



Universidade do Estado do Rio de Janeiro

centro

unidade patrono

Nome Sobrenome

Título do trabalho acadêmico

Cidade

aaaa

Nome Sobrenome

Título do trabalho acadêmico

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Curso, da Universidade do Estado do Rio de Janeiro.



Orientador: Cargo Titulação Nome Sobrenome
Coorientador: Cargo Titulação Nome Sobrenome

Cidade

aaaa

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/D

D979

Sobrenome, Nome

Título do trabalho acadêmico / Nome Sobrenome. – Cidade, aaaa-
?? f.

Orientador: Cargo Titulação Nome Sobrenome

Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro,
unidade, Programa de Pós-Graduação em Curso, aaaa.

1. primeira palavra chave.. 2. segunda palavra chave.. 3. terceira
palavra chave.. I. Cargo Titulação Nome Sobrenome. II. Universidade do
Estado do Rio de Janeiro. III. unidade. IV. Título

CDU 02:141:005.7

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta
dissertação, desde que citada a fonte.

Assinatura

Data

Nome Sobrenome

Título do trabalho acadêmico

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Curso, da Universidade do Estado do Rio de Janeiro.

Aprovada em dd de Mês de aaaa.

Banca Examinadora:

Cargo Titulação Nome Sobrenome (Orientador)
Unidade – Instituição

Cargo Titulação Nome Sobrenome (Coorientador)
Unidade – Instituição

primeiro membro titular da banca
instituição

segundo membro titular da banca
instituição

terceiro membro titular da banca
instituição

primeiro membro suplente da banca
instituição

segundo membro suplente da banca
instituição

terceiro membro suplente da banca
instituição instituição instituição instituição

Cidade

aaaa

DEDICATÓRIA

Texto da dedicatória

AGRADECIMENTOS

Texto de agradecimento

RESUMO

SOBRENOME, Iniciais. Do. Nome. *Título do trabalho acadêmico*. aaaa. ?? f.
Dissertação (Mestrado em Curso) – unidade, Universidade do Estado do Rio de Janeiro,
Cidade, aaaa.

Texto do resumo em português.

Palavras-chave: primeira palavra chave. segunda palavra chave. terceira palavra chave.

ABSTRACT

SOBRENOME, Iniciais. Do. Nome. *Title of dissertation.* aaaa. ?? f. Dissertação (Mestrado em Curso) – unidade, Universidade do Estado do Rio de Janeiro, Cidade, aaaa.

Abstract in English.

Keywords: first keyword. second keyword. third keyword.

LIST OF FIGURES

LIST OF TABLES

LISTA DE ALGORITMOS

LISTA DE ABREVIATURAS E SIGLAS

sigla1	por extenso
sigla2	por extenso
sigla3	por extenso

LISTA DE SÍMBOLOS

<i>simbolo1</i>	significado e/ou valor
<i>simbolo2</i>	significado e/ou valor
<i>simbolo3</i>	significado e/ou valor

CONTENTS

INTRODUÇÃO

Texto da introdução. Texto, texto texto (??), texto ??). Texto ?? em ??, texto ??, texto.

1 BOLTZMANN MACHINES

Here we begin by explaining the theory behind Boltzmann Machines

EXPLICAR OS ELEMENTOS DE PROBABILIDADE COM OS QUAIS ESTAMOS LIDANDO: QUEM SÃO AS VARIÁVEIS ALEATÓRIAS E COMO VAMOS DOMINÁ-LAS NO TEXTO, COMENTAR QUE ESTAMOS TRATANDO COM VARIÁVEIS DISCRETAS, COMO VAMOS IDENTIFICAR AS PROBABILIDADES,...

EXPLICAR O QUE OS ÍNDICES REPRESENTAM, e as variáveis,

MUDAR de alfa e beta para v e h.

