(v, h) = \frac{e^{-E(v, h)}}{Z} \quad (3)$$

- Ou seja, uma RBM fica totalmente especificada pelos seus vieses e pesos de conexão, pelo conjunto de parâmetros **W**, **b**, **c**

- Por meio de uma abordagem não supervisionada, o modelo é treinado com o Método do Gradiente Descendente Estocástico - SGD, aplicado na seguinte **Função Perda**.

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{v}^{(t)}) = -\ln P(\mathbf{v}^{(t)}) \quad (4)$$

- Gradiente

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}; \mathbf{v}^{(t)})}{\partial \boldsymbol{\theta}} = \mathbb{E}_{P(\mathbf{h}|\mathbf{v})} \left[ \frac{\partial E(\mathbf{v}^{(t)}, \mathbf{H})}{\partial \boldsymbol{\theta}} \Big| \mathbf{v}^{(t)} \right] - \mathbb{E}_{P(\mathbf{v}, \mathbf{h})} \left[ \frac{\partial E(\mathbf{V}, \mathbf{H})}{\partial \boldsymbol{\theta}} \right] \quad (5)$$

- Algoritmo CD (*Contrastive Divergence*)

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}; \mathbf{v}^{(t)})}{\partial \boldsymbol{\theta}} \approx \frac{\partial \mathcal{F}(\mathbf{v}^{(t)})}{\partial \boldsymbol{\theta}} - \frac{\partial \mathcal{F}(\tilde{\mathbf{v}})}{\partial \boldsymbol{\theta}} \quad (6)$$

- A partir do **gradiente** calculado, os parâmetros são atualizados a cada rodada de treinos, respeitando a seguinte regra:

$$\theta \leftarrow \theta - \lambda \left( \frac{\partial \mathcal{F}(\mathbf{v}^{(t)})}{\partial \theta} - \frac{\partial \mathcal{F}(\tilde{\mathbf{v}})}{\partial \theta} \right) \quad (7)$$

- Ao desenvolver a equação acima para cada um dos parâmetros, temos que:

$$\begin{aligned} b_j &\leftarrow b_j + \lambda \left[ P(h_j = 1 | \mathbf{v}^{(t)}) - P(h_j = 1 | \tilde{\mathbf{v}}) \right] \\ c_i &\leftarrow c_i + \lambda \left[ v_i^{(t)} - \tilde{v}_i \right] \\ w_{ji} &\leftarrow w_{ji} + \lambda \left[ P(h_j = 1 | \mathbf{v}^{(t)}) v_i^{(t)} - P(h_j = 1 | \tilde{\mathbf{v}}) \tilde{v}_i \right] \end{aligned} \quad (8)$$

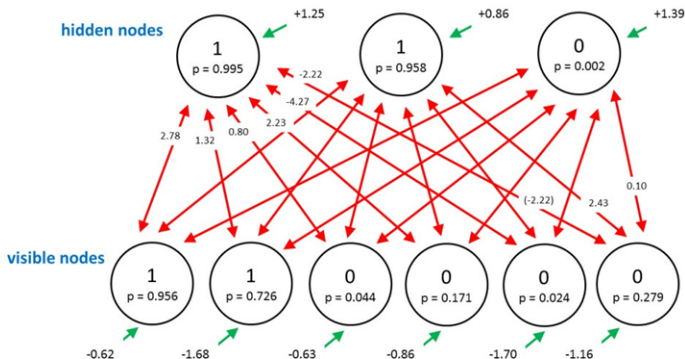
- Para acompanhar a evolução do treinamento, geralmente se utiliza uma medida de **erro de reconstrução**:

$$\varepsilon_r = \left\| \tilde{\mathbf{v}} - \mathbf{v}^{(t)} \right\|^2 \quad (9)$$



# Exemplo Numérico

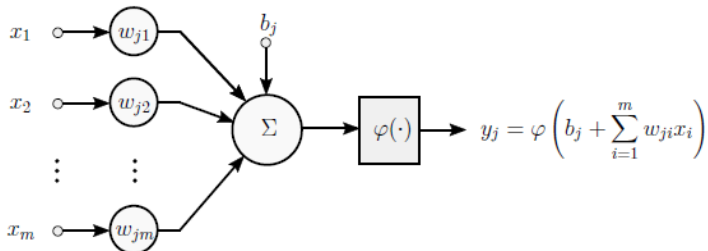
Novamente o Sistema de Recomendação de Filmes.



