Máquinas de Boltzmann

Aluno: Artur Chiaperini Grover
Orientadora: Dra. Roseli Suzi Wedemann

PPG-CComp
Universidade do Estado do Rio de Janeiro

21 de novembro de 2019

Sumário

Introdução

Redes neuronais artificiais são algoritmos computacionais que se inspiram na rede neuronal biológica. Estes algoritmos podem ser usados para resolver problemas onde uma solução analítica é difícil de ser encontrada, por exemplo, no caso de reconhecimento de padrão em imagens.

A inspiração nas redes neuronais biologicas, é devido ao fato de que as redes neuronais artificiais tem dois elementos basicos: os neuronios (unidades) e as sinapses (conexoes).

Introdução

Há diferentes algoritmos de redes neuronais artificiais, suas diferenças estão relacionadas a forma como esses dois elementos básicos são agrupados (como são feitas as conexões entre eles), e como cada unidade opera.

Exemplos de redes neuronais conhecidas são: perceptron, MLP, CNN, redes de Hopfield, máquina de Boltzmann, entre outras ...

Motivação e Objetivo

As diferentes redes neuronais possuem cada uma as suas peculiaridades, e problemas para os quais possuem melhor desempenho se comparadas com seus outros parentes.

Neste trabalho o nosso objetivo é entender o lado teórico da máquina de Boltzmann, e de suas derivadas, como a máquina restrita de Boltzmann, e aplicá-las em um problema simples afim de comprovar seu mecanismo de funcionamento.

Redes Neuronais Artificiais (ANN)

Precisamos de dois elementos básicos para definir uma rede neuronal: as unidades (os neurônios) e as conexões (as sinápses).

Redes Neuronais Artificiais (ANN)

Precisamos de dois elementos básicos para definir uma rede neuronal: as unidades (os neurônios) e as conexões (as sinápses).

Determinar como que as unidades ficam ativas ou não.

Redes Neuronais Artificiais (ANN)

Precisamos de dois elementos básicos para definir uma rede neuronal: as unidades (os neurônios) e as conexões (as sinápses).

Determinar como que as unidades ficam ativas ou não. Isso vai do tipo de rede neuronal.

ANN - A Unidade

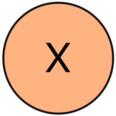

Figura 2.1: Neurônio

ANN - A Unidade

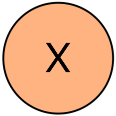

Figura 2.1: Neurônio

Neurônio – Unidade

ANN - Unidades e Conexões

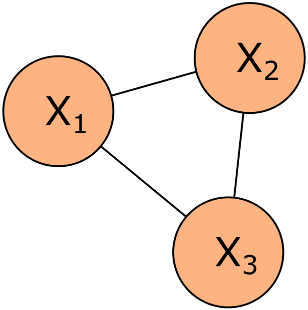

Figura 2.2: Rede neuronal artificial.

ANN - Unidades e Conexões

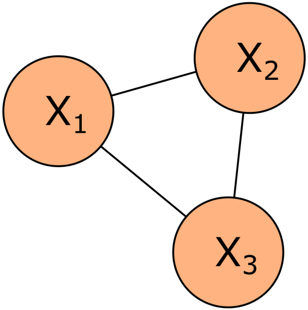

Figura 2.2: Rede neuronal artificial.

Conexões – Sinápse – Pesos

Treinamento e Padrões

Treinar uma rede neuronal significa determinar o valor das conexões entre as unidades baseado num conjunto de observações que são mostrados para a rede.

O conjunto de treinamento é composto de padrões, que são estados (configurações) da rede.

Redes de Hopfield

Uma rede de Hopfield é uma rede que armazena padrões de tal forma que quando um padrão novo é mostrado para a rede, ela responde devolvendo o padrão que tem armazenado mais próximo ao novo padrão.

Hopfield é uma rede determinística.

Quais os elementos básicos para Hopfield?

Vamos considerar que cada uma das **unidades** da rede é denominada por x_i, onde $i = 1, \ldots, N$, para uma rede com N unidades.

Em Hopfield, cada x_i pode assumir um dos valores $x_i \in \{0, 1\}$. Rede binária!

Se $x_i = 0$, a unidade i desativada; se $x_i = 1$, unidade i ativada.

Quais os elementos básicos para Hopfield?

Os **pesos** chamaremos de ω.

Na rede de Hopfield os pesos são simétricos, isto é, $\omega_{ij} = \omega_{ji}$ (conexão entre as unidades i e j).

Rede de Hopfield diagrama

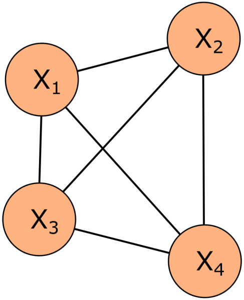

Figura 3.3: Distribuição dos neurônios de uma rede de Hopfield com 4 unidades.

Rede de Hopfield diagrama

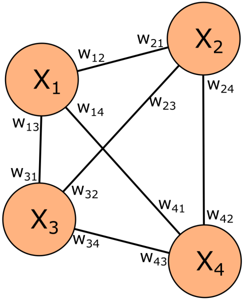

Figura 3.4: Identificação das conexões de uma rede de Hopfield com 4 unidades.

Estado da Rede

Exemplo de um estado $\mathbf{x}$ para uma rede binária com 4 unidades:

Estado da Rede

Exemplo de um estado $\mathbf{x}$ para uma rede binária com 4 unidades:

$$\begin{aligned}\mathbf{x} &= (x_1, x_2, x_3, x_4)\\ &= (0, 1, 1, 0)\end{aligned}$$

Função de Energia

Segundo **HOPFIELD (1982)**, como as conexões são simétricas, existe uma função de energia H para a rede, dado o estado $\mathbf{x}$ em que ela se encontra.

$$H(\mathbf{x}) = -\frac{1}{2}\sum_i \sum_j \omega_{ij} x_i x_j - \sum_i \phi_i x_i, \qquad (3.1)$$

Equação (3.1) depende da correlação entre as unidades (primeiro termo), e depende do limiar de ativação de cada unidade (segundo termo). Por simplificação podemos considerar apenas o primeiro termo.

Aprendizagem

Fazer a rede de Hopfield aprender corresponde a determinar os valores das conexões ω_{ij}, a partir de um conjuto de padrões apresentado para a rede. Isto é, aprender quais são os mínimos da rede para o conjunto de treinamento em questão,

$$\omega_{ij} = \frac{1}{N} x_i x_j. \tag{3.2}$$

(considerando que a rede aprende apenas um padrão.)

Função de Ativação

O valor x_i que a unidade x_i possui é dada pela função degrau.

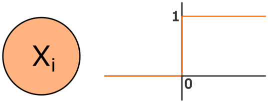

Figura 3.5: Função de ativação de uma unidade na rede de Hopfield.

$$x_i = \Theta\left(\sum_j \omega_{ij} x_j\right) \tag{3.3}$$

Funcionamento da rede de Hopfield

Uma vez que a rede determinou o valor de suas conexões, podemos mostrar um novo estado para a rede, por exemplo, $\mathbf{x} = (1, 1, 0, 0)$.

Partindo de uma unidade escolhida aleatoriamente, determinamos se a unidade fica ativa ou não (muda de valor) avaliando a diferença de energia

$$\Delta H_i = H_i(x_i = 0) - H_i(x_i = 1). \tag{3.4}$$

Funcionamento da rede de Hopfield

Hopfield tem característica determinista porque a rede sempre opera indo em direção a menor energia.

Pode acontecer de ficar preso em mínimo local.

Em um problema com muitas variáveis, isso pode ser tornar um problema, pois podemos não chegar ao melhor resultado. Há outras formas de se contornar isso!

Redes Estocásticas

Podemos considerar também o caso em que as unidades x_i de uma rede são variáveis aleatórias.

A unidade x_i tem agora certa probabilidade $g(h_i)$ de ser encontrado com valor $x_i = 1$, e probabilidade $1 - g(h_i)$, com valor $x_i = 0$.

Função de Ativação

Outra forma para a função de ativação:

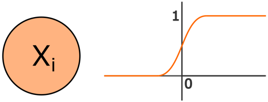

Figura 4.6: Função de ativação no caso estocástico. Trata-se de uma função logistica, e não mais uma função degrau.

Função de Ativação

Função logística, ao invés de função degrau ...

$$x_i = g(h_i) = \frac{1}{1 + e^{(-\beta h_i)}} \tag{4.1}$$

$$h_i = \sum_j \omega_{ij} x_j \tag{4.2}$$

Temperatura – Ruído

Na equação (4.1), temos o parâmetro β, que corresponde ao inverso da temperatura T.

Esta temperatura não tem o mesmo significado físico, mas é usado aqui para introduzir ruído. Controlar a inclinação da função logística.

Se β for muito grande, ou seja, temperatura muito baixa, o sistema retorna à situação determinística. A função logística passa a ser a função degrau.

Possibilidade de escapar de mínimos locais!

Rede Estocásticas diagrama

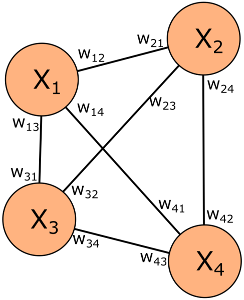

Figura 4.7: Rede estocástica com 4 unidades, e seus respectivas pesos.

Funcionamento da rede Estocástica

Novamente, mostramos um novo estado para a rede, por exemplo, $\mathbf{x} = (1, 1, 1, 0)$.

Partimos de uma unidade aleatória para ser atualizada. Agora para determinarmos se ela está ativa ou não, pela função logistica, equação (4.1).

Funcionamento da rede Estocástica

Considerando que todas as unidades são atualizadas de forma aleatória, a rede vai alcançar um equilíbrio, isto é, o valor das unidades deixará de mudar. Dizemos que o sistema encontra-se estacionário.

Nesta condição, a probabilidade de encontrarmos a rede no estado em que se manteve estacionária, $\mathbf{x}_s$, é dada pela distribuição de Boltzmann

$$P(\mathbf{x}_s) = \frac{e^{-\beta H(\mathbf{x}_s)}}{Z}, \tag{4.3}$$

$$Z = \sum_{\mathbf{u}} e^{-\beta H(\mathbf{u})}. \tag{4.4}$$

Máquina de BOLTZMANN

Máquina de Boltzmann

Rede estocástica com $\omega_{ij} = \omega_{ji}$.

Possui dois tipos de unidades distintas: unidades **visíveis** e unidades **escondidas**.

Visíveis: ligadas ao meio externo; correspondem às variáveis do problema.

Escondidas: não tem ligação do o meio externo; determinam a relação entre as variáveis do problema.

Arquitetura de BMs

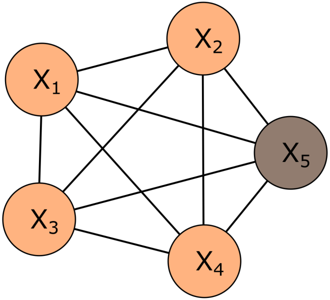

Figura 5.8: Diagrama de uma máquina de Boltzmann.

Arquitetura de BMs

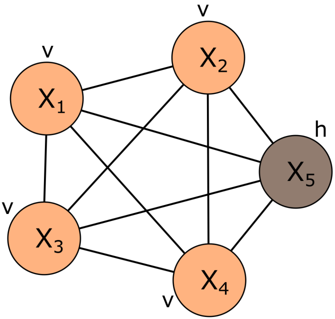

Figura 5.9: Diagrama de uma máquina de Boltzmann; unidades mais claras são as visíveis, e unidades mais escuras, as escondidas.

O sistema

Vamos considerar que temos uma rede com N unidades visíveis, e K unidades escondidas,

O sistema

Vamos considerar que temos uma rede com N unidades visíveis, e K unidades escondidas, em que as unidades visíveis estão no estado v, e as escondidas, no estado h.

O sistema

Vamos considerar que temos uma rede com N unidades visíveis, e K unidades escondidas, em que as unidades visíveis estão no estado v, e as escondidas, no estado h.

Temos $2^{(N+K)}$ possibidades de estados em que a rede pode ser encontrada.

O sistema

Vamos considerar que temos uma rede com N unidades visíveis, e K unidades escondidas, em que as unidades visíveis estão no estado v, e as escondidas, no estado h.

Temos $2^{(N+K)}$ possibilidades de estados em que a rede pode ser encontrada.

Uma unidade $\mathrm{x}_i = x_i$, sendo que $x_i \in \{0, 1\}$.

Objetivo da Máquina de Boltzmann

Na máquina de Boltzmann fazemos o ajuste das conexões ω_{ij} para termos um certo estado nas unidades visíveis com uma distribuição de probabilidade desejada.