1.1 Boltzmann Machines

Boltzmann Machines (BM) are a type of stochastic neural networks (SNN) where the connections between units, which are described by w , are symmetrical, i.e., $w_{ij} = w_{ji}$ [HERTZ]. This kind of stochastic neural networks are capable of learning internal representation and to model an input distribution. Boltzmann Machines were named after the Boltzmann distribution. Due to its stochastic behaviour, the probability of the state of the system to be found in a certain configuration is given by previous mentioned distribution [HERTZ]. According to [MONTUFAR, 2018], BM can be seen as an extension of Hopfield networks to include hidden units.

Boltzmann Machines have visible and hidden units. The visible units are linked to the external world and they correspond to the components of an observation. On the other hand, the hidden units do not have any connection outside of the network and model the dependencies between the components of the observations [FISCHER, 2012]. In BM, there is no connection restriction, this means that every unit, visible or hidden, can be connected to every other unit as in a complete graph, this pattern is not mandatory as some of the connections may not exist depending on the network layout.

Training Boltzmann Machines means finding the right connection between the units.

Boltzmann Machines (BM) are stochastic neural networks with symmetric connections, i.e., $w_{ij} = w_{ji}$. Boltzmann Machines use the Boltzmann distribution to determine the probability of the state of the system of the network. BM resembles the Hopfield networks with the inclusion of hidden units. Finding the right connections between the hidden units without knowing it from the training patterns what the hidden units represent is part of solving the Boltzmann Machine problem.

Units x_i in BM are split into two kinds: visible and hidden units. The visible units have connection to the outside world and are the units that receive the data input.

On the other hand, the hidden units do not have any connection to the outside of the network and they are responsible to find the data relation from the input. In a BM, the connections between units can be complete or not. Regardless of how the connections are, every connection in a BM is symmetric.

BM are made of stochastic units x_i which each of them can assume a binary value with a certain probability as follows:

$$x_i = \begin{cases} 1 & \text{with probability } g(h_i) \\ 0 & \text{with probability } 1 - g(h_i) \end{cases}, \quad (1)$$

where

$$h_i = \sum_j w_{ij} x_j, \quad (2)$$

and

$$g(h_i) = \frac{1}{1 + e^{-2\beta h_i}}. \quad (3)$$

Due to the symmetrical connections, there is an energy function give by

$$H(\mathbf{x}) = - \sum_i \sum_j w_{ij} x_i x_j - \sum_i w_{ii} x_i, \quad (4)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$, and n is equal to the number of units in the network, and the above energy function has minimum when there is a stable state characterised by

$$x_i = \text{sgn}(h_i). \quad (5)$$

The probability P of finding the system in a given state \mathbf{x} after the equilibrium is reached can be computed as follows:

$$P(\mathbf{x}) = \frac{1}{Z} e^{-\beta H(\mathbf{x})}, \quad (6)$$

where

$$Z = \sum_{\mathbf{x}'} e^{-\beta H(\mathbf{x}')} \quad (7)$$

is the partition function.

The learning process of a Boltzmann Machine consists in adjusting the connections w_{ij} in such a way that the state of the visible units have a particular desired probability distribution.

Let us identify the state of the visible units by an index v and the state of the hidden units by an index h . Considering a system which has N visible units and K hidden units, the whole system have 2^{N+K} possibilities of states in which it can be found.

The joint probability P_{vh} is the probability of finding the visible and hidden units in the states v and h , respectively. This probability measurement is given by the Boltzmann distribution:

$$P_{vh} = \frac{e^{-\beta H_{vh}}}{Z}, \quad (8)$$

where

$$Z = \sum_u \sum_k e^{-\beta H_{uk}}, \quad (9)$$

and

$$H_{vh} = - \sum_i \sum_j w_{ij} x_i^{vh} x_j^{vh} - \sum_i w_{ii} x_i^{vh}. \quad (10)$$