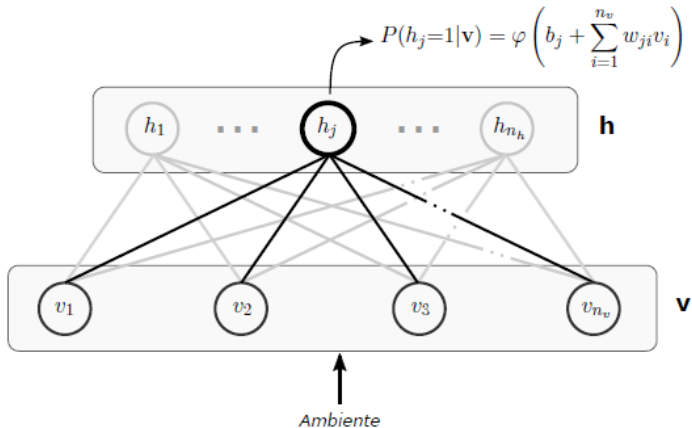
# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA**
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências

No **neurônio clássico**,  $y_i$  representa efetivamente a saída do neurônio, um valor real.



# RBM vs. RNA

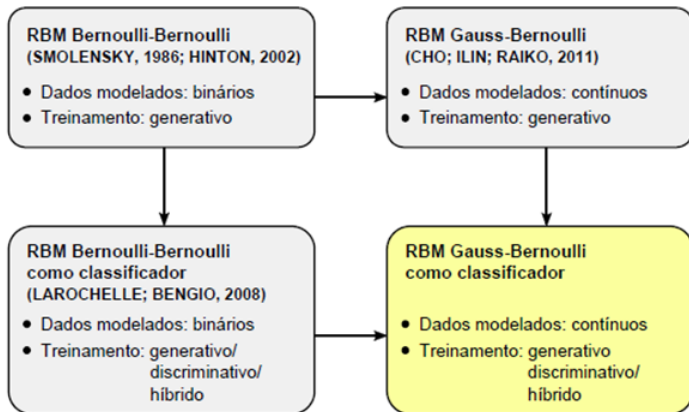


RBM é formada por **neurônios estocásticos**, pois expressa a probabilidade de que o neurônio assumo valor igual a 1.

# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM**
- 8 Aplicação
- 9 Referências

# Variantes da Estrutura

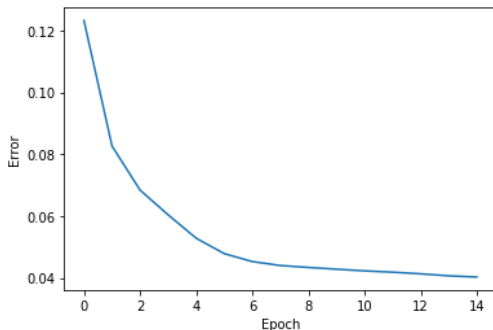


# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação**
- 9 Referências

# Exemplo

- Base de Dados do site movielens.org.
- Ratings que 6000 usuários atribuíram a 3883 filmes.
- Dados do Modelo:
  - 20 Camadas Ocultas e 3883 Camadas Visíveis
  - Epochs = 15
  - batchsize = 100





# Sumário

- 1 Motivação
- 2 Termodinâmica
- 3 Máquina de Boltzmann
- 4 Estrutura
- 5 Formalização Matemática
- 6 Comparação com uma RNA
- 7 Variantes da RBM
- 8 Aplicação
- 9 Referências**

## Referência da Apresentação

- Borin, R. G. (2016). Detecção de atividade vocal empregando máquinas de Boltzmann restritas (Doctoral dissertation, Universidade de São Paulo).

## Referência RBM Bernoulli-Bernoulli

- Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. Colorado Univ at Boulder Dept of Computer Science.
- Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8), 1771-1800.