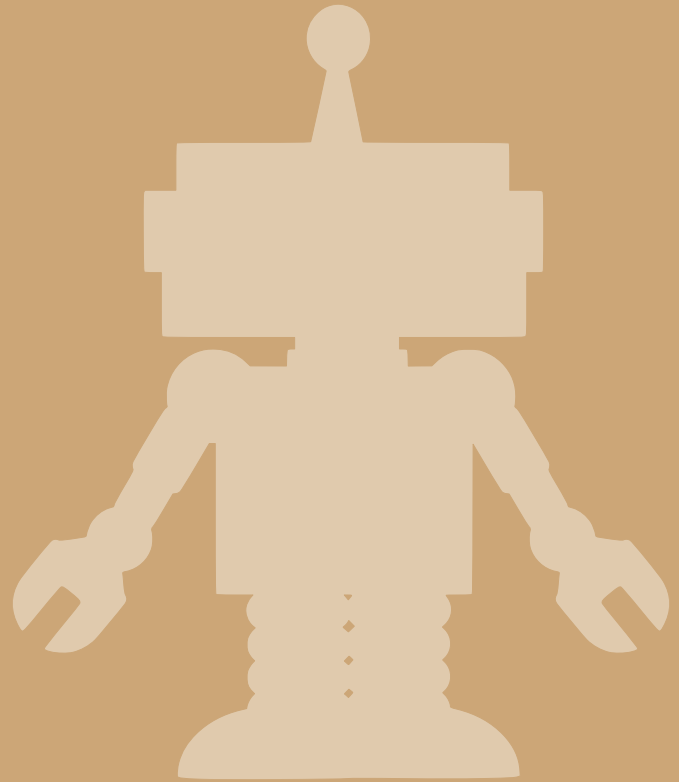


# Aprendizado de Máquina 2

Aula 9

Professora: Patrícia Pampanelli

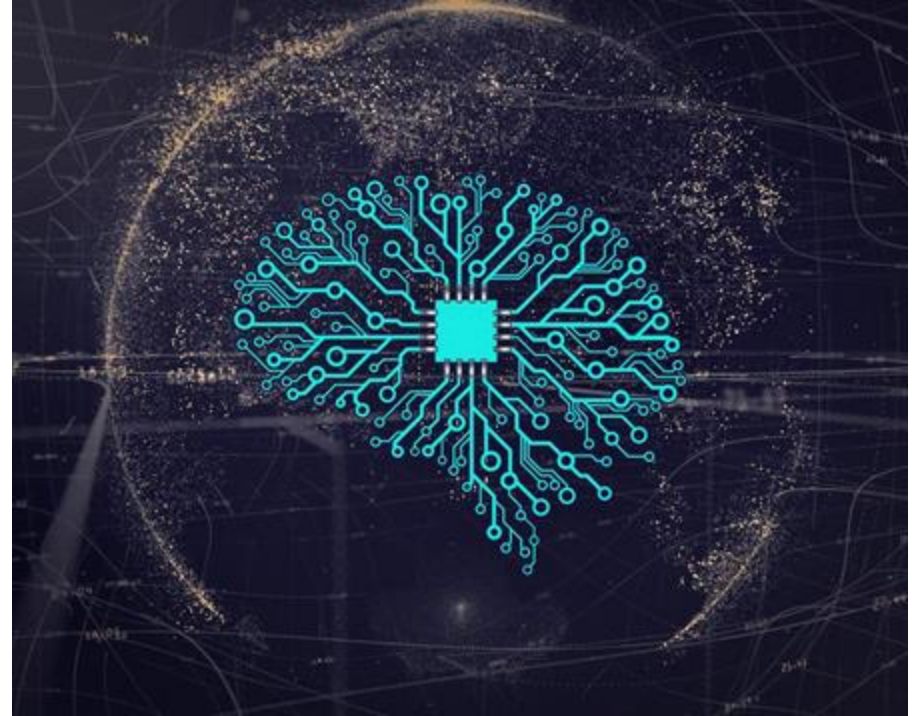
[patricia.pampanelli@usp.br](mailto:patricia.pampanelli@usp.br)



Dúvidas da última aula?

# Aula de Hoje

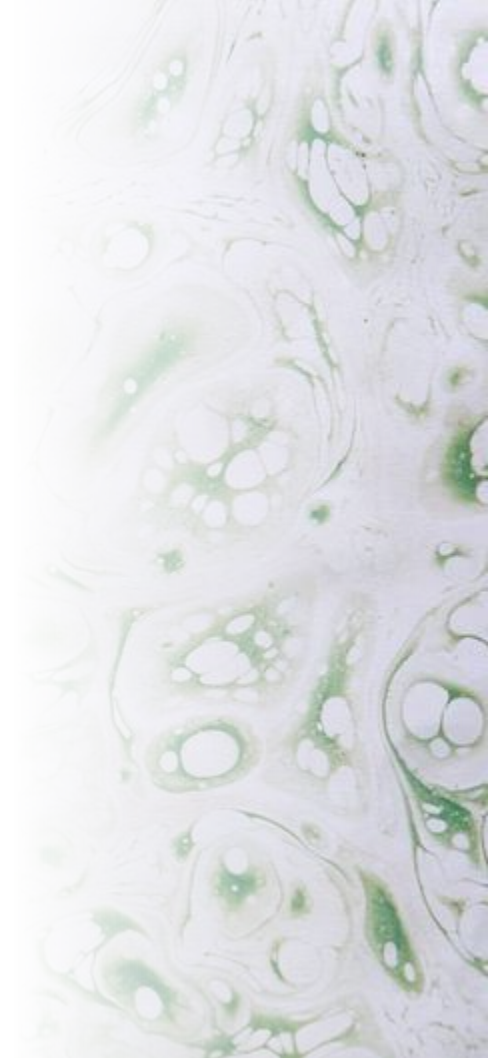
- **Máquina de Boltzmann**
- **Máquina Restrita de Boltzmann**



# Máquina de Boltzmann

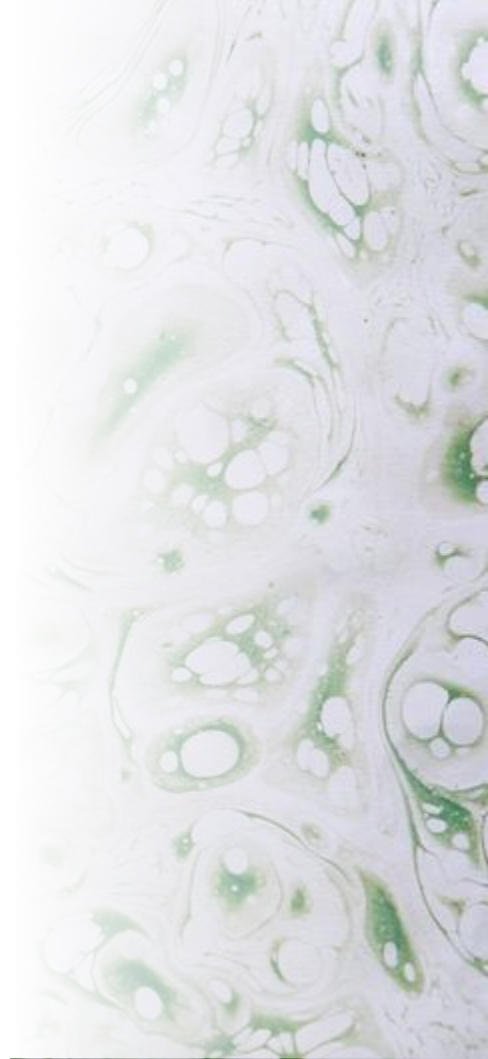
# Maquina de Boltzmann

- Proposto por Geoffrey Hinton (conhecido como o pai do Deep Learning) e Terry Sejnowski
- Tipo de treinamento **não supervisionado** capaz de aprender os padrões de um determinado conjunto de dados.
- A máquina de Boltzmann é um tipo especial de rede neural



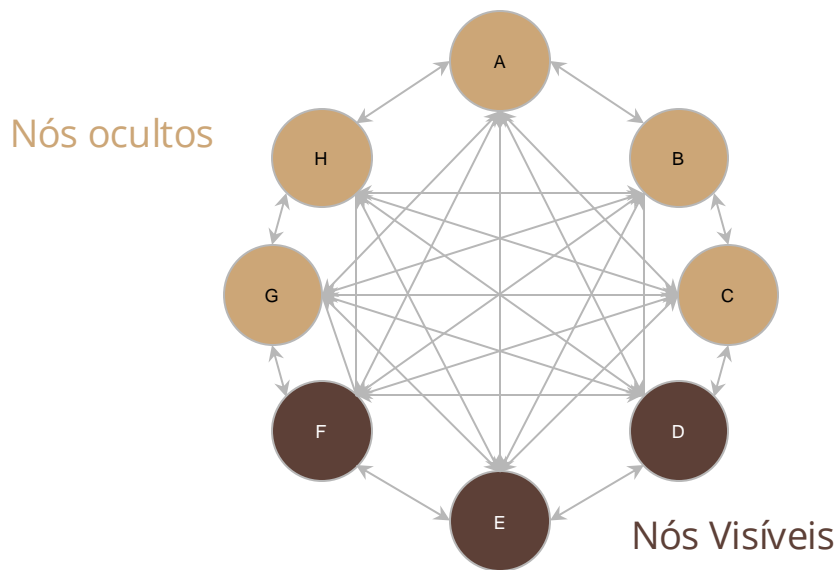
# Maquina de Boltzmann

- A máquina de Boltzmann também pode ser vista como um tipo de rede capaz de extrair padrões dos dados de entrada.
- Desta forma, as aplicações destas redes se assemelham aos autoencoders
- Essas redes também podem ser utilizadas para redução de dimensionalidade. Nestes casos, a camada oculta apresenta número menor de neurônios se comparado a camada de entrada



# Maquina de Boltzmann

- Composta somente de *hidden nodes* e *input nodes*

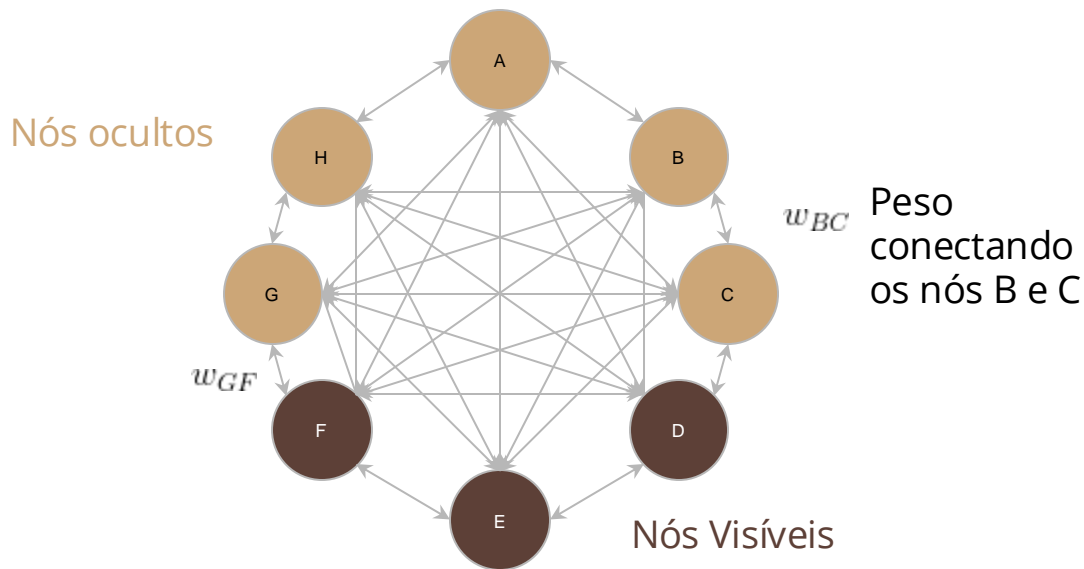






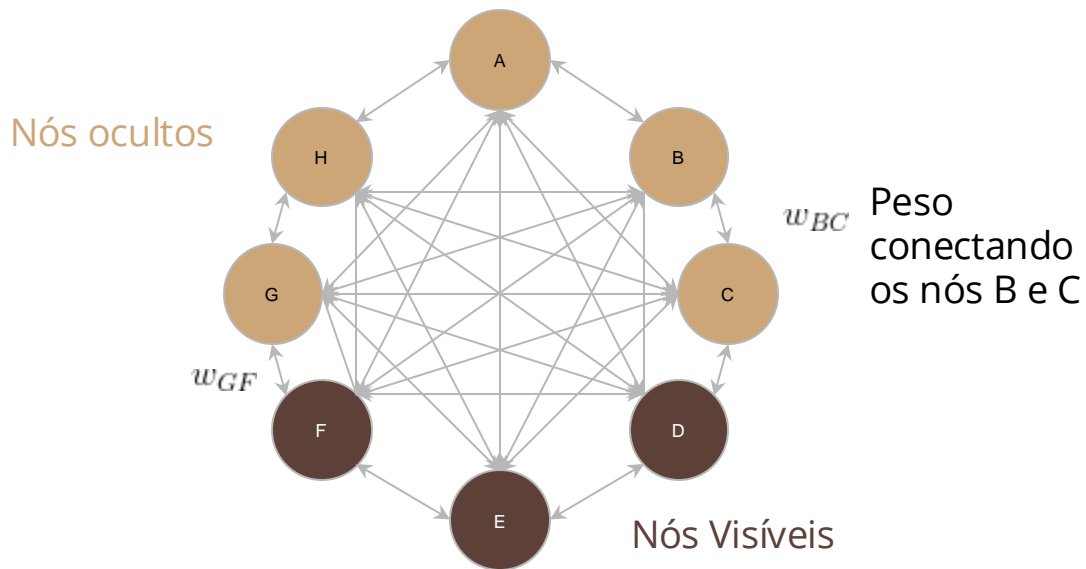
# Maquina de Boltzmann

- Cada conexão entre os nós tem um peso associado à ela.
- Os nós são responsáveis por tomar decisões probabilísticas sobre um determinado evento.



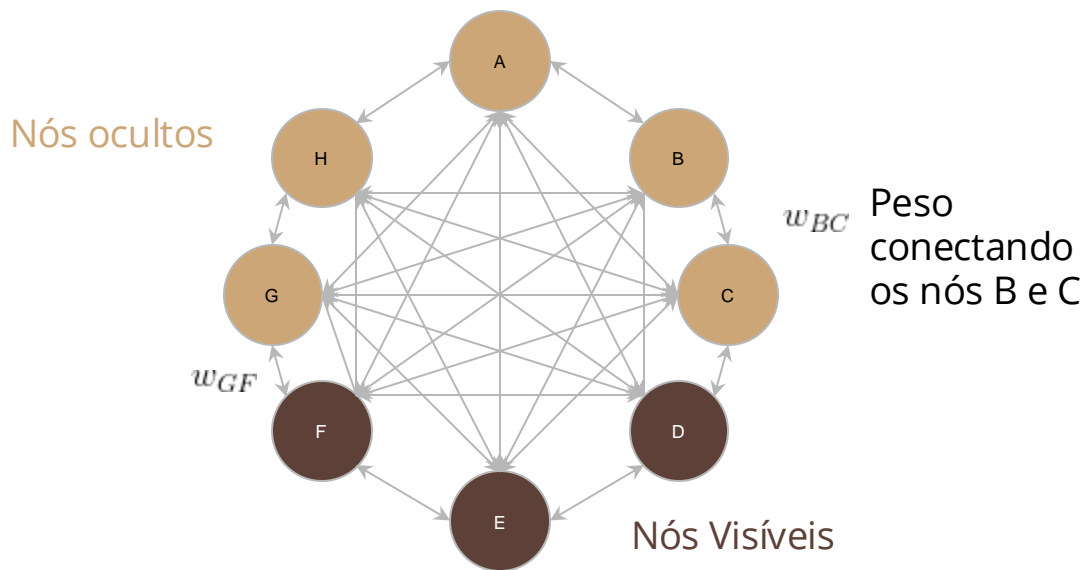
# Maquina de Boltzmann

- Os nós visíveis são aqueles com os quais nós conseguimos interagir. Os demais nós não são acessíveis externamente



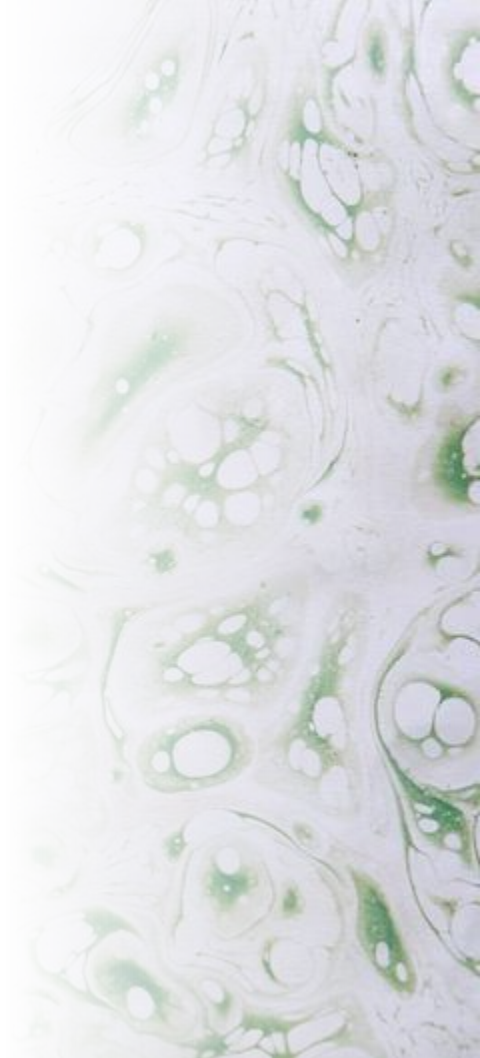
# Maquina de Boltzmann

- Os nós são responsáveis por tomar decisões



# Maquina de Boltzmann

- Ineficiência no treinamento:
  - O número de **conexões** estabelecidas **entre os neurônios (todos para todos)** faz com que o treinamento de uma máquina de Boltzmann seja um processo bastante **ineficiente**
  - Durante o período dos anos 1980s e 1990s as máquinas não restritas de Boltzmann tiveram um uso muito limitado devido à esta ineficiência no treinamento
- Os autores decidiram então modificar a arquitetura proposta originalmente para que se tornar-se mais eficiente



# Máquina Restrita de Boltzmann

# Máquina Restrita de Boltzmann (RBM)

- Modificações propostas por Geoffrey Hinton e os demais autores propostas nos anos de 2000s
- Foram propostas algumas restrições:
  - A principal diferença é que as RBMs tem somente uma camada de entrada e uma camada escondida
  - As conexões também são restritas. Neurônios de uma mesma camada não estão conectados mais
  - Em outras palavras, não existe conexões: *hidden-hidden layers* e *visible-visible layers*.
- A abordagem de treinamento das RBMs é estocástica. Diferente da abordagem do auto-encoder e das redes neurais é determinística.

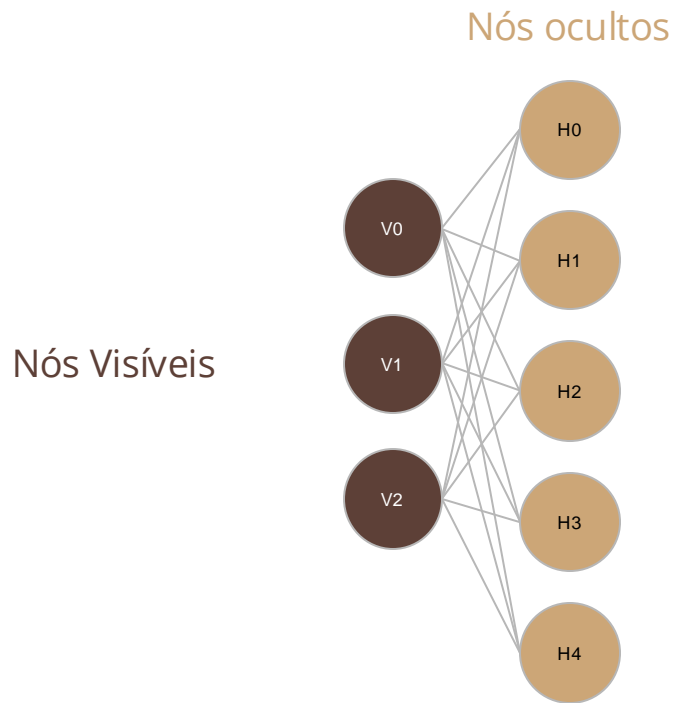


# Máquina Restrita de Boltzmann (RBM)

Nós Visíveis

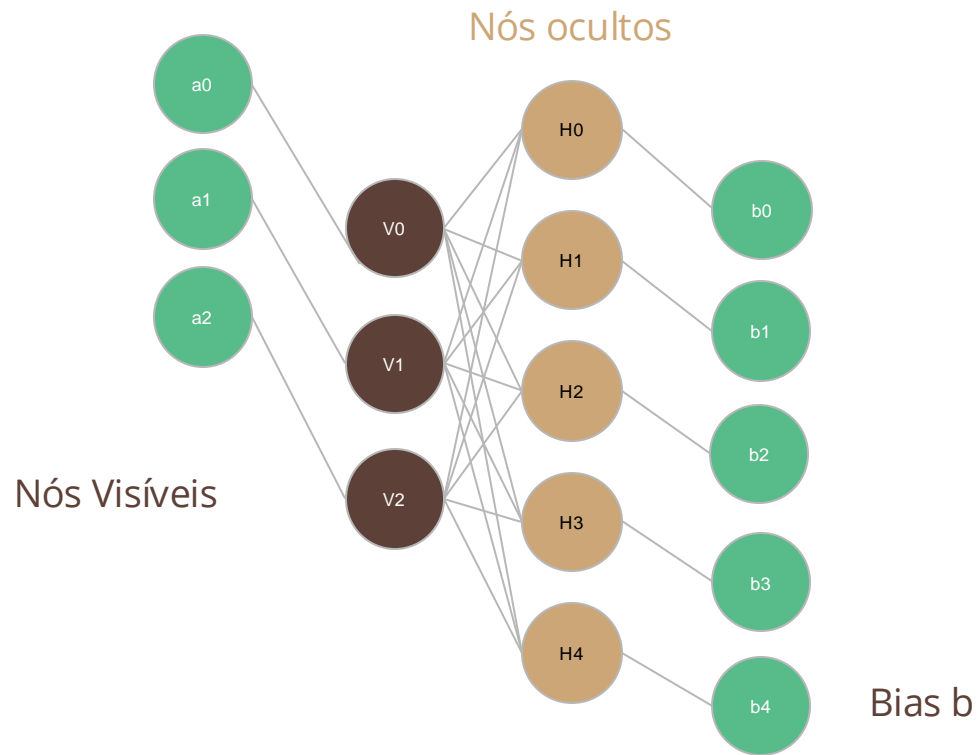


# Máquina Restrita de Boltzmann (RBM)

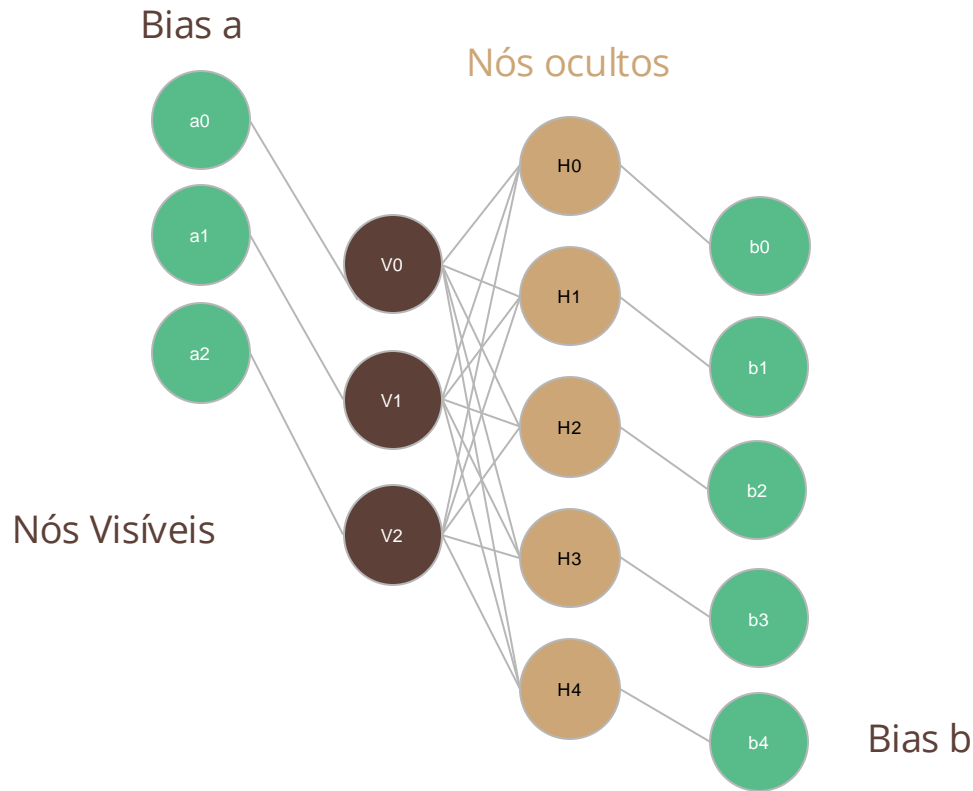




# Máquina Restrita de Boltzmann (RBM)



# Máquina Restrita de Boltzmann (RBM)



- Os nós visíveis são aqueles sobre os quais nós temos conhecimento. Em outras palavras, são estes eventos que temos definidos no dataset.
- Os nós invisíveis são aqueles sobre os quais não temos conhecimento. Esses nós são utilizados para nos ajudar a explicar os eventos conhecidos.

# Máquinas de Boltzmann vs Redes Neurais Feedforward

As RBMs são usadas para reconstruir entradas, enquanto redes feedforward são usadas para prever saídas

Característica	Máquina de Boltzmann	Rede Neural Feedforward
Tipo	Generativa	Discriminativa
Estrutura	Estocástica, sem camada de saída	Determinística, com camadas de saída
Aplicação	Modelagem probabilística	Classificação/Regressão

# Exemplo numérico

# Exemplo numérico

- Vamos utilizar um exemplo numérico de uma Máquina de Boltzmann Restrita.
- Nós temos as informações sobre a presença de três alunos em uma aula de pós graduação. Estas são as informações que compõem o nosso dataset.
- Estas informações representam os estados da camada visível da RBM.



Augusto



Bárbara



Priscila

Nós Visíveis

# Exemplo numérico

- Para nos ajudar a explicar o comparecimento destes alunos, nós temos dois estados invisíveis. Estes estados da camada escondida representam o tipo de aula ministrada (prática ou teórica).



Nós ocultos



Augusto



Bárbara



Priscila

Nós Visíveis

# Exemplo numérico

- Nós não sabemos *à priori* qual o tipo de aula que foi ministrada em cada dia.
- Em outras palavras, estes dados não se encontram em nosso dataset.



Nós ocultos



Augusto



Bárbara



Priscila

Nós Visíveis

# Exemplo numérico

- Nós temos em nosso dataset as informações de comparecimento dos alunos
- Além disso, nós temos pesos associados aos estados e também as interações entre a camada visível e a camada oculta



Nós ocultos



Augusto



Bárbara



Priscila

Nós Visíveis



# Exemplo numérico

- Nós temos em nosso dataset as informações de comparecimento dos alunos
- Além disso, nós temos pesos associados aos estados e também as interações entre a camada visível e a camada oculta



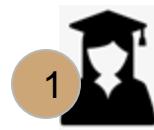
Nós ocultos



Augusto



Bárbara

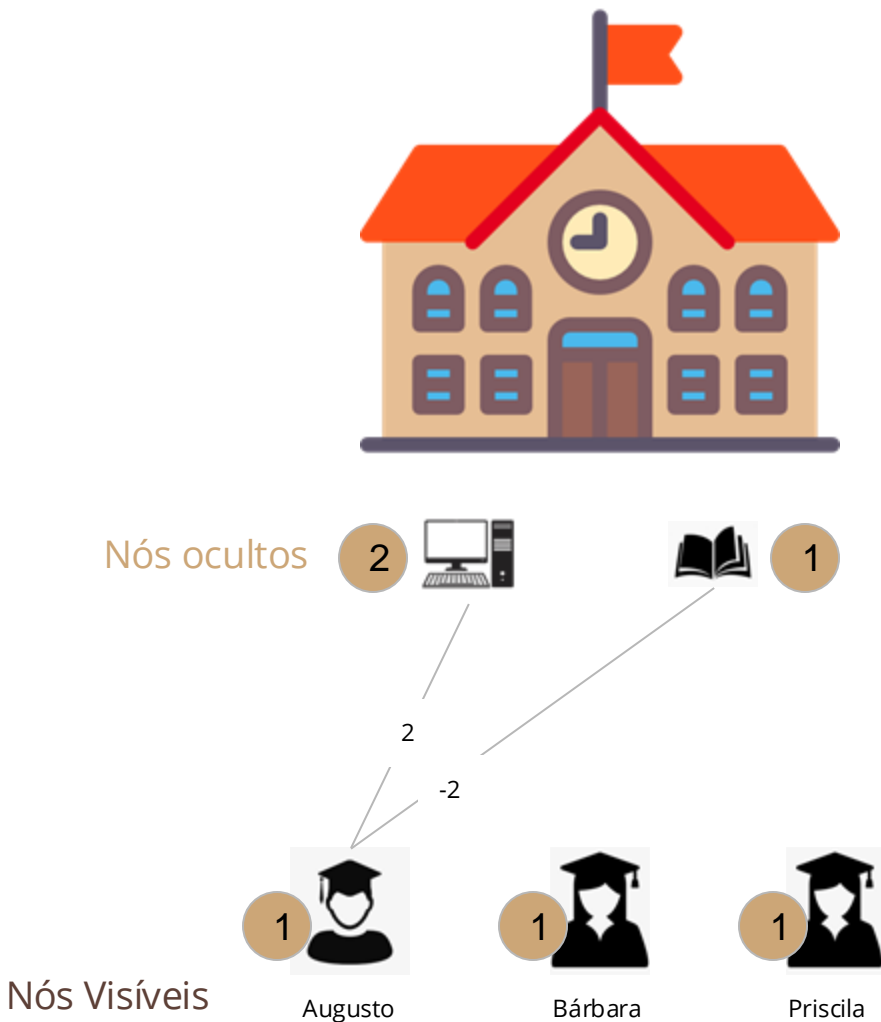


Priscila

Nós Visíveis

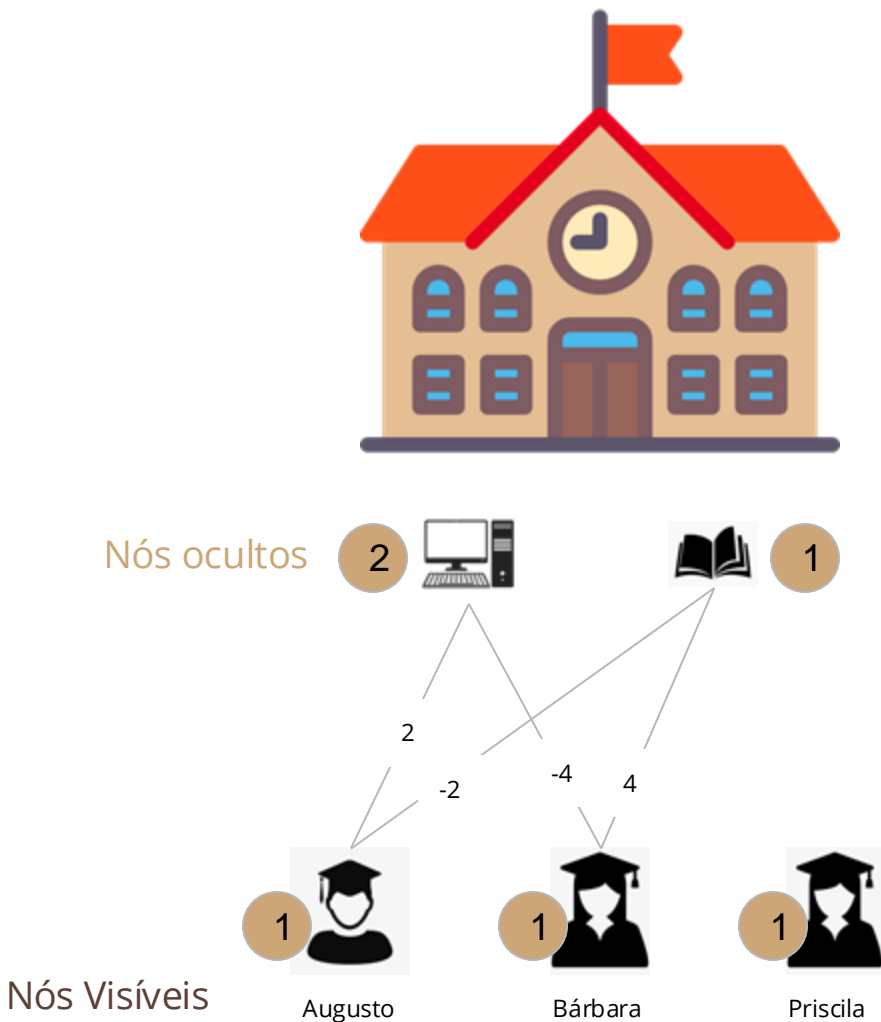
# Exemplo numérico

- Nós temos em nosso dataset as informações de comparecimento dos alunos
- Além disso, nós temos pesos associados aos estados e também as interações entre a camada visível e a camada oculta



# Exemplo numérico

- Nós temos em nosso dataset as informações de comparecimento dos alunos
- Além disso, nós temos pesos associados aos estados e também as interações entre a camada visível e a camada oculta

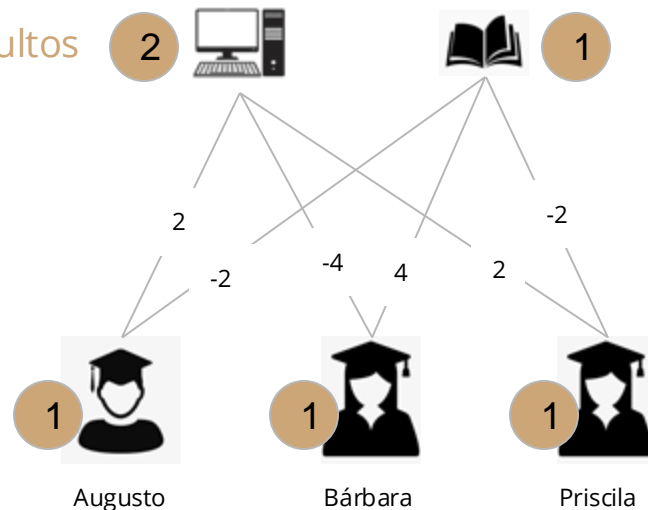


# Exemplo numérico

- Nós temos em nosso dataset as informações de comparecimento dos alunos
- Além disso, nós temos pesos associados aos estados e também as interações entre a camada visível e a camada oculta



Nós ocultos

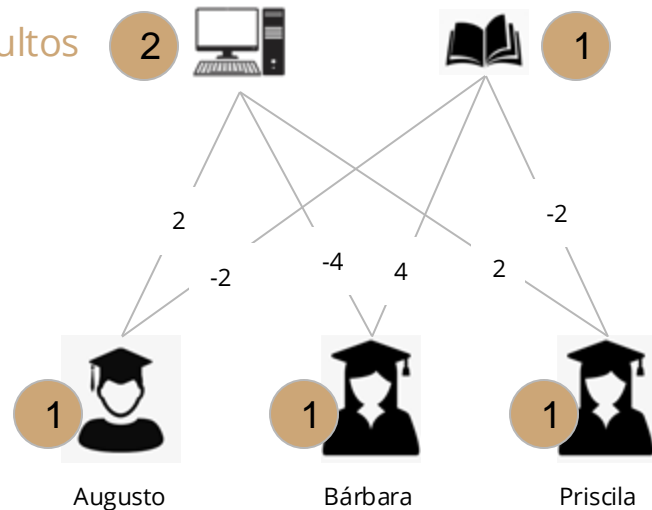


# Exemplo numérico

- O dataset é composto pelos estados de cada neurônio visível no dataset:

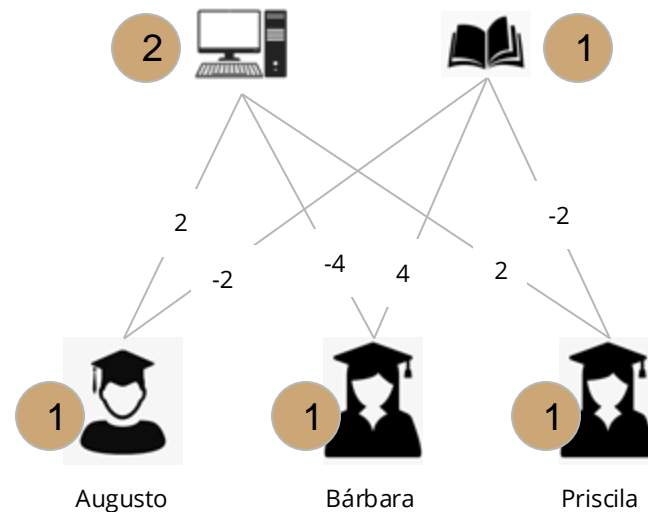
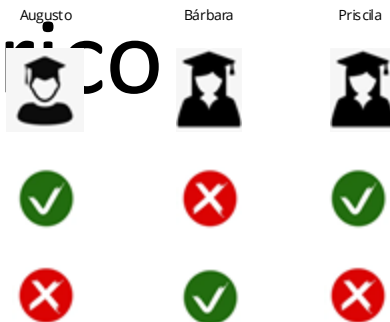


Nós ocultos



# Exemplo numérico

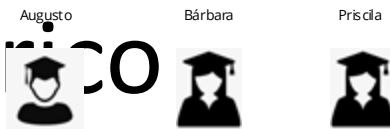
- O dataset é composto pelos estados de cada neurônio visível no dataset:



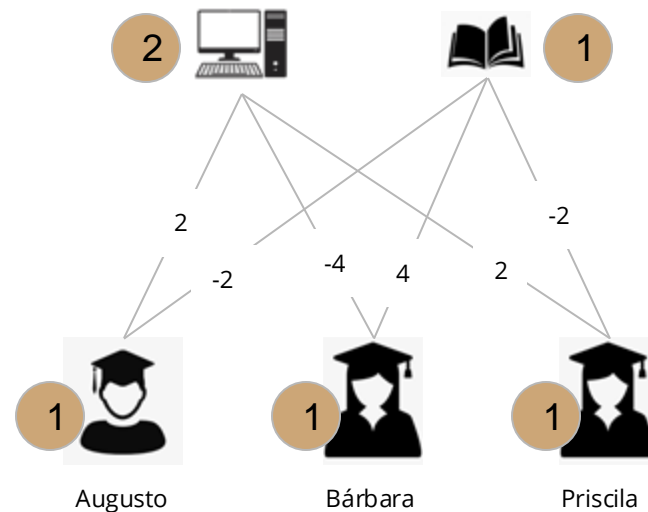
Nós Visíveis

# Exemplo numérico

- O dataset é composto pelos estados de cada neurônio visível no dataset:



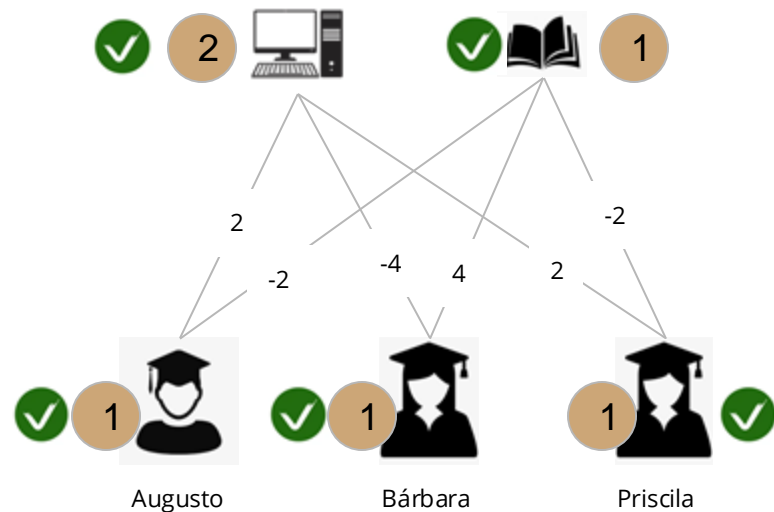
Nós Visíveis



# Exemplo numérico

- Podemos avaliar o score de cada cenário considerando os nós invisíveis e os nós visíveis da RBM.
- O cálculo do Score é feito fazendo simplesmente a soma dos pesos e dos estados verdadeiros.

Score  
???





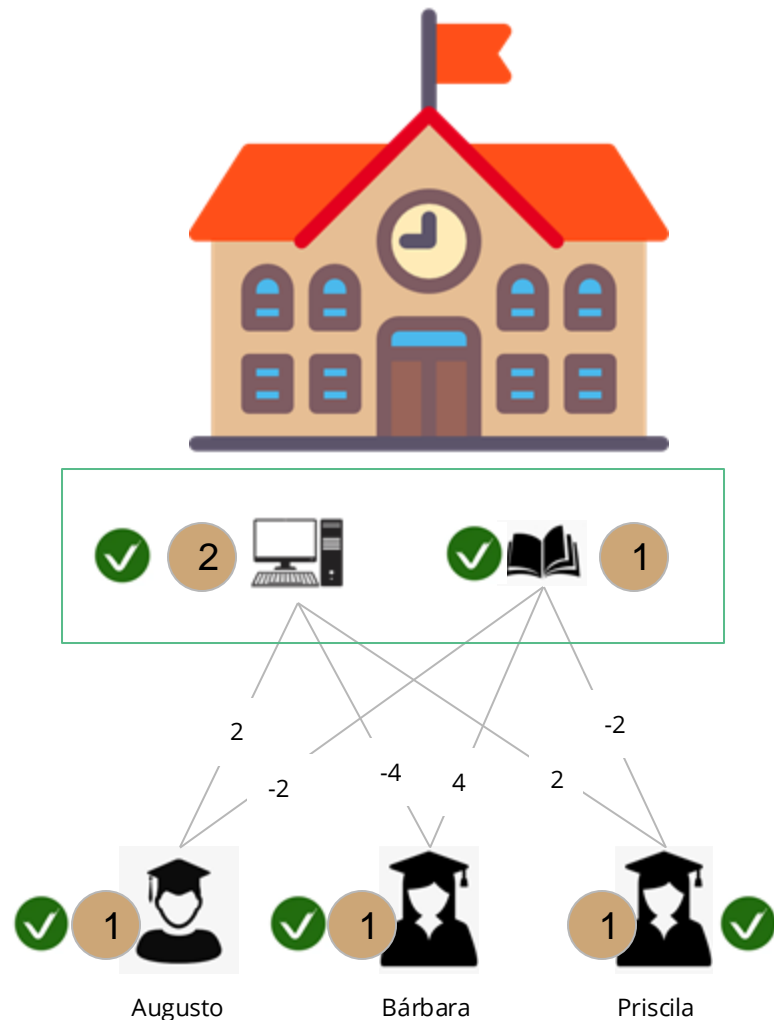
# Exemplo numérico

- Vamos verificar este estado onde **Augusto, Bárbara e Priscila compareceram a uma aula que teve teoria e prática:**



Score

???



# Exemplo numérico

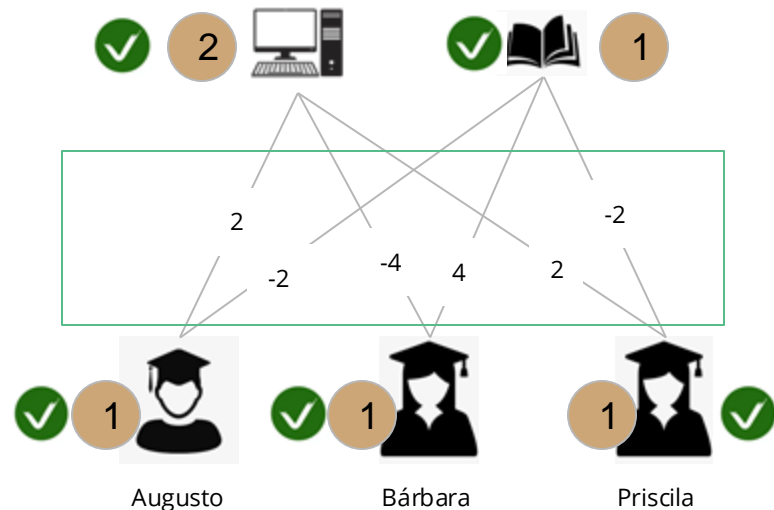
- Vamos verificar este estado onde **Augusto, Bárbara e Priscila compareceram a uma aula que teve teoria e prática:**



$$2 + (-2) + (-4) + 4 + 2 + (-2)$$

Score

???



# Exemplo numérico

- Vamos verificar este estado onde **Augusto, Bárbara e Priscila compareceram a uma aula que teve teoria e prática:**

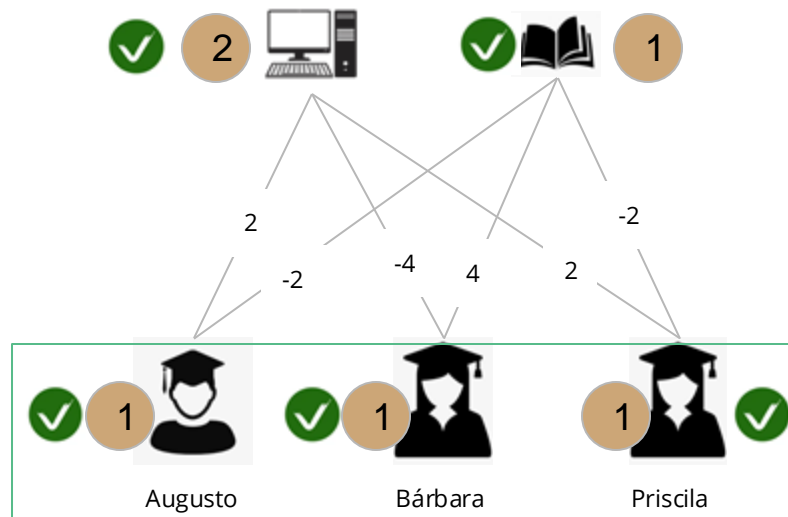
$$2 + 1$$

$$2 + (-2) + (-4) + 4 + 2 + (-2)$$

$$1 + 1 + 1$$

Score

???



# Exemplo numérico

- Vamos verificar este estado onde **Augusto, Bárbara e Priscila compareceram a uma aula que teve teoria e prática:**

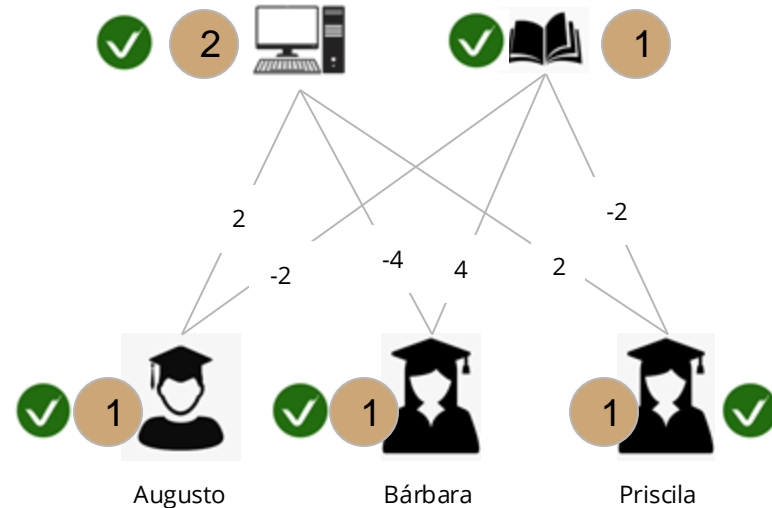
$$2 + 1$$

$$2 + (-2) + (-4) + 4 + 2 + (-2)$$

$$1 + 1 + 1$$

Score

6

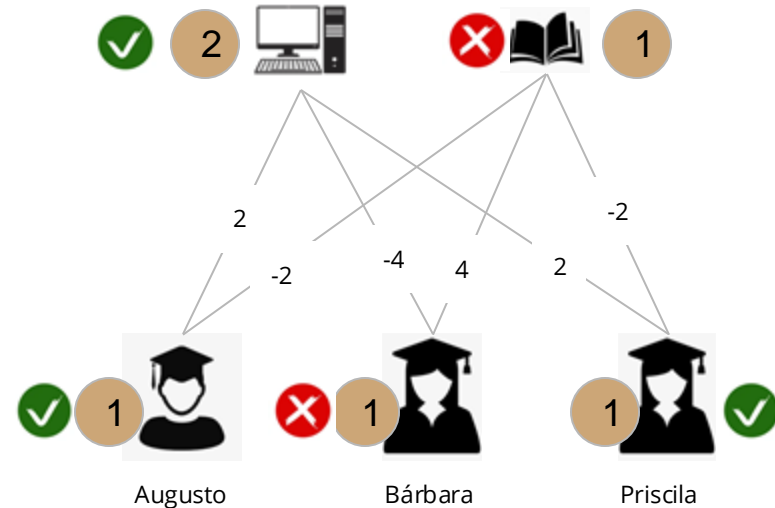


# Exemplo numérico

- Agora vamos avaliar um cenário onde **somente Augusto e Priscila** comparecem à uma aula prática:

Score

???

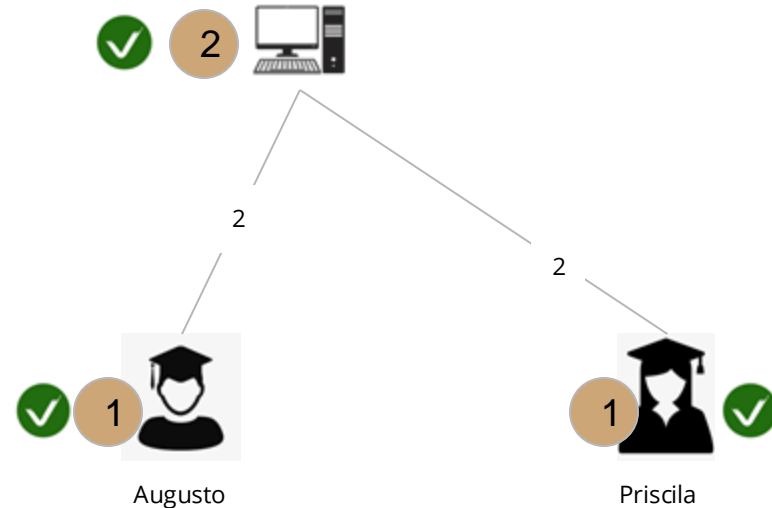


# Exemplo numérico

- Agora vamos avaliar um cenário onde **somente Augusto e Priscila** comparecem à uma aula prática:

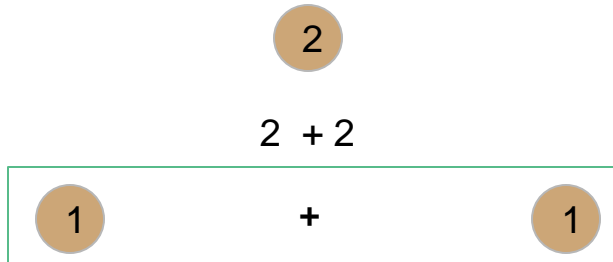
Score

???

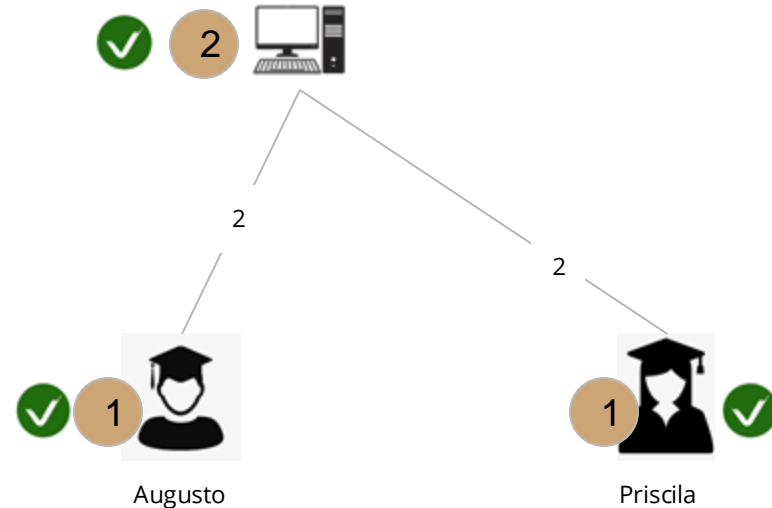


# Exemplo numérico

- Agora vamos avaliar um cenário onde **somente Augusto e Priscila** comparecem à uma aula prática:



Score  
???

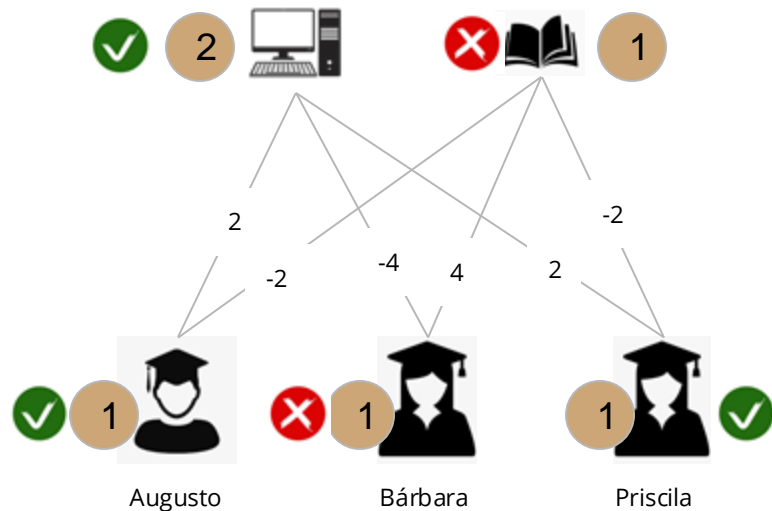


# Exemplo numérico

- Este é um cenário **mais provável de acontecer**, como vimos no dataset.
- Por conta disso, temos um **score mais alto para este cenário**.

Score

8



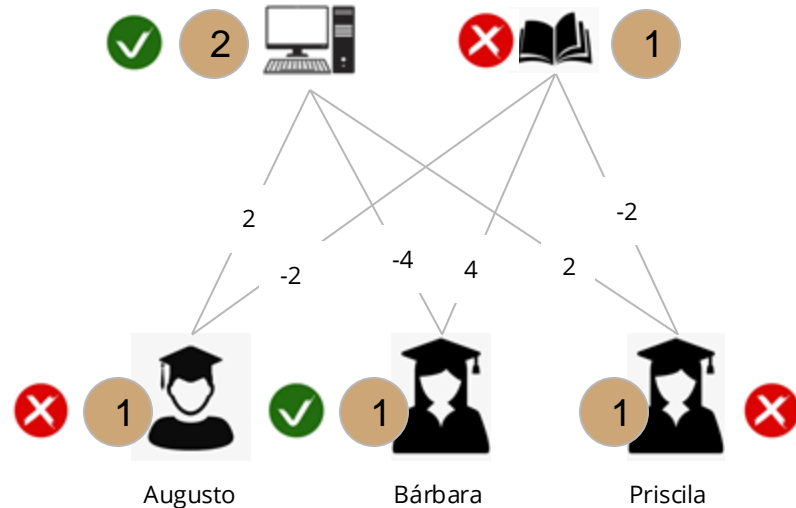


# Exemplo numérico

- Vamos verificar um cenário bastante **improvável**:
  - A Bárbara comparece em uma aula prática!

Score

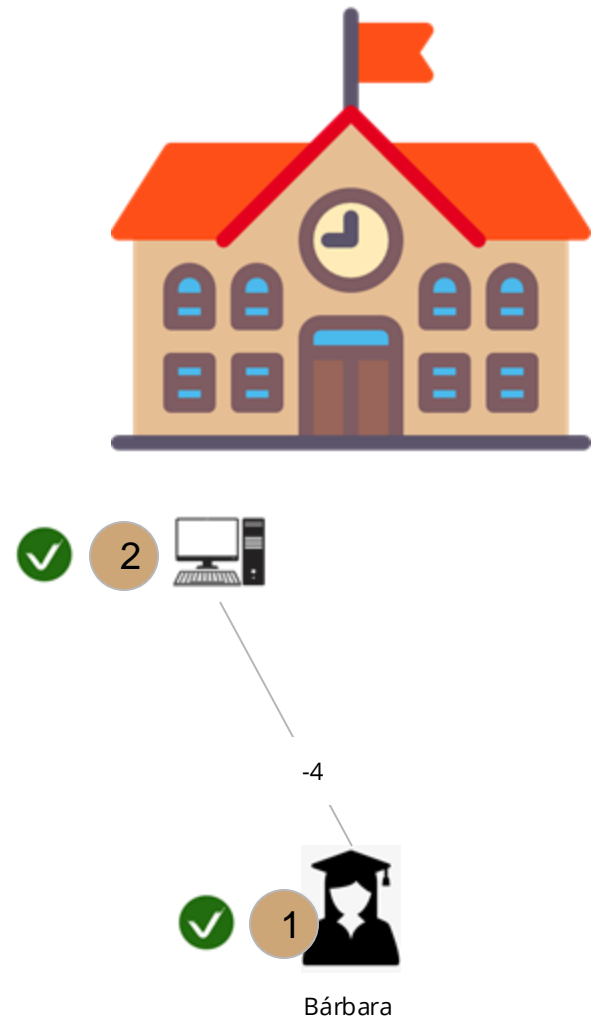
???



# Exemplo numérico

- Vamos verificar um cenário bastante **improvável**:
  - A Bárbara comparece em uma aula prática!

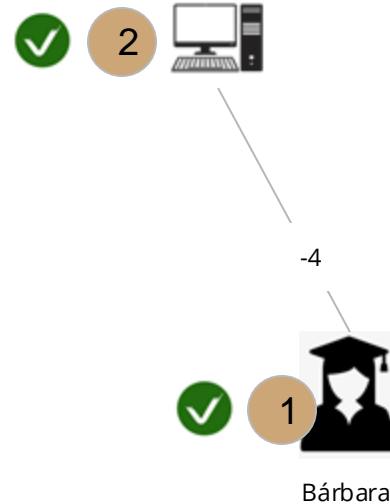
Score  
???



# Exemplo numérico

- Vamos verificar um cenário bastante **improvável**:
  - A Bárbara comparece em uma aula prática!

Score
-1

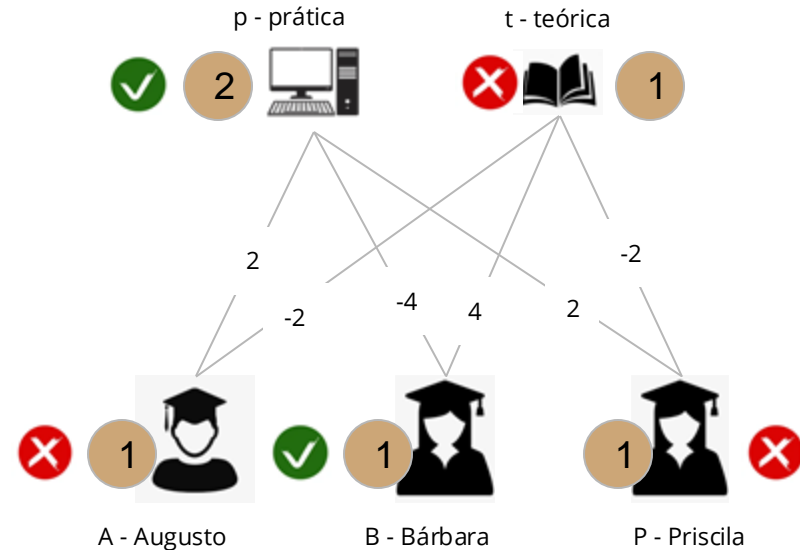


# Exemplo numérico

- Se fizermos isso para todos os 32 cenários possíveis, temos uma tabela:

Cenário	Score
ABP pt	6
AP p	8
B p	-1
A	1
A t	0
AB p	3
A p	5
BP p	2
...	

- A - Augusto
- B - Bárbara
- P - Priscila
- p - Prática
- t - Teórica

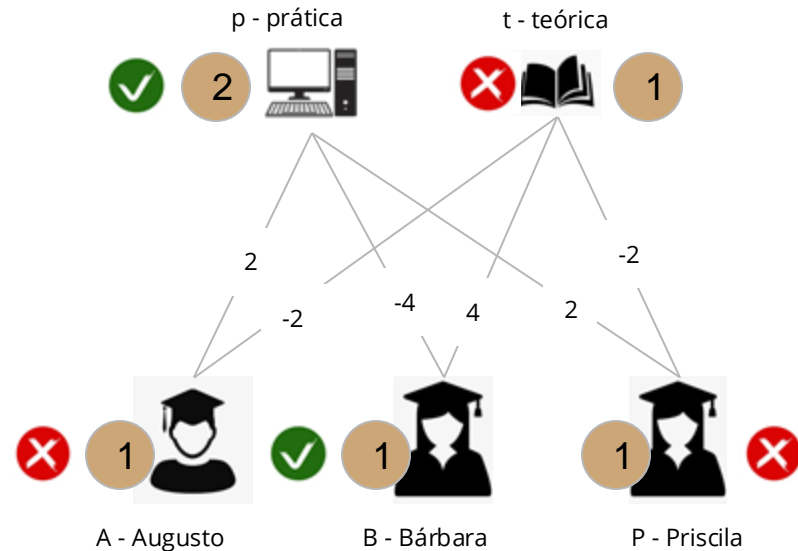


# Exemplo numérico

- Se fizermos isso para todos os 32 cenários possíveis, temos uma tabela:

Cenário	Score
ABP pt	6
AP p	8
B p	-1
A	1
A t	0
AB p	3
A p	5
BP p	2
...	

- A - Augusto
- B - Bárbara
- P - Priscila
- p - Prática
- t - Teórica

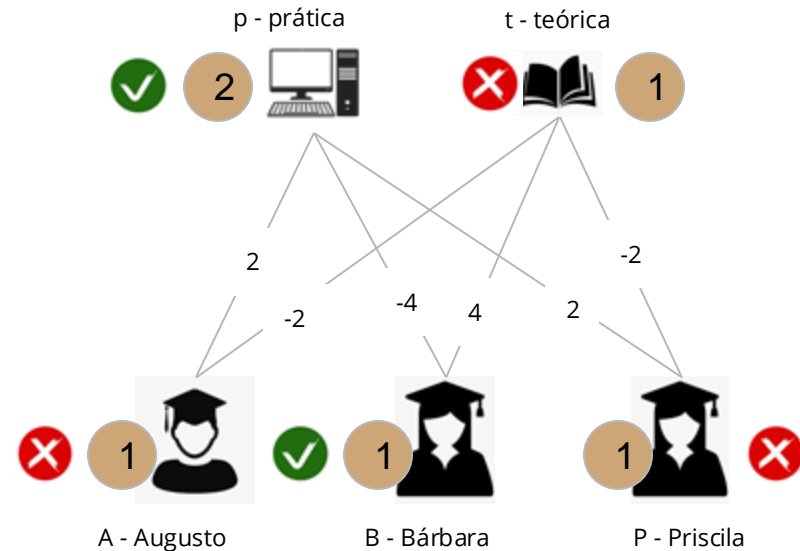


# Exemplo numérico

- Se fizermos isso para todos os 32 cenários possíveis, temos uma tabela:

Cenário	Score
ABP pt	6
AP p	8
B p	-1
A	1
A t	0
AB p	3
A p	5
BP p	2
...	

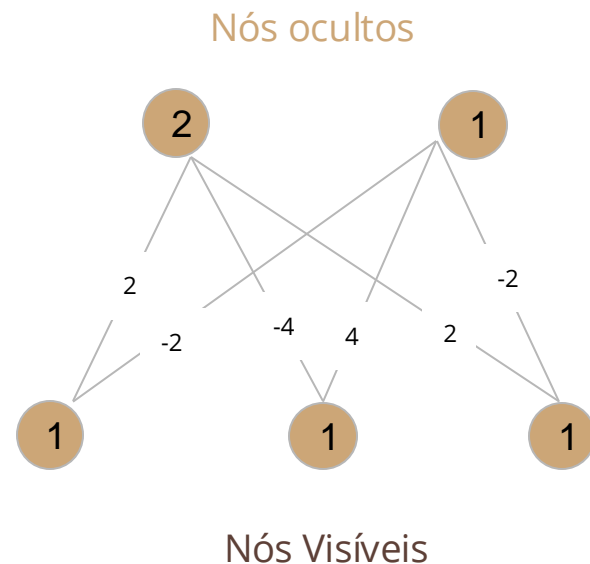
- A - Augusto
- B - Bárbara
- P - Priscila
- p - Prática
- t - Teórica



# Exemplo numérico

- Para uma RBM nós calculamos uma medida de **energia** através da seguinte fórmula:

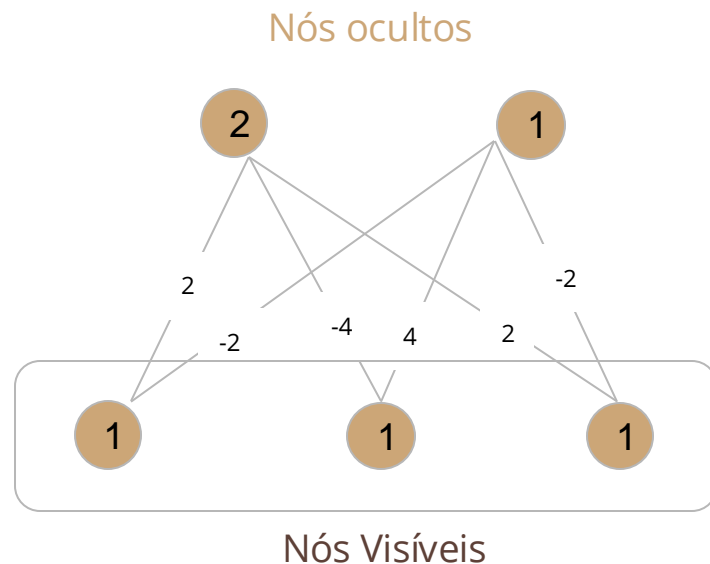
$$E(V, H) = - \sum_i aV_i - \sum_j bH_j - \sum_{i,j} V_i H_j w_{ij}$$



# Exemplo numérico

- Multiplicamos os pesos associados a **camada visível**:

$$E(V, H) = - \sum_i aV_i - \sum_j bH_j - \sum_{i,j} V_i H_j w_{ij}$$

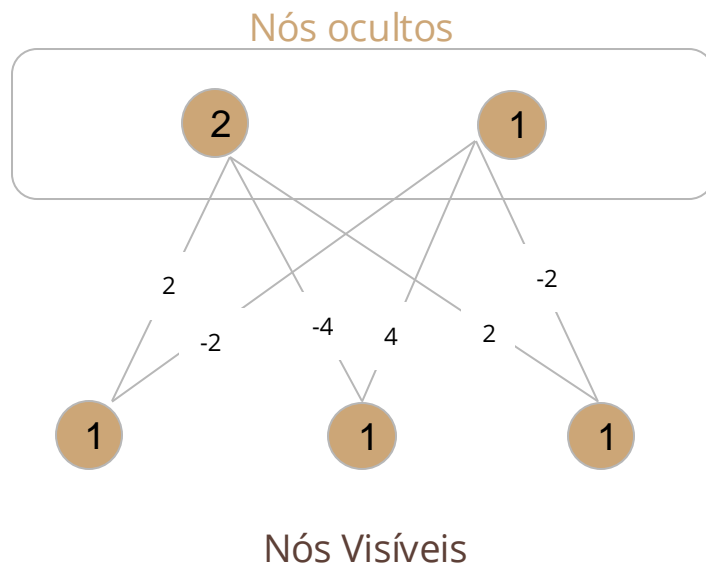




# Exemplo numérico

- Multiplicamos os pesos associados a **camada oculta**:

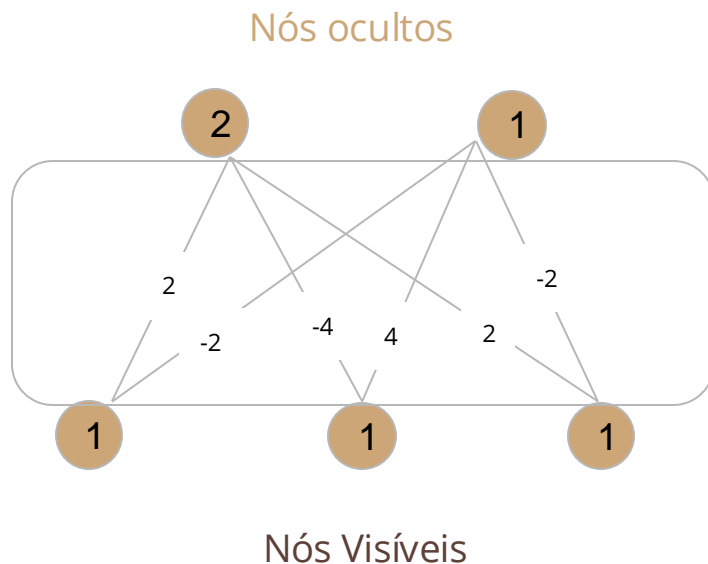
$$E(V, H) = - \sum_i aV_i - \underbrace{\sum_j bH_j}_{\text{camada oculta}} - \sum_{i,j} V_i H_j w_{ij}$$



# Exemplo numérico

- Multiplicamos os pesos associados a **camada visível e oculta com relação às conexões**:

$$E(V, H) = - \sum_i aV_i - \sum_j bH_j - \sum_{i,j} V_i H_j w_{ij}$$



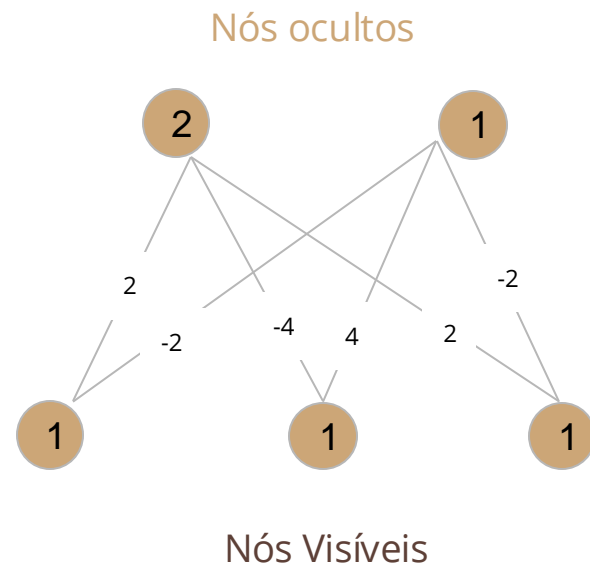
# Exemplo numérico

$$E(V, H) = - \sum_i aV_i - \sum_j bH_j - \sum_{i,j} V_i H_j w_{ij}$$

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$



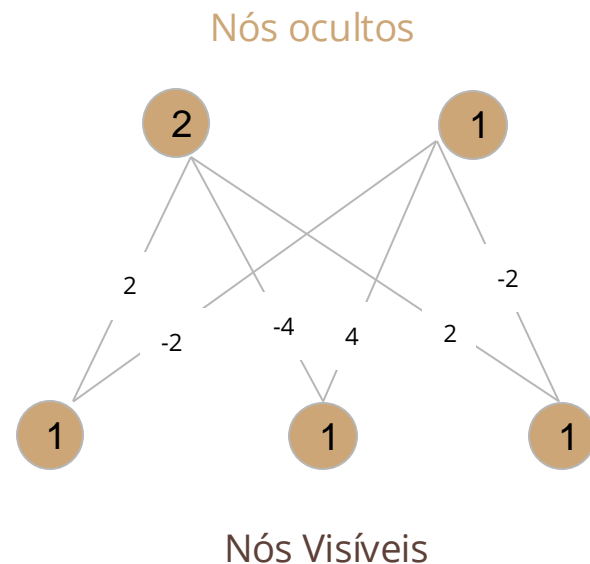
# Exemplo numérico

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$

Cenário	Score	$e^{\text{score}}$	
ABP pt	6		
AP p	8		
B p	-1		
A	1		
A t	0		
AB p	3		
A p	5		
BP p	2		



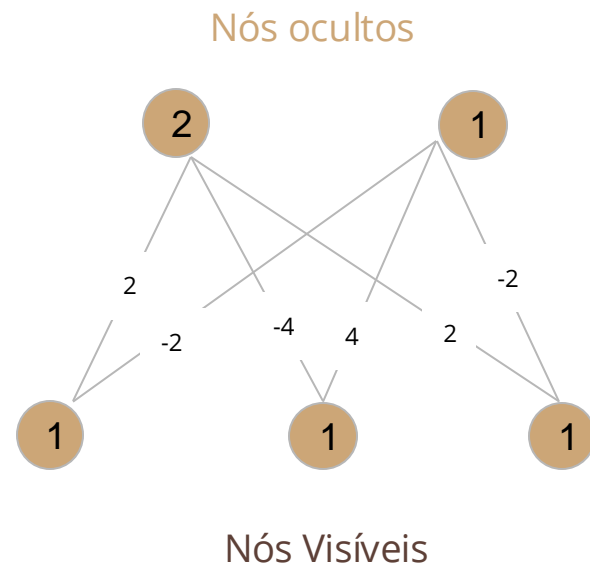
# Exemplo numérico

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$

Cenário	Score	$e^{\text{score}}$	
ABP pt	6	403	
AP p	8	2980	
B p	-1	0.36	
A	1	2.71	
A t	0	1	
AB p	3	20	
A p	5	148	
BP p	2	7.38	



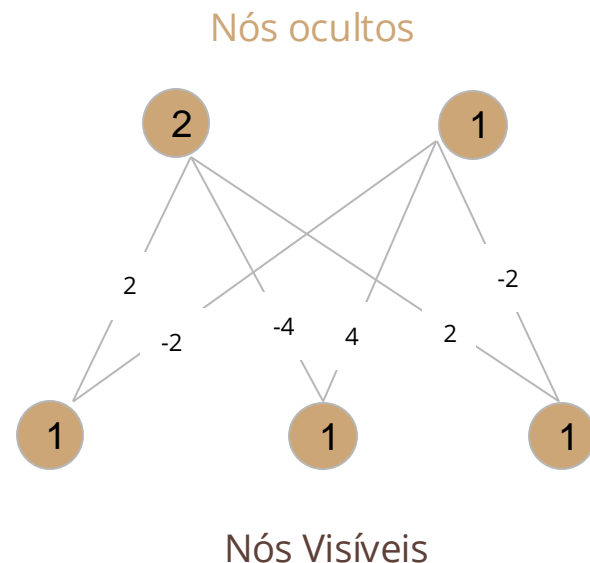
# Exemplo numérico

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	6	403	0.11
AP p	8	2980	0.84
B p	-1	0.36	0.00
A	1	2.71	0.00
A t	0	1	0.00
AB p	3	20	0.01
A p	5	148	0.04
BP p	2	7.38	0.00



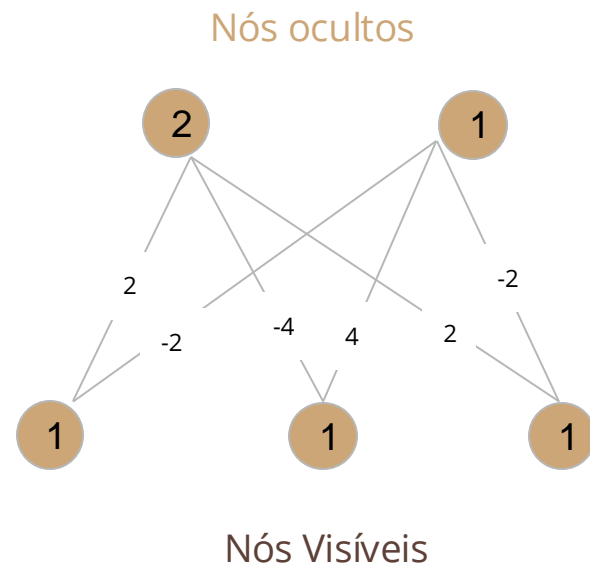
# Exemplo numérico

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	6	403	0.11
AP p	8	2980	0.84
B p	-1	0.36	0.00
A	1	2.71	0.00
A t	0	1	0.00
AB p	3	20	0.01
A p	5	148	0.04
BP p	2	7.38	0.00



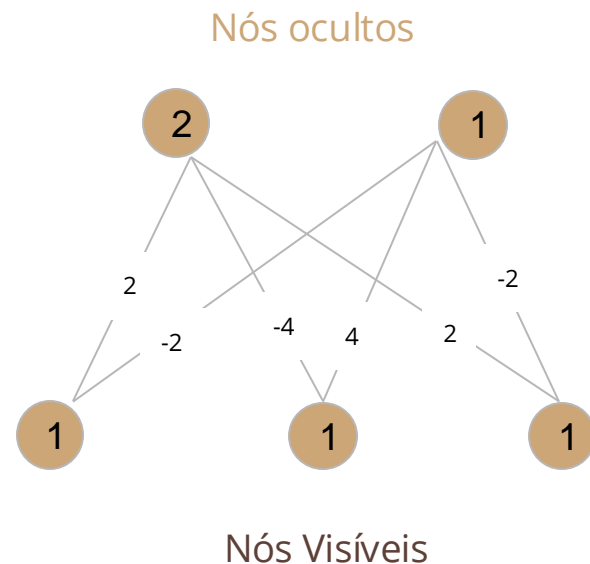
# Exemplo numérico

- Podemos fazer paralelo da **energia (score)** com relação às **probabilidades** de cada cenário.
- Vamos transformar este valor de energia em probabilidades:

$$p(V, H) = \frac{1}{Z} e^{-E(V, H)}$$

$$Z = \sum_{V, H} e^{-E(V, H)}$$

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	6	403	0.11
AP p	8	2980	0.84
B p	-1	0.36	0.00
A	1	2.71	0.00
A t	0	1	0.00
AB p	3	20	0.01
A p	5	148	0.04
BP p	2	7.38	0.00

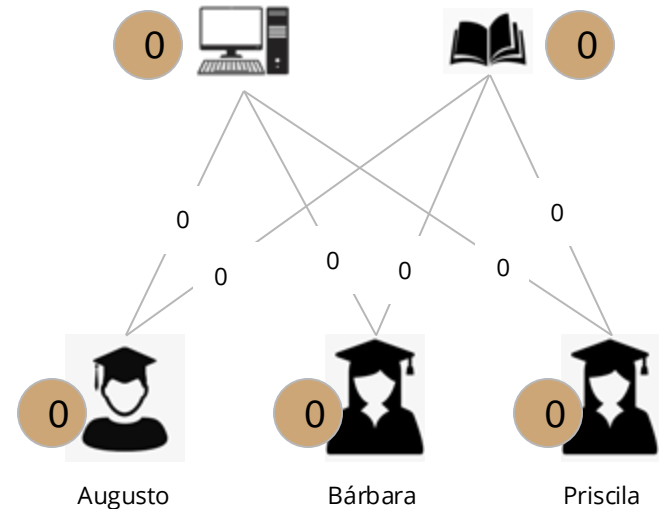




Como treinar uma Máquina Restrita de Boltzmann?

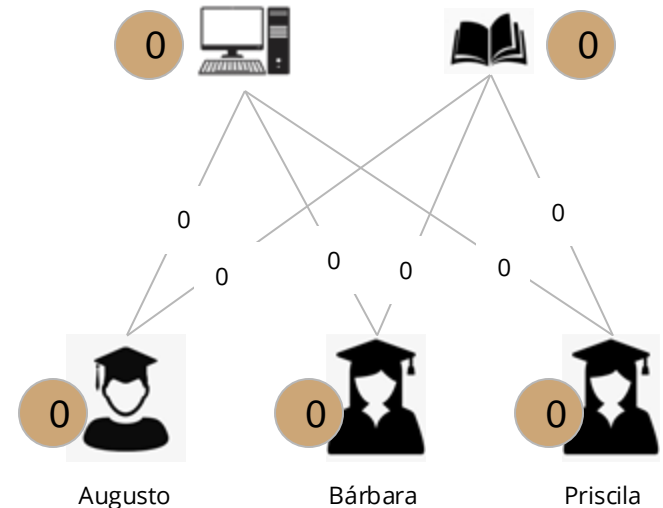
# Como treinar uma Máquina Restrita de Boltzmann?

- No treinamento nos queremos encontrar os **pesos** associados aos neurônios na **camada de entrada (visível)**, na **camada oculta** e nas **conexões entre os neurônios**



# Como treinar uma Máquina Restrita de Boltzmann?

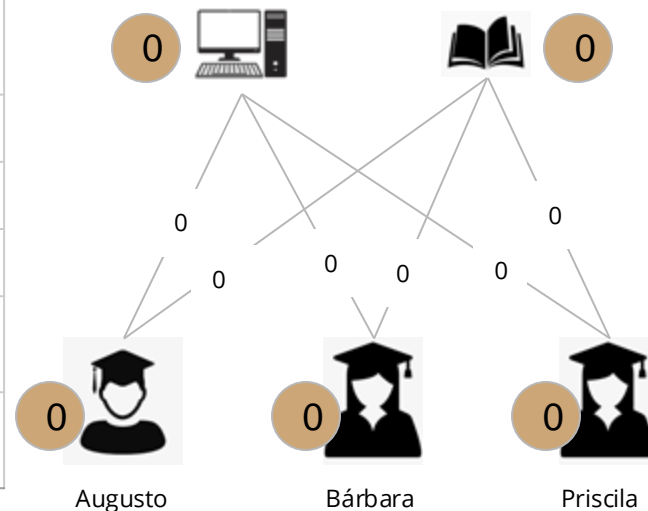
- Qual o número de possibilidades de estados que temos nesta rede com 5 neurônios na camada visível e 2 neurônios na camada oculta?
  - $2^5$  possíveis cenários = 32



# Como treinar uma Máquina Restrita de Boltzmann?

Vamos avaliar as probabilidades de cada cenário quando a rede está inicializada com pesos iguais a 0:

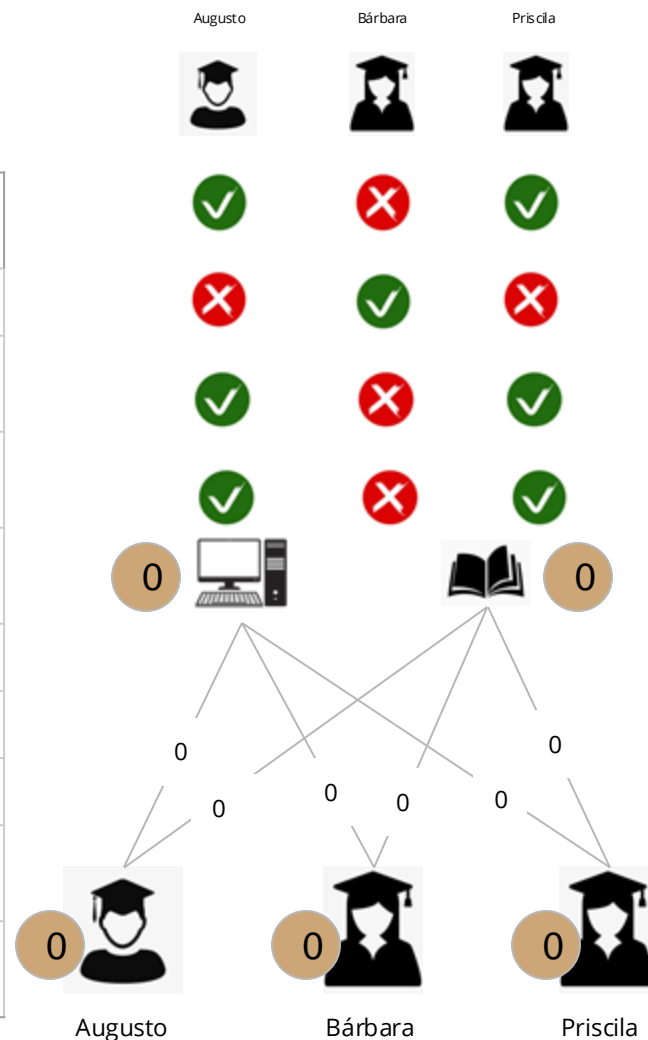
Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	0	1	0.03
AP p	0	1	0.03
B p	0	1	0.03
A	0	1	0.03
A t	0	1	0.03
AB p	0	1	0.03
A p	0	1	0.03
BP p	0	1	0.03
...			



# Como treinar uma Máquina Restrita de Boltzmann?

No entanto, o que nós gostaríamos é que os pesos refletissem as probabilidades presentes no nosso dataset:

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	0	1	0.03
AP p	0	1	0.03
B p	0	1	0.03
A	0	1	0.03
A t	0	1	0.03
AB p	0	1	0.03
A p	0	1	0.03
BP p	0	1	0.03
...			

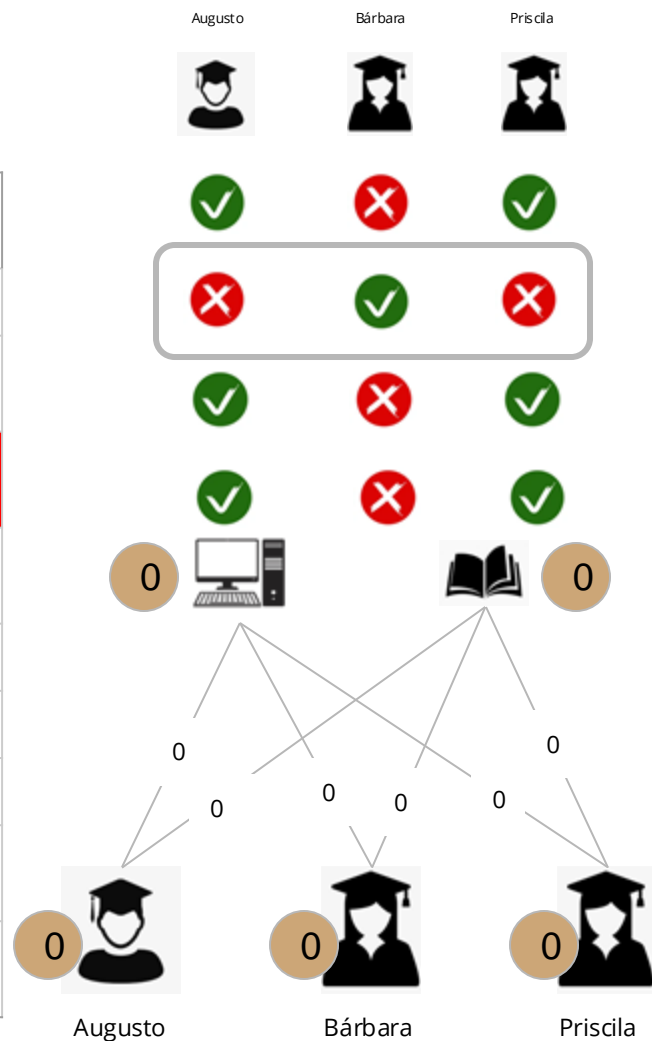


# Como treinar uma Máquina Restrita de Boltzmann?

Gostaríamos que as situações onde só a Bárbara aparecem tivessem maior probabilidade:



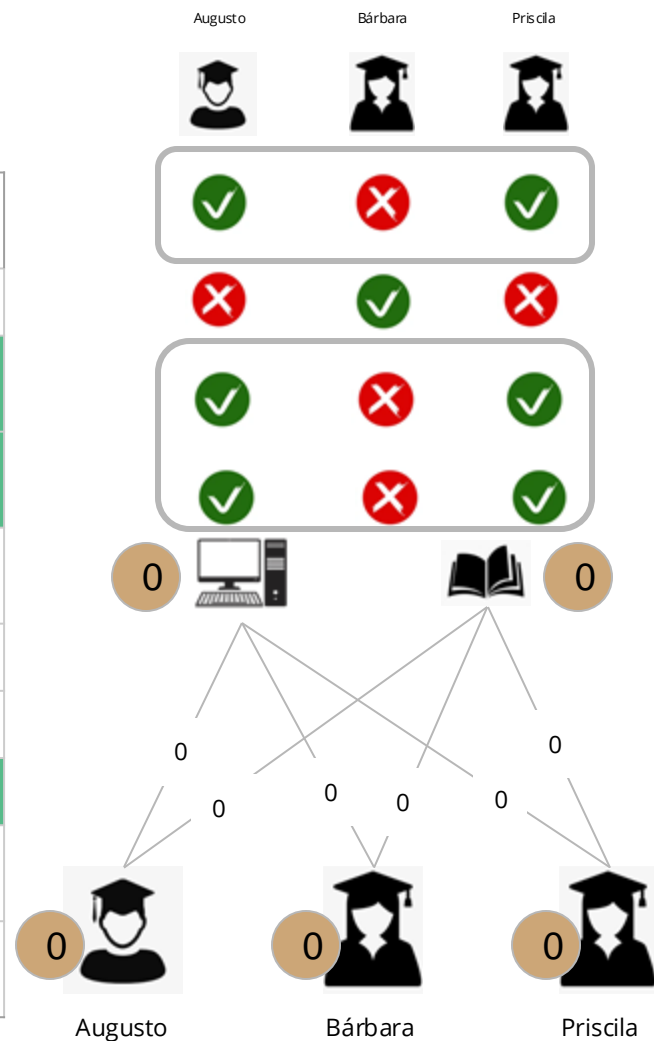
Cenário	Score	e <sup>score</sup>	p(V, H)
ABP pt	0	1	0.03
AP p	0	1	0.03
<b>B p</b>	<b>0</b>	<b>1</b>	<b>0.03</b>
A	0	1	0.03
A t	0	1	0.03
AB p	0	1	0.03
A p	0	1	0.03
BP p	0	1	0.03
...			



# Como treinar uma Máquina Restrita de Boltzmann?

Situações onde a Priscila e o Augusto aparecem juntos:

Cenário	Score	e <sup>score</sup>	p(V, H)
ABP pt	0	1	0.03
AP p	0	1	0.03
B p	0	1	0.03
A	0	1	0.03
A t	0	1	0.03
AB p	0	1	0.03
A p	0	1	0.03
BP p	0	1	0.03
...			



# Como treinar uma Máquina Restrita de Boltzmann?

Para treinar esta rede utilizamos o método chamado ***Contrastive Divergence***.

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	0	1	0.03
AP p	0	1	0.03
B p	0	1	0.03
A	0	1	0.03
A t	0	1	0.03
AB p	0	1	0.03
A p	0	1	0.03
BP p	0	1	0.03
...			

Augusto



Bárbara



Priscila





# Como treinar uma Máquina Restrita de Boltzmann?

- Neste método, selecionamos uma amostra do dataset. Por exemplo, selecionamos uma amostra onde **Antônio e Priscila** comparecem a aula.
- Então **aumentamos a probabilidade deste evento**.
- **Diminuímos de todos os demais**

Cenário	Score	e <sup>score</sup>	p(V, H)
ABP pt	-0.1	0.90	0.03
AP p	0.1	1.11	0.04
B p	-0.1	0.90	0.03
A	-0.1	0.90	0.03
A t	-0.1	0.90	0.03
AB p	-0.1	0.90	0.03
A p	-0.1	0.90	0.03
BP p	-0.1	0.90	0.03
...			

Augusto



Bárbara



Priscila



# Como treinar uma Máquina Restrita de Boltzmann?

- Seleccionamos o próximo evento e repetimos a operação:

Cenário	Score	e <sup>score</sup>	p(V, H)
ABP pt	-0.2	0.90	0.03
AP p	0	1.11	0.04
B p	0	1	0.04
A	-0.2	0.90	0.03
A t	-0.2	0.90	0.03
AB p	-0.2	0.90	0.03
A p	-0.2	0.90	0.03
BP p	-0.2	0.90	0.03
...			

Augusto

Bárbara

Priscila



# Como treinar uma Máquina Restrita de Boltzmann?

- À medida que fazemos isso para todo o dataset, a tendência é que as probabilidades reflitam os cenários do dataset de treinamento.

Cenário	Score	$e^{\text{score}}$	$p(V, H)$
ABP pt	-0.2	0.90	0.03
AP p	0	1.11	0.04
B p	0	1	0.04
A	-0.2	0.90	0.03
A t	-0.2	0.90	0.03
AB p	-0.2	0.90	0.03
A p	-0.2	0.90	0.03
BP p	-0.2	0.90	0.03
...			

Augusto



Bárbara



Priscila



# Como treinar uma Máquina Restrita de Boltzmann?

- Existe um problema de lidarmos com o ajuste das probabilidades desta forma. O problema está no número de eventos possíveis:
  - $2^5 = 32$
- O número de possibilidades facilmente se torna proibitivo. Se tivermos 25 neurônios:
  - $2^{25} = 33.554.432$

Cenário	Score	e <sup>score</sup>	p(V, H)
ABP pt	-0.2	0.90	0.03
AP p	0	1.11	0.04
B p	0	1	0.04
A	-0.2	0.90	0.03
A t	-0.2	0.90	0.03
AB p	-0.2	0.90	0.03
A p	-0.2	0.90	0.03
BP p	-0.2	0.90	0.03
...			

Augusto

Bárbara

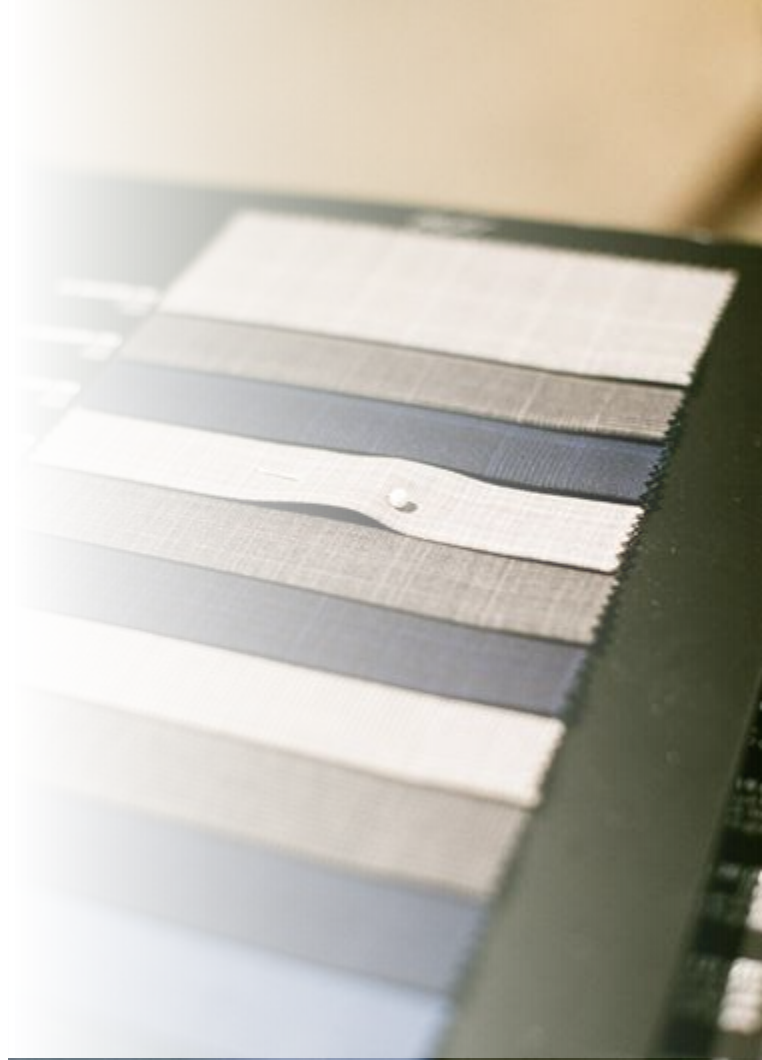
Priscila



# Amostragem de Gibbs

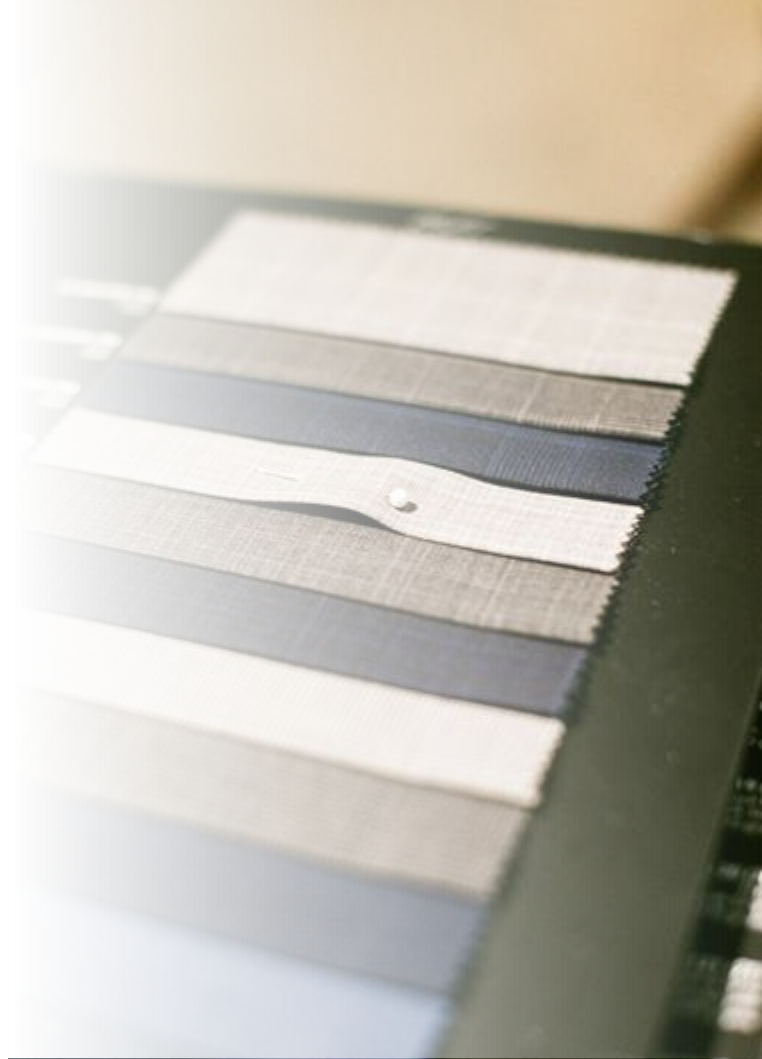
# Amostragem de Gibbs

- A amostragem de Gibbs constrói uma cadeia de Markov para a qual os valores convergem para uma determinada distribuição.
- Nós precisamos desta abordagem por que queremos que a seleção dos cenários possíveis seja feita respeitando a probabilidade dos eventos



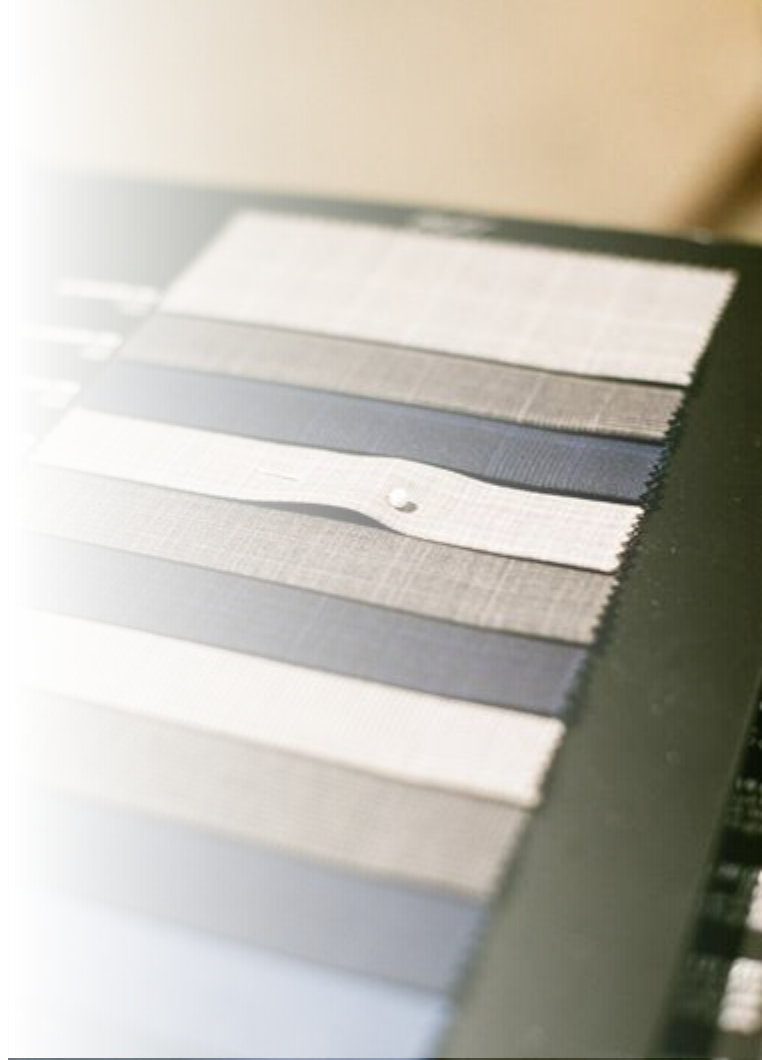
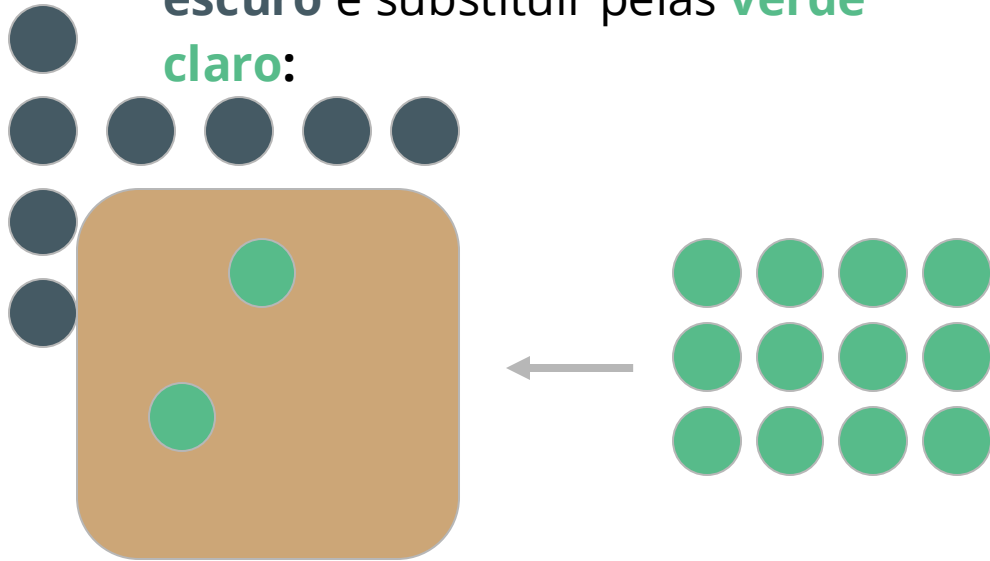
# Como funciona?

- Vamos imaginar que nós temos uma caixa onde a maioria das amostras são **verde escuro** e algumas **verde claro**:



# Como funciona?

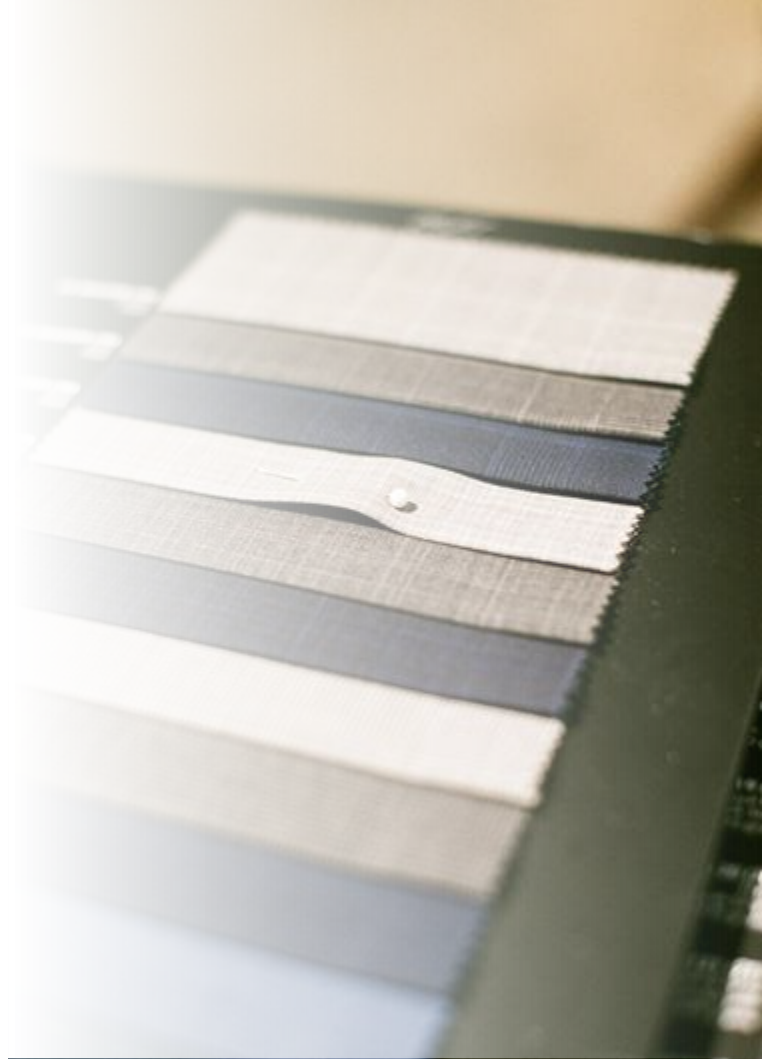
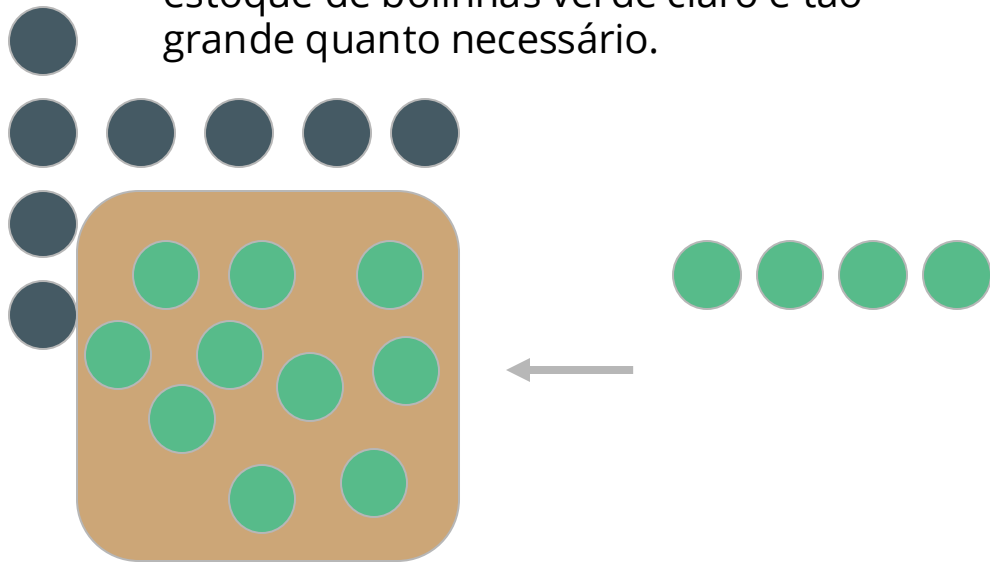
- Agora nós gostaríamos de remover todas as esferas **verde escuro** e substituir pelas **verde claro**:





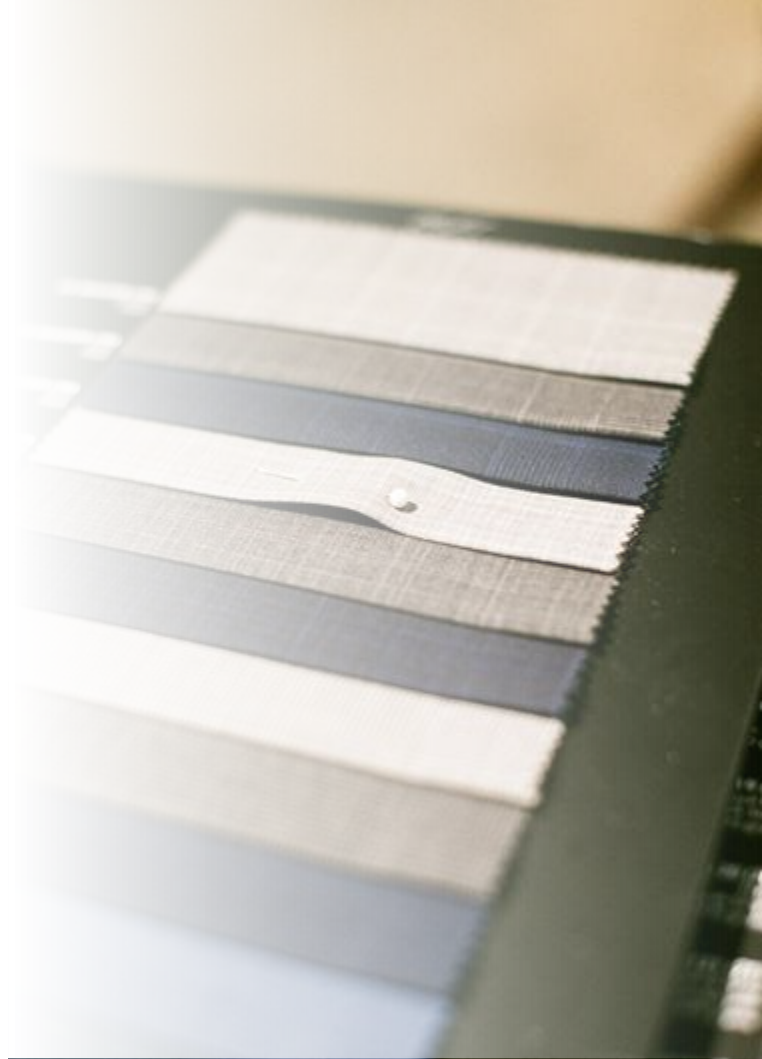
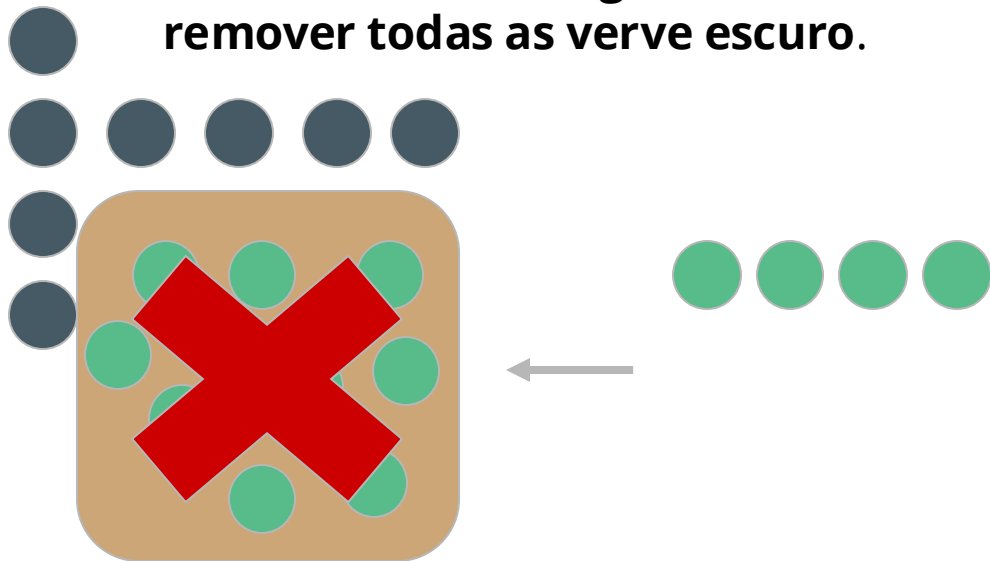
# Como funciona?

- Nós poderíamos simplesmente remover todas as esferas **verde escuro** e substituir pelas **verde claro**. Nosso estoque de bolinhas verde claro é tão grande quanto necessário.



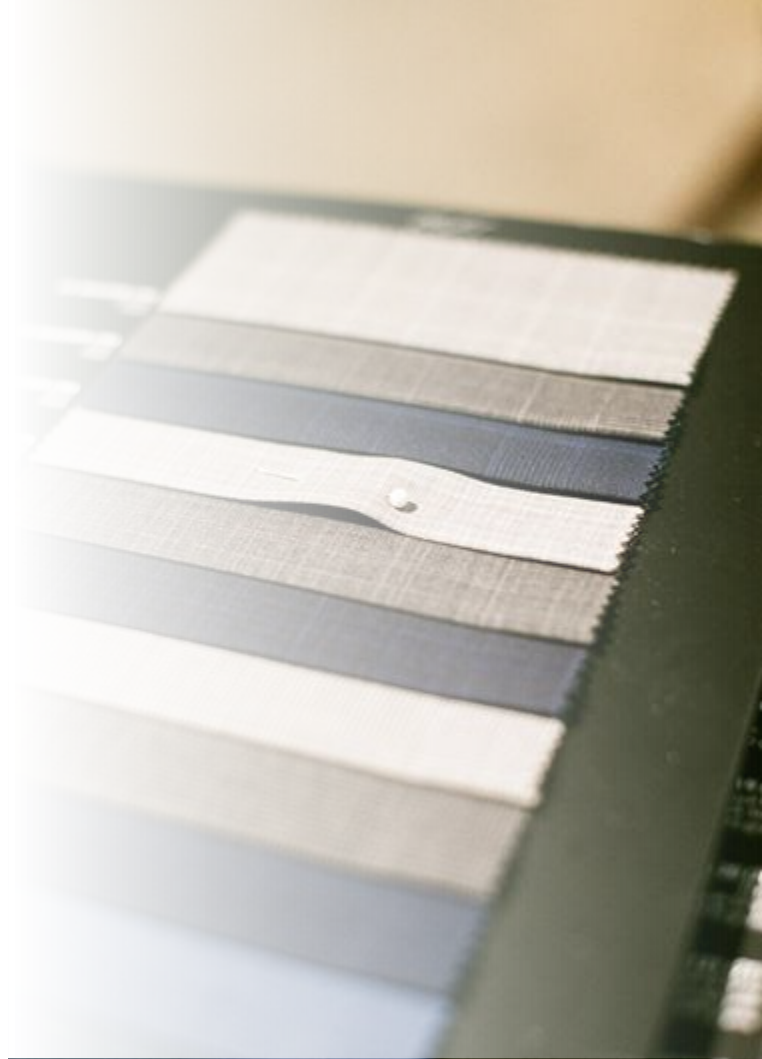
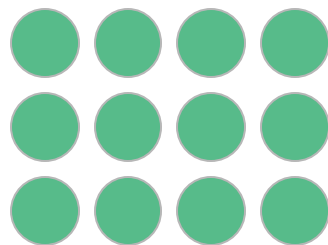
# Como funciona?

- Um ponto importante é que não conseguimos ver dentro da caixa. Portanto, **não conseguimos remover todas as verve escuro.**



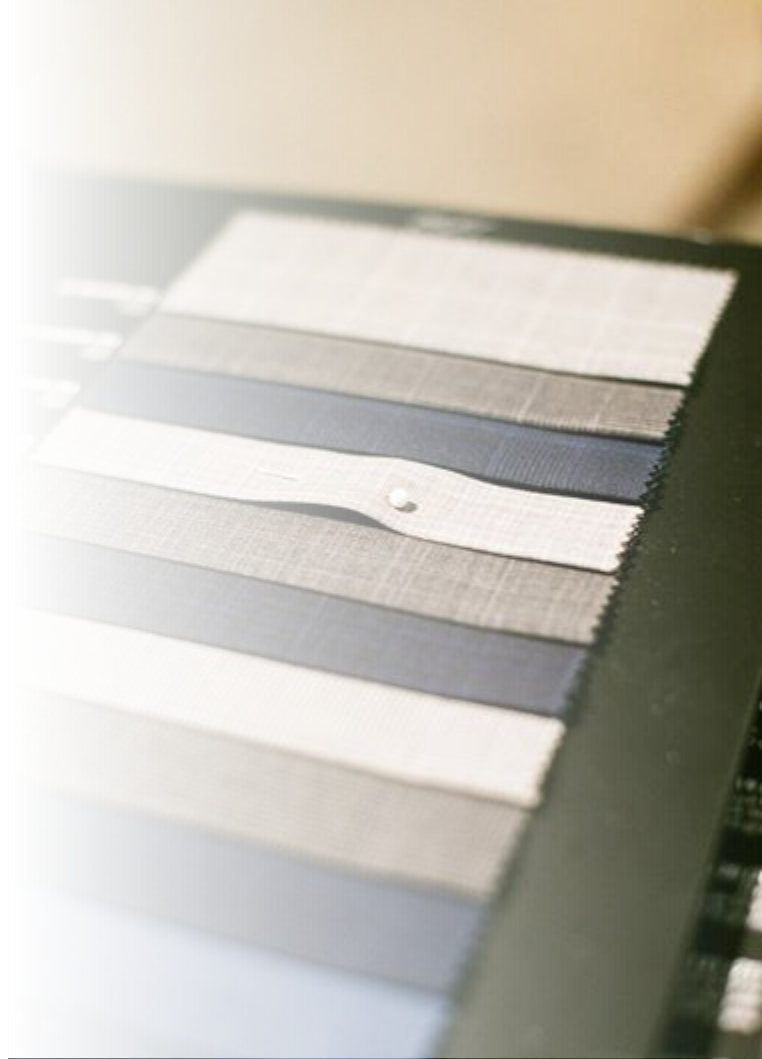
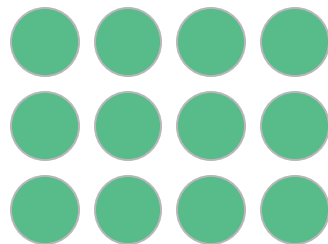
# Como funciona?

- Por outro lado, podemos remover seguidamente uma bolinha da caixa e substituir por uma verde claro.
- Nós temos que manter sempre o mesmo número de amostras dentro da caixa. Para cada bolinha que removemos, colocamos uma amostra verde claro no lugar.



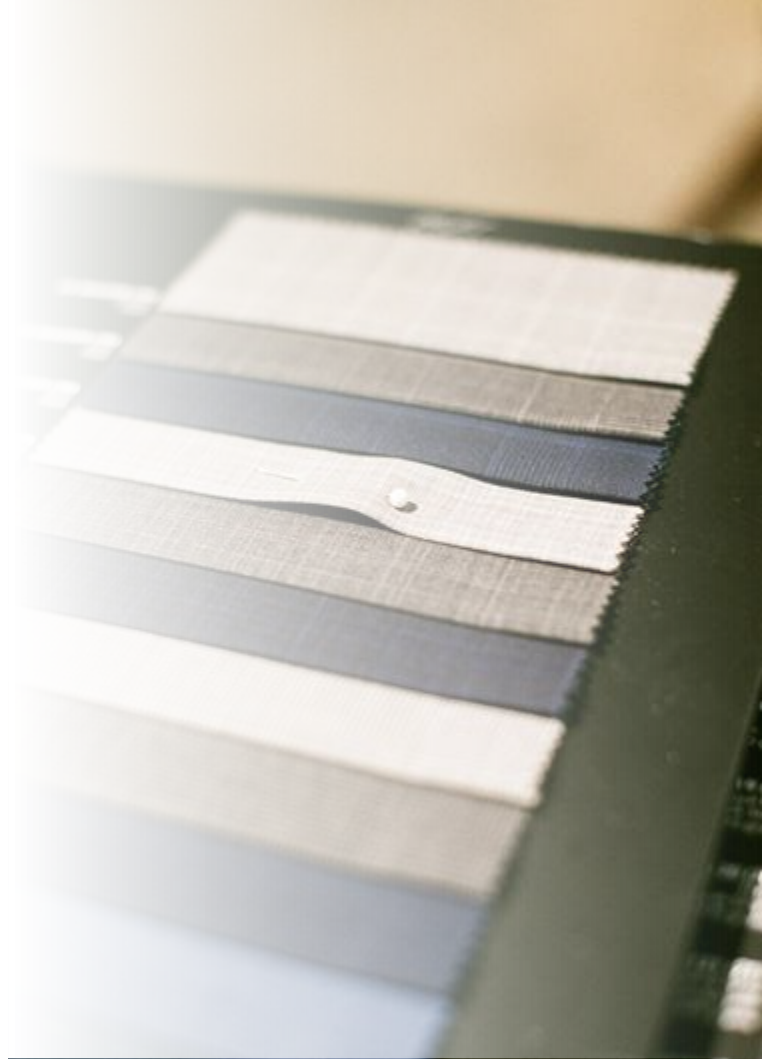
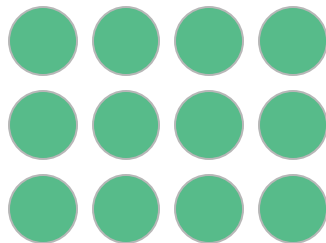
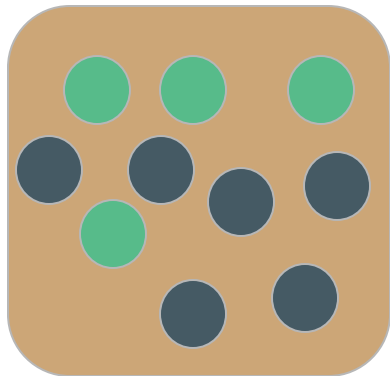
# Como funciona?

- Por outro lado, podemos remover seguidamente uma bolinha da caixa e substituir por uma verde claro.
- Nós temos que manter sempre o mesmo número de amostras dentro da caixa. Para cada bolinha que removemos, colocamos uma amostra verde claro no lugar.



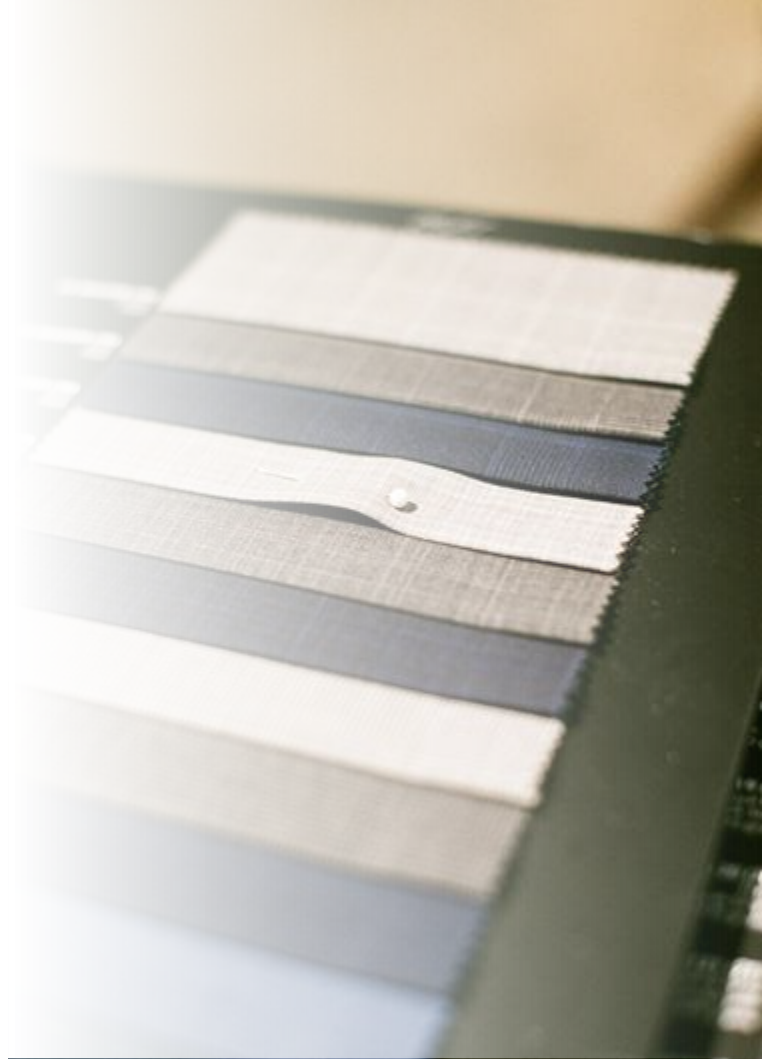
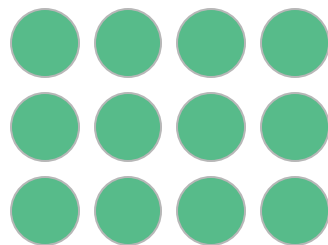
# Como funciona?

- Por outro lado, podemos remover seguidamente uma bolinha da caixa e substituir por uma verde claro.
- Nós temos que manter sempre o mesmo número de amostras dentro da caixa. Para cada bolinha que removemos, colocamos uma amostra verde claro no lugar.



# Como funciona?

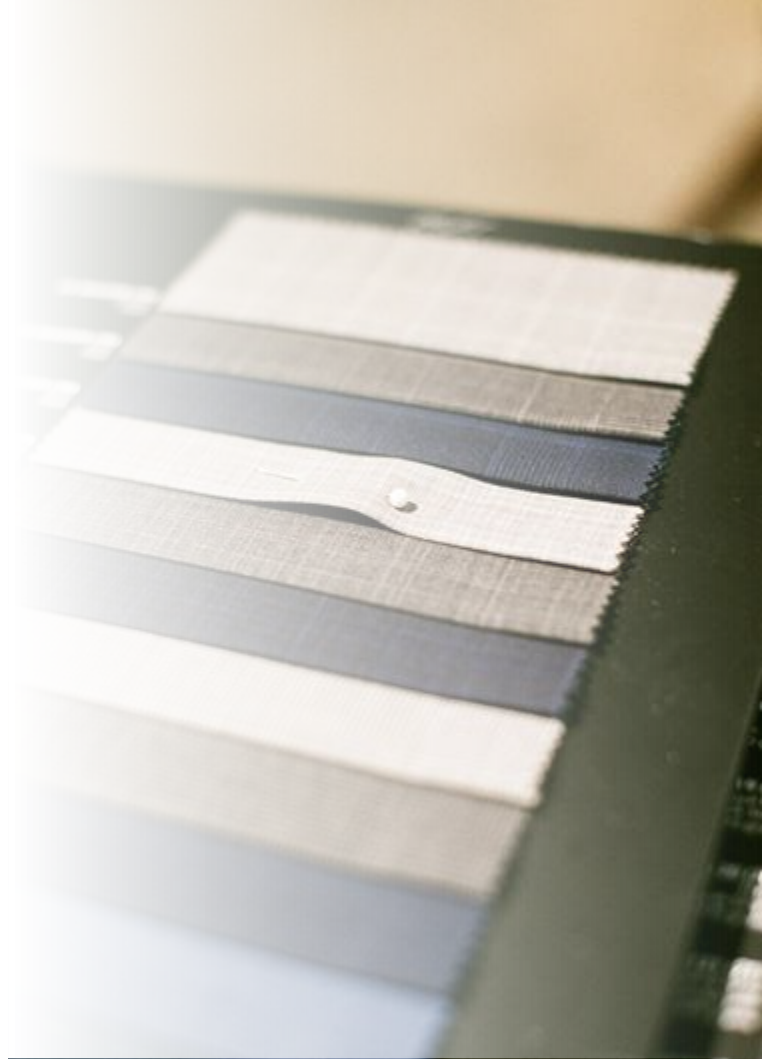
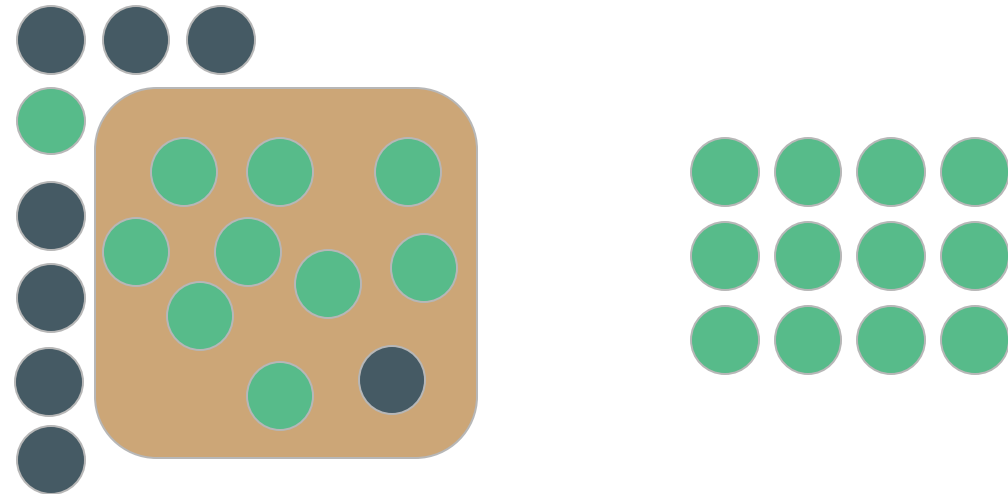
- Por outro lado, podemos remover seguidamente uma bolinha da caixa e substituir por uma verde claro.
- Nós temos que manter sempre o mesmo número de amostras dentro da caixa. Para cada bolinha que removemos, colocamos uma amostra verde clara no lugar.





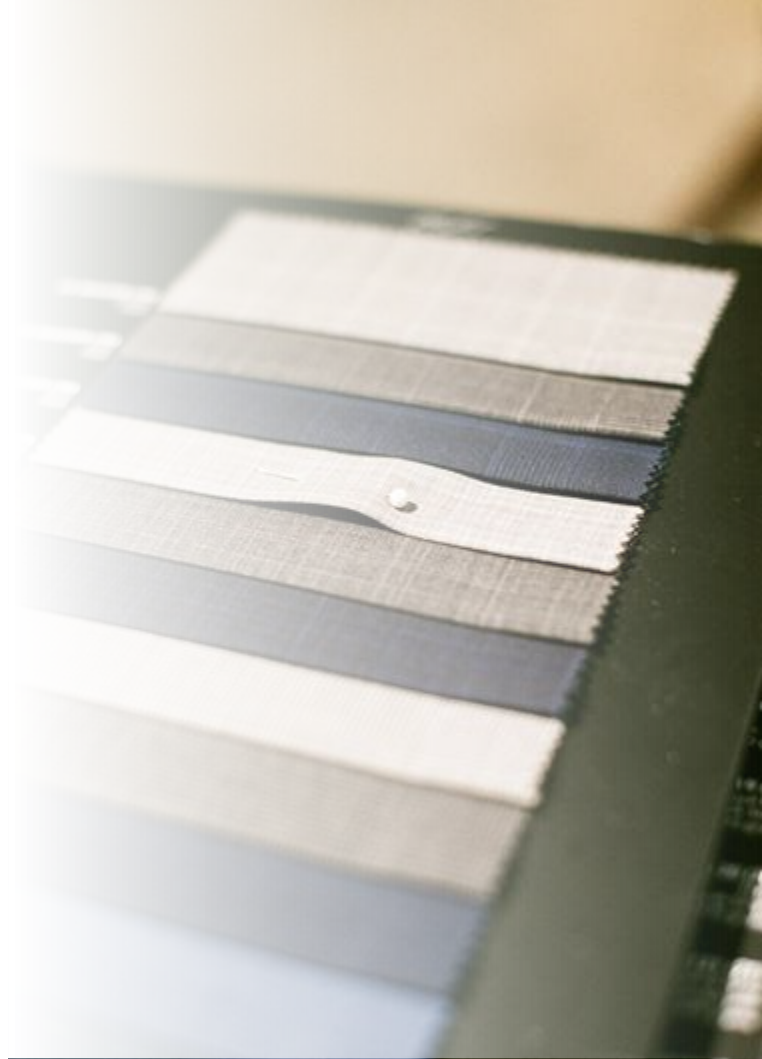
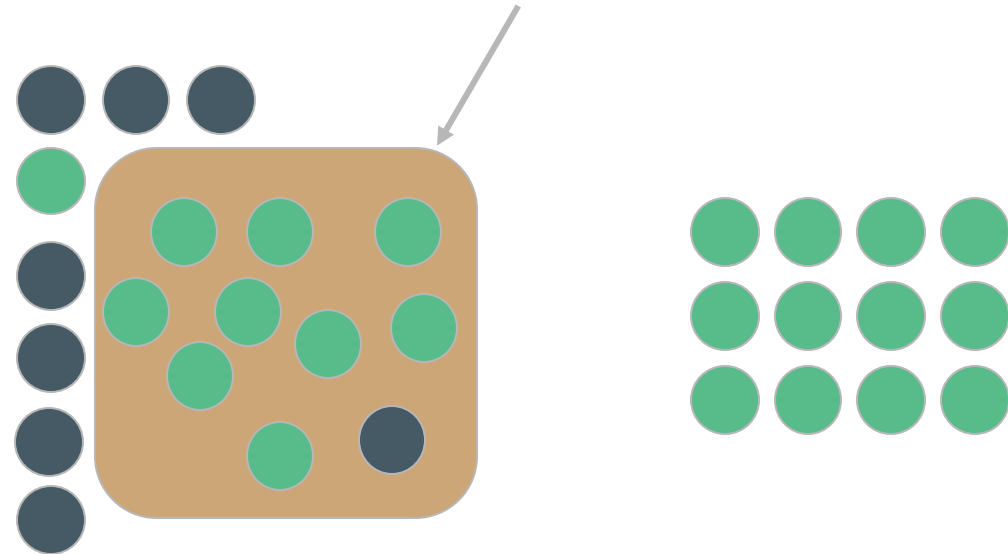
# Como funciona?

- A cada passo que removemos uma amostra da caixa e colocamos outra verde claro no lugar **estamos mais próximos do objetivo ou, pelo menos, no mesmo estado anterior.**



# Como funciona?

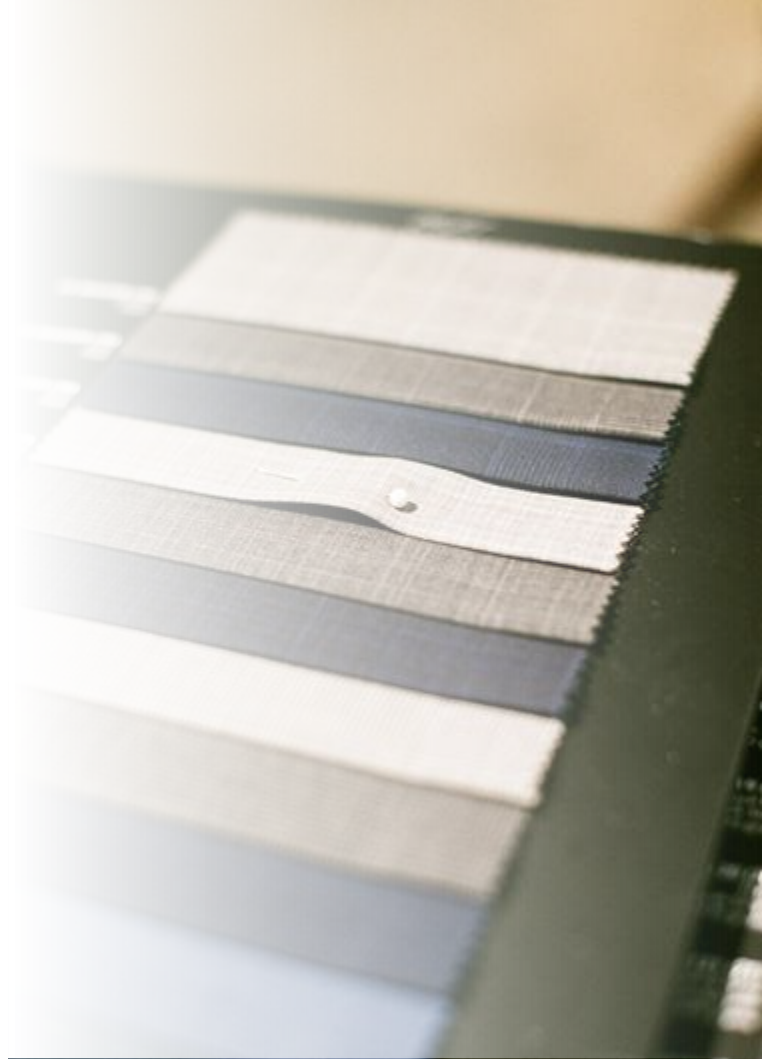
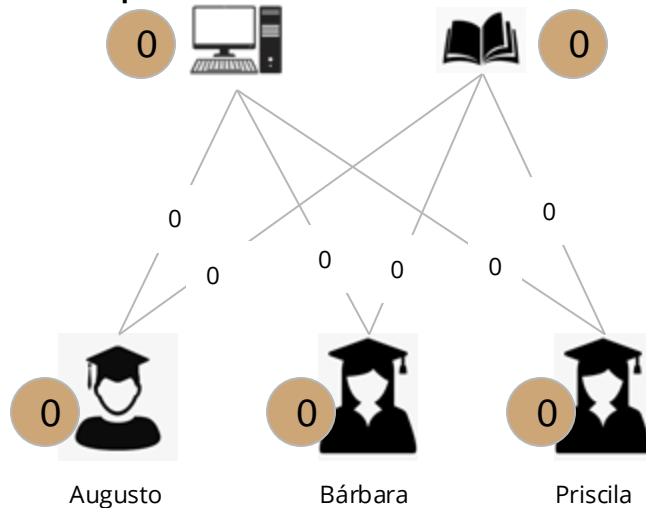
- Ao final de muitas repetições estaremos com quase todas as esferas verdes claro na caixa.





# Como funciona?

- Este é o processo que iremos utilizar para selecionar uma amostra dentre todo universo de cenários possíveis

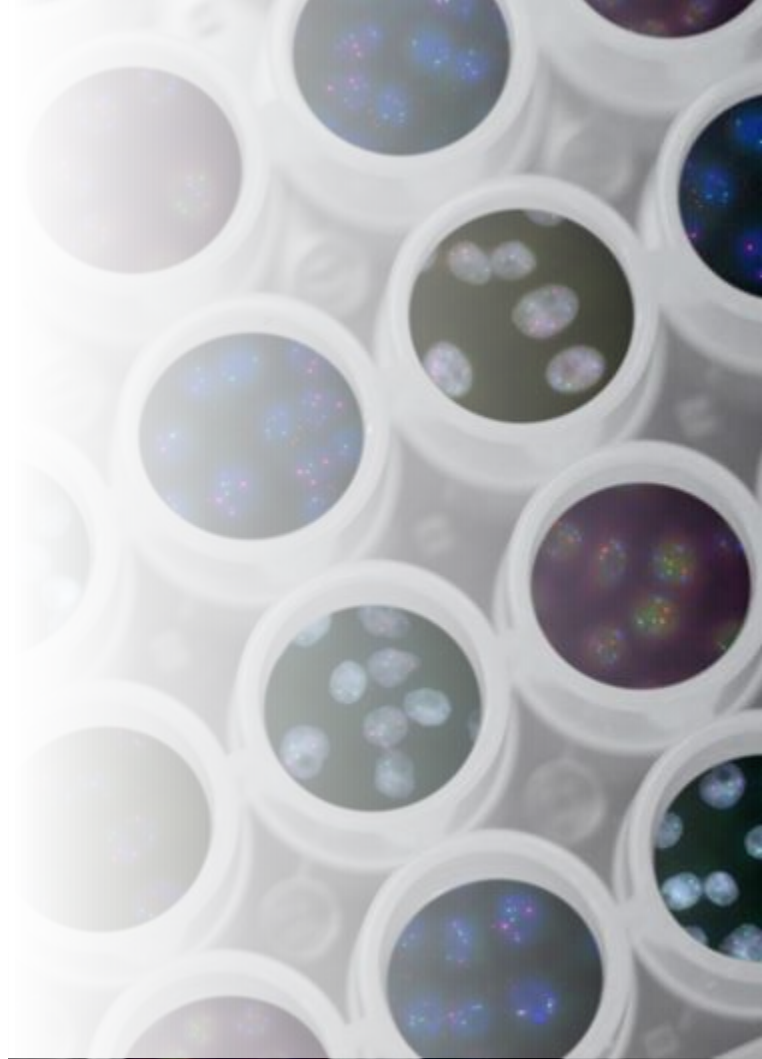
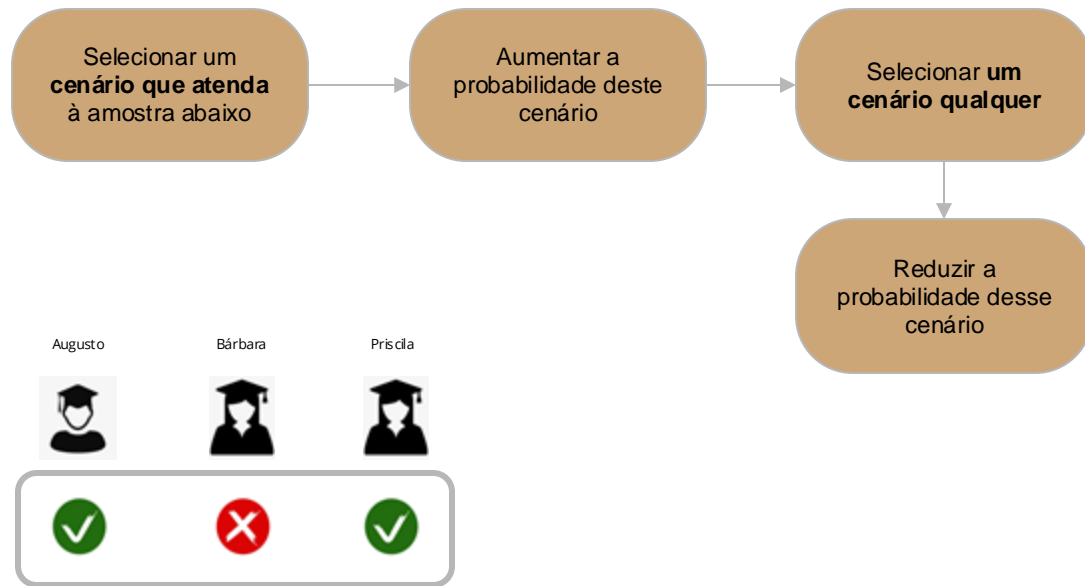




# Treinamento da RBM com amostragem de Gibbs



# Treinamento da RBM com amostragem de Gibbs



# Treinamento da RBM com amostragem de Gibbs

Remover uma amostra da caixa

Selecionar um  
**cenário que atenda**  
à amostra abaixo

Aumentar a  
probabilidade deste  
cenário

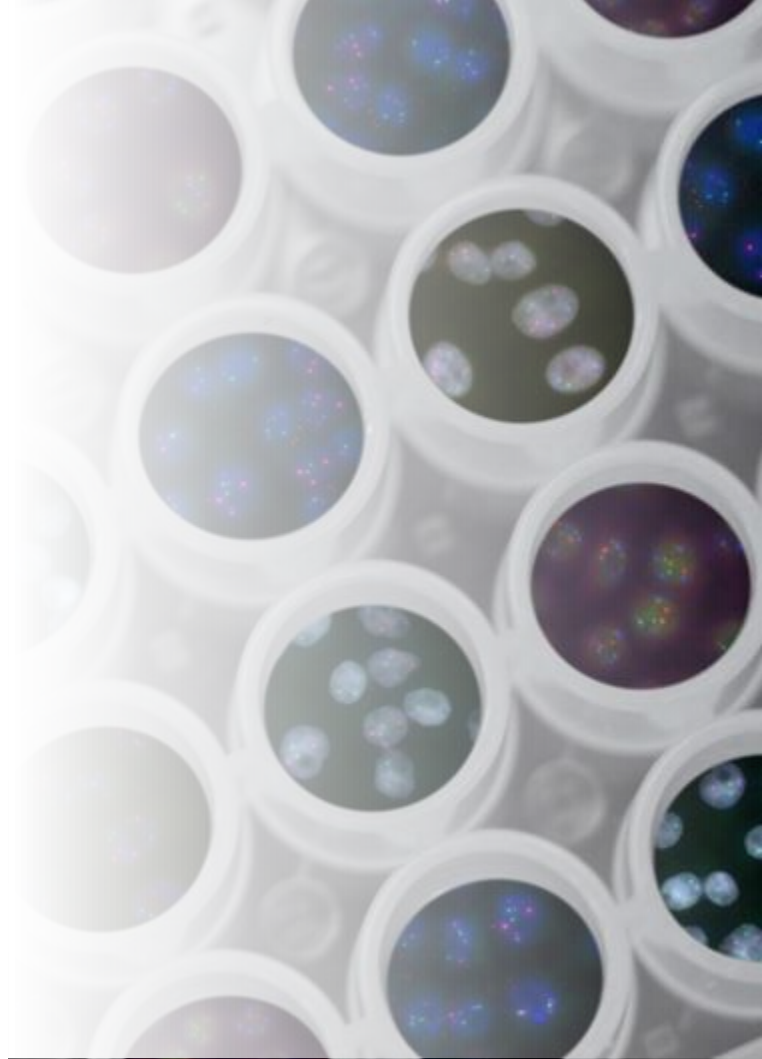
Selecionar um  
**cenário qualquer**

Reduzir a  
probabilidade desse  
cenário

Augusto

Bárbara

Priscila



# Treinamento da RBM com amostragem de Gibbs

Remover uma amostra da caixa

Selecionar um **cenário que atenda** à amostra abaixo

Aumentar a probabilidade deste cenário

Inserir uma **amostra verde claro** na caixa

Selecionar **um cenário qualquer**

Reduzir a probabilidade desse cenário

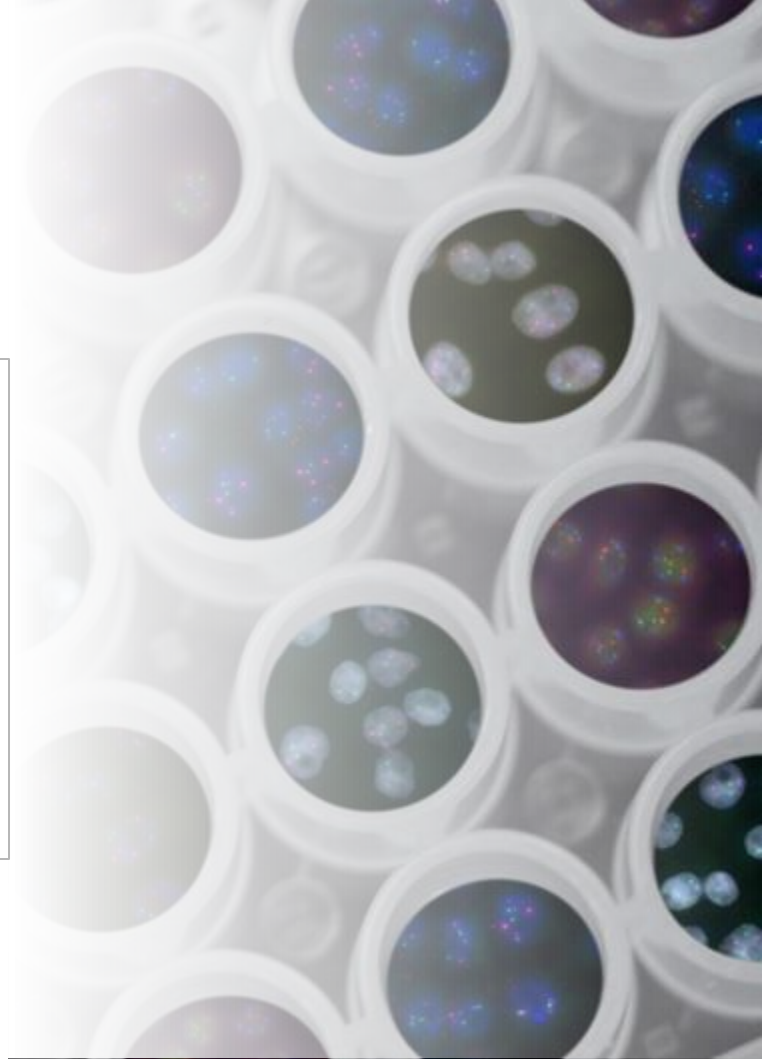
Augusto



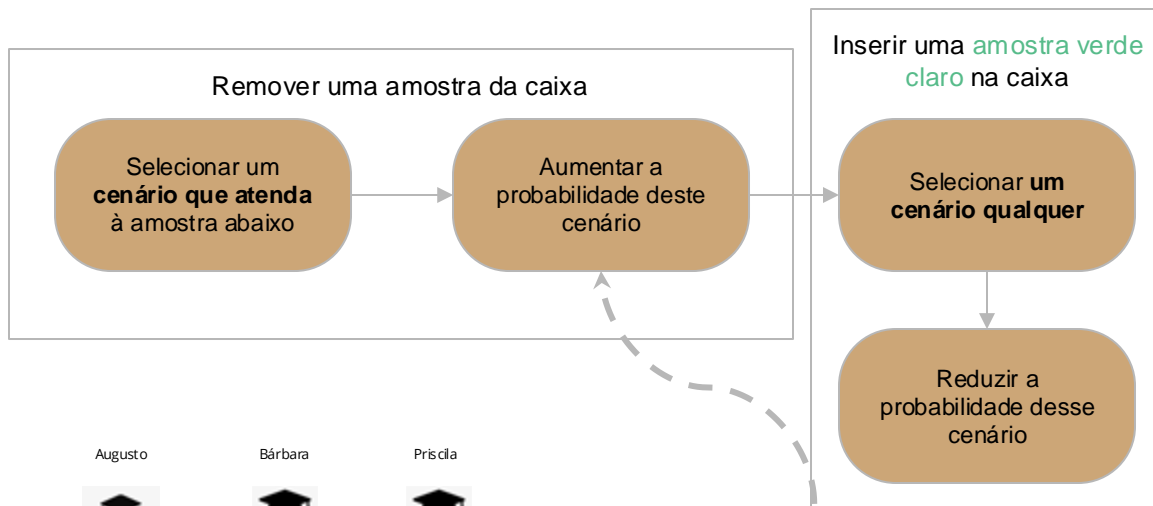
Bárbara



Priscila



# Treinamento da RBM com amostragem de Gibbs



Augusto



Bárbara

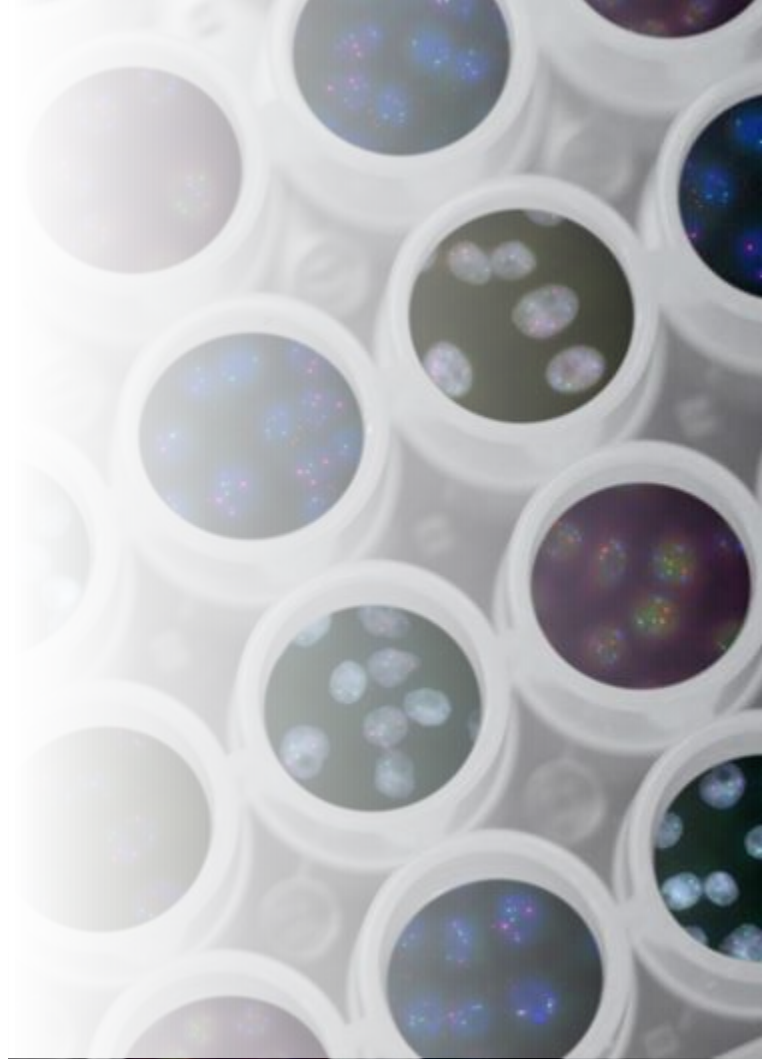
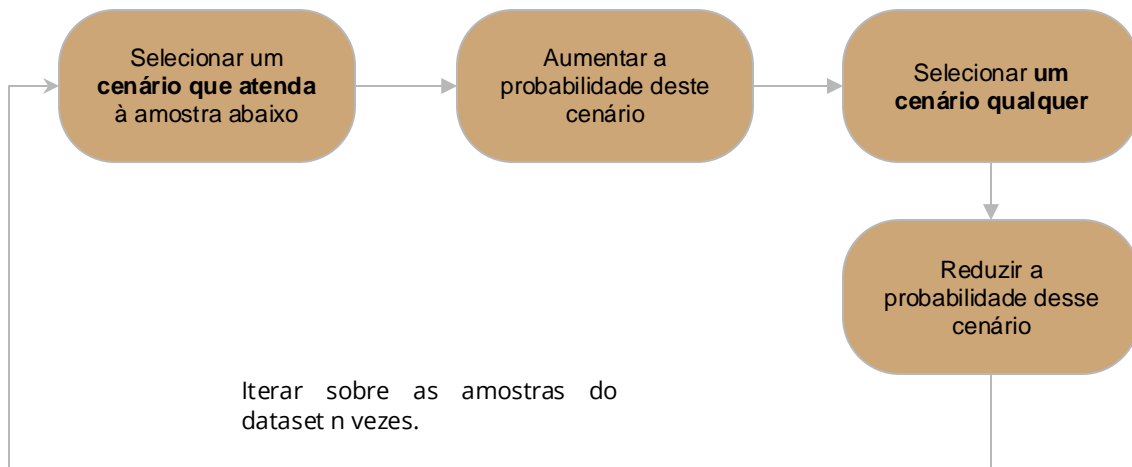


Priscila



Pode acontecer de reduzirmos a probabilidade da mesma amostra que incrementamos na primeira etapa, mas isso tem baixa probabilidade de acontecer.

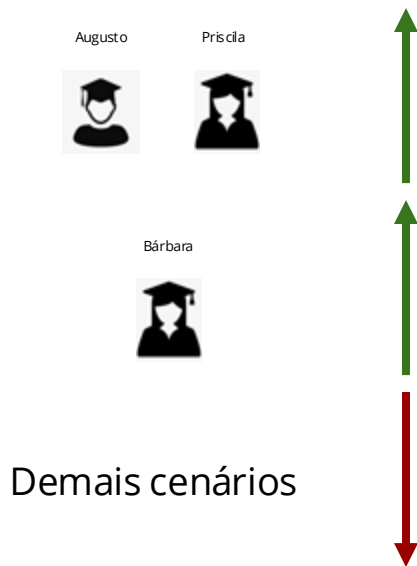
# Treinamento da RBM com amostragem de Gibbs





# Treinamento da RBM com amostragem de Gibbs

- O que esperamos obter no final do treinamento é que a probabilidade dos cenários encontrados no dataset seja alta







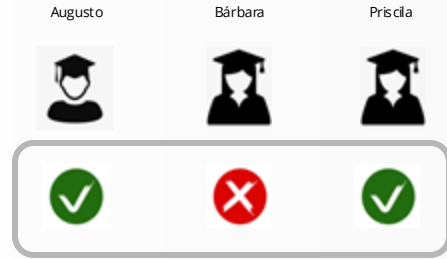
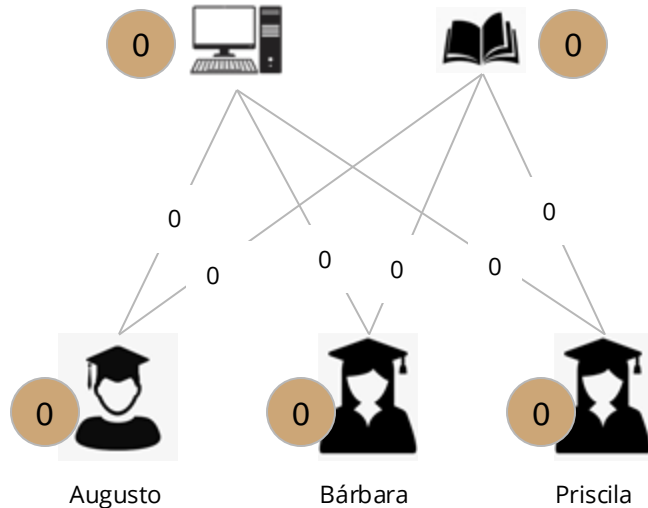
Atualizando os pesos da rede



# Atualizando os pesos da rede

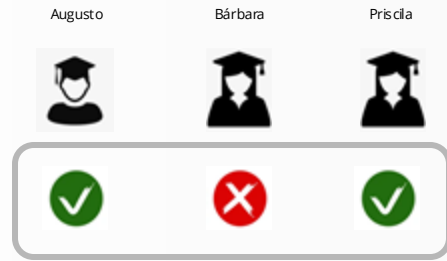