Se precisamos determinar as conexões ω_{ij}, precisamos de um mecanismo para atualizar os pesos, temos que resolver:

Objetivo da Máquina de Boltzmann

Na máquina de Boltzmann fazemos o ajuste das conexões ω_{ij} para termos um certo estado nas unidades visíveis com uma distribuição de probabilidade desejada.

Se precisamos determinar as conexões ω_{ij}, precisamos de um mecanismo para atualizar os pesos, temos que resolver:

$$\omega_{ij}^{(final)} = \omega_{ij}^{(inicial)} + \Delta\omega_{ij}. \tag{5.1}$$

Objetivo da Máquina de Boltzmann

Na máquina de Boltzmann fazemos o ajuste das conexões ω_{ij} para termos um certo estado nas unidades visíveis com uma distribuição de probabilidade desejada.

Se precisamos determinar as conexões ω_{ij}, precisamos de um mecanismo para atualizar os pesos, temos que resolver:

$$\omega_{ij}^{(final)} = \omega_{ij}^{(inicial)} + \Delta\omega_{ij}. \tag{5.1}$$

Lembrando que os pesos são ajustados baseados nos exemplos de treinamento que a rede recebe.

Grandezas Importantes

Energia para rede com unidades visíveis no estado v, e escondidas, no estado h:

$$H_{vh} = -\frac{1}{2} \sum_i \sum_j \omega_{ij} x_i^{(vh)} x_j^{(vh)} - \sum_i \phi_i x_i^{(vh)}. \tag{5.2}$$

Podemos calcular a soma de todos os possíveis estados da rede pela função partição:

$$Z = \sum_u \sum_k e^{(-\beta H_{uk})}. \tag{5.3}$$

Grandezas Importantes

A probabilidade P_v de achar as unidades visíveis de uma máquina de Boltzmann em um determinado estado v, é

$$P_v = \sum_h \frac{1}{Z} e^{(-\beta H_{vh})}. \tag{5.4}$$

Também pode ser escrito como

$$P_v = \frac{\sum_h e^{-\beta H_{vh}}}{\sum_u \sum_k e^{-\beta H_{uk}}}. \tag{5.5}$$

Entropia Relativa - Divergência Kullback-Leibler

Se temos duas distribuições de probabilidade distintas, por exemplo, P e R, sobre um mesmo espaço de estados, podemos calcular a diferença entre estas distribuições pela entropia relativa, E, (ou divergência de Kullback-Leibler, $D_{KL}(R||P)$).

Entropia Relativa - Divergência Kullback-Leibler

Com as máquinas de Boltzmann queremos determinar a probabilidade P_v de achar as unidades visíveis em determinado estado v, como mostra a equação (5.4). Porém para este estado temos uma probabilidade deseja R_v.

$$E = \sum_v R_v \ln \left[\frac{R_v}{P_v} \right] \tag{5.6}$$

Atualização de $\Delta\omega_{ij}$

Usamos a nossa entropia relativa como função custo.

Pelo método do grandiente descendente, podemos calcular $\Delta\omega_{ij}$,

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial \omega_{ij}}, \tag{5.7}$$

onde η é a taxa de aprendizagem.

$E(P_v)$; $P_v(H_{vh})$; $H_{vh}(\omega_{ij})$.

Atualização de $\Delta\omega_{ij}$

Por simplificação, vamos considerar que a função de energia para a rede no estado v e h é dada apenas pela correlação entre as unidades (primeiro termo da equação (5.2)):

$$H_{vh} = -\frac{1}{2}\sum_i \sum_j \omega_{ij} x_i^{(vh)} x_j^{(vh)} \tag{5.8}$$

Atualização de $\Delta\omega_{ij}$

Para se determinar $\Delta\omega_{ij}$, temos que passar por uma série de derivações...

$$\Delta\omega_{ij} = \eta \sum_{v} \frac{R_v}{P_v} \frac{\partial P_v}{\partial \omega_{ij}} \tag{5.9}$$

(5.11)

Atualização de $\Delta\omega_{ij}$

Para se determinar $\Delta\omega_{ij}$, temos que passar por uma série de derivações...

$$\Delta\omega_{ij} = \eta \sum_v \frac{R_v}{P_v} \frac{\partial P_v}{\partial \omega_{ij}} \tag{5.9}$$

$$\frac{\partial P_v}{\partial \omega_{ij}} = \beta \left[- \sum_h \frac{e^{-\beta H_{vh}}}{Z} \frac{\partial H_{vh}}{\partial \omega_{ij}} + \sum_h \frac{e^{-\beta H_{vh}}}{Z} \sum_u \sum_k \frac{e^{-\beta H_{uk}}}{Z} \frac{\partial H_{uk}}{\partial \omega_{ij}} \right] \tag{5.10}$$

$$\tag{5.11}$$

Atualização de $\Delta\omega_{ij}$

Para se determinar $\Delta\omega_{ij}$, temos que passar por uma série de derivações...

$$\Delta\omega_{ij} = \eta \sum_{v} \frac{R_v}{P_v} \frac{\partial P_v}{\partial \omega_{ij}} \tag{5.9}$$

$$\frac{\partial P_v}{\partial \omega_{ij}} = \beta \left[-\sum_{h} \frac{e^{-\beta H_{vh}}}{Z} \frac{\partial H_{vh}}{\partial \omega_{ij}} + \sum_{h} \frac{e^{-\beta H_{vh}}}{Z} \sum_{u} \sum_{k} \frac{e^{-\beta H_{uk}}}{Z} \frac{\partial H_{uk}}{\partial \omega_{ij}} \right] \tag{5.10}$$

$$\frac{\partial H}{\partial \omega_{ij}} = -x_i x_j \tag{5.11}$$

Atualização de $\Delta\omega_{ij}$

Com as devidas substituições na equação (5.10), temos

$$\frac{\partial P_v}{\partial \omega_{ij}} = \beta \left[\sum_h P_{vh} x_i^{(vh)} x_j^{(vh)} - P_v \sum_u \sum_k P_{uk} x_i^{(uk)} x_j^{(uk)} \right] \quad (5.12)$$

Atualização de $\Delta\omega_{ij}$

Com as devidas substituições na equação (5.10), temos

$$\frac{\partial P_v}{\partial \omega_{ij}} = \beta \left[\sum_h P_{vh} x_i^{(vh)} x_j^{(vh)} - P_v \sum_u \sum_k P_{uk} x_i^{(uk)} x_j^{(uk)} \right] \quad (5.12)$$

No segundo termo, temos o valor médio da correlação entre as unidades i e j, que é dado por

$$\langle x_i x_j \rangle = \sum_u \sum_k P_{uk} x_i^{(uk)} x_j^{(uk)} \quad (5.13)$$

Atualização de $\Delta\omega_{ij}$

Substituindo as equações (5.12) e (5.13) em (5.9):

$$\Delta\omega_{ij} = \eta\beta\left[\sum_v \sum_h R_v \frac{P_{vh}}{P_v} x_i^{(vh)} x_j^{(vh)} - \sum_v R_v \langle x_i x_j \rangle\right] \quad (5.14)$$

Considerando o probabilidade condicional

$$P_{h|v} = \frac{P_{vh}}{P_v}, \quad (5.15)$$

podemos rearranjar o primeiro termo de (5.14),

$$\sum_v R_v \sum_h P_{h|v} x_i^{(vh)} x_j^{(vh)} = \sum_v R_v \langle x_i x_j \rangle^{(v)} = \langle\langle x_i x_j \rangle^{(v)}\rangle \quad (5.16)$$

$\Delta\omega_{ij}$

Por último, temos que o termo responsável pela atualização dos pesos é

$$\Delta\omega_{ij} = \eta\beta \left[\langle \langle x_i x_j \rangle^{(v)} \rangle_{fixo} - \langle x_i x_j \rangle_{llivre} \right] \tag{5.17}$$

O primeiro termo corresponde ao valor médio de acordo com a distribuição de probabilidade R_v do valor médio das correlações entre as unidades i e j quando as unidades visíveis estão no estado fixo v.

O segundo termo corresponde a média da correlação entre as unidades i e j para todas as combinações possíveis dos estados v e h do sistema.

Cronograma

- Finalizar a parte teócia da máquina de Boltzmann e estudar as máquinas restritas de Boltzmann (RBM) – **Janeiro 2020**
- Definir um caso para ser resolvido com RBM – **Janeiro 2020**
- Implementar e testar o algoritmo da RBM – **May 2020**
- Em paralelo continuar com o desenvolvimento da dissertação – **Junho 2020**

Obrigado! ☺