As mentioned above, the problem a Boltzmann Machine is trying to solve is determining the connections w_{ij} between units such that the visible units have a certain probability distribution. In order to do that, we need to find the marginal probability of the state v in which the visible units are found regardless of the state h of the hidden units. The marginal probability P_v is given by

$$P_v = \sum_h P_{vh} = \sum_h \frac{e^{-\beta H_{vh}}}{Z}. \quad (11)$$

Although we know that P_v is a function of the connections w_{ij} , and that this is the probability of finding the visible units in the state v . We want the states to have a certain probability Q_v , i.e., a desired probability. This means that ideally we would like to match the empirical distribution of the data, even though we do not have access to the correct distribution, only to what the observed data has given us as an input to training the model.

One way to evaluate the difference between two probability distribution, for example, P_v and Q_v , is using the Kullback-Leibler divergence, which can also be referred to relative entropy, E , which will be our cost function. (EXPLICAR FUNÇÃO DE CUSTO e D_{KL} !!!).

$$E = \sum_v Q_v \ln \left(\frac{Q_v}{P_v} \right). \quad (12)$$

The relative entropy E has the property of always being equal or greater than

zero. It reaches zero only if $P_v = Q_v$, which means that we are able to retrieve the exactly probability distribution of the input data at the visible units.

$$\begin{aligned}
E &= \sum_v Q_v \ln \left(\frac{Q_v}{P_v} \right) \\
&\geq \sum_v Q_v \left(1 - \frac{P_v}{Q_v} \right) \\
&= \sum_v (Q_v - P_v) \\
&= \sum_v Q_v - \sum_v P_v = 1 - 1 \\
&\Rightarrow E \geq 0.
\end{aligned} \tag{13}$$

From the gradient descent equation

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}, \tag{14}$$

where

$$\begin{aligned}
E &= \sum_v Q_v \ln \left(\frac{Q_v}{P_v} \right) \\
&= \sum_v [\ln(Q_v) - \ln(P_v)].
\end{aligned} \tag{15}$$

In the following steps, we present the gradient descent derivation

$$\begin{aligned}
\Delta w_{ij} &= -\eta \frac{\partial E}{\partial w_{ij}} \\
&= -\eta \frac{\partial}{\partial w_{ij}} \left[\sum_v Q_v (\ln(Q_v) - \ln(P_v)) \right] \\
&= \eta \frac{\partial}{\partial w_{ij}} \left[\sum_v Q_v \ln(P_v) \right] \\
&= \eta \sum_v Q_v \frac{\partial}{\partial w_{ij}} [\ln(P_v)] \\
&\Rightarrow \Delta w_{ij} = \eta \sum_v \frac{Q_v}{P_v} \frac{\partial P_v}{\partial w_{ij}}.
\end{aligned} \tag{16}$$

To continue with the computation of Δw_{ij} , we have to find the derivative of $\partial P_v / \partial w_{ij}$, from the marginal probability, equation ??,

$$P_v = \frac{\sum_h e^{-\beta H_{vh}}}{\sum_u \sum_k e^{-\beta H_{uk}}}, \tag{17}$$

thus the derivative of P_v follows

$$\begin{aligned}
\frac{\partial P_v}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}} \left[\frac{\sum_h e^{-\beta H_{vh}}}{\sum_u \sum_k e^{-\beta H_{uk}}} \right] \\
&= \frac{1}{\sum_u \sum_k e^{-\beta H_{uk}}} \sum_h (-\beta) e^{-\beta H_{vh}} \frac{\partial H_{vh}}{\partial w_{ij}} - \sum_h e^{-\beta H_{vh}} \frac{1}{(\sum_u \sum_k e^{-\beta H_{uk}})^2} \sum_u \sum_k e^{-\beta H_{uk}} (-\beta) \frac{\partial H_{uk}}{\partial w_{ij}}
\end{aligned} \tag{18}$$

2 TÍTULO DO CAPÍTULO 1

Texto do capítulo. Texto, texto, Figura ?? . Texto Figura ????

Figure 1 - Título da figura.



Legend: Texto da legenda.

Source: Citação da fonte ou 'O autor.'.

Table 1 - Título da tabela.

X	Y
1,20	15,7
1,23	15,6
1,19	15,3
1,26	15,1
1,22	15,5
1,16	15,3
1,37	15,7

Legend: Texto da legenda.

Source: Citação da fonte ou 'O autor.'.

Figure 2 - Título da figura.



(a)



(b)



(c)

Legend: Texto da legenda. ?? Texto da imagem. ?? Texto da imagem. ?? Texto da imagem.

Source: Citação da fonte ou 'O autor'.

3 TÍTULO DO CAPÍTULO 2

Texto do capítulo. Texto, texto Algoritmo ??. Texto.

Algoritmo 1 - Título do algoritmo.

DOCUMENTAÇÃO

TÍTULO

Nome do algoritmo

PROPÓSITO

Propósito do algoritmo.

MÉTODO

Método utilizado no algoritmo.

ENTRADAS

a, m: multiplicador e módulo

n0: semente

i: contador auxiliar

SAÍDAS

n: número aleatório

OBSERVAÇÕES, RESTRIÇÕES, REQUISITOS

Observações, restrições e requisitos.

ALGORITMO IDENTIFICAÇÃO

declarar a, m, i numéricos

declarar $n0, n$ numéricos

```

1.   $m \leftarrow 13$ 
2.   $n0 \leftarrow 1$ 
3.  para  $a$  de 2 até  $m - 1$  , fazer      {para cada possível valor de 'a'}
4.  |   escrever "a = ",  $a$ , ":",  $n = \{$ 
5.  |    $n \leftarrow n0$       {reinicia a geração com a semente  $n0$ }
6.  |   para  $i$  de 0 até  $m - 1$  , fazer
7.  |   |    $n \leftarrow \text{resto}(a * n, m)$       {gerador de números aleatórios}
8.  |   |   se ( $n == n0$ ), então      {se fim da sequencia...}
9.  |   |   |   escrever  $n, \{$ 
10. |   |   |   parar
11. |   |   senão
12. |   |   |   escrever  $n$ 
13. |   |   fim se
14. |   fim para
15. fim para
— continua —

```

Algoritmo 1 - Título do algoritmo. (continuação)

```
— continuação —  
a ← 1  
enquanto ( $a < 10$ ), fazer      {comentário}  
|   escrever a  
|   a ← a + 1  
fim enquanto  
a ← 1  
repetir      {comentário}  
|   escrever a  
|   a ← a + 1  
até que ( $a \geq 10$ )  
  
a ← 1  
fazer      {comentário}  
|   escrever a  
|   a ← a + 1  
enquanto ( $a < 10$ )  
FIM ALGORITMO  
FIM DOCUMENTAÇÃO
```

CONCLUSÃO

Texto da conclusão.

GLOSSÁRIO

termo	significado
termo	significado
termo	significado

APPENDIX A – Primeiro apêndice

A.1 Primeira seção

Texto da primeira seção.

A.1.1 Primeira subseção

Texto da primeira subseção.

A.1.1.1 Primeira subsubseção

Texto da primeira subsubseção.

APPENDIX B – Segundo apêndice

B.1 Primeira seção

Texto da primeira seção.

B.1.1 Primeira subseção

Texto da primeira subseção.

B.1.1.1 Primeira subsubseção

Texto da primeira subsubseção.

ANNEX A – Primeiro anexo

A.1 Primeira seção

Texto da primeira seção.

A.1.1 Primeira subseção

Texto da primeira subseção.

A.1.1.1 Primeira subsubseção

Texto da primeira subsubseção.

ANNEX B – Segundo anexo

B.1 Primeira seção

Texto da primeira seção.

B.1.1 Primeira subseção

Texto da primeira subseção.

B.1.1.1 Primeira subsubseção

Texto da primeira subsubseção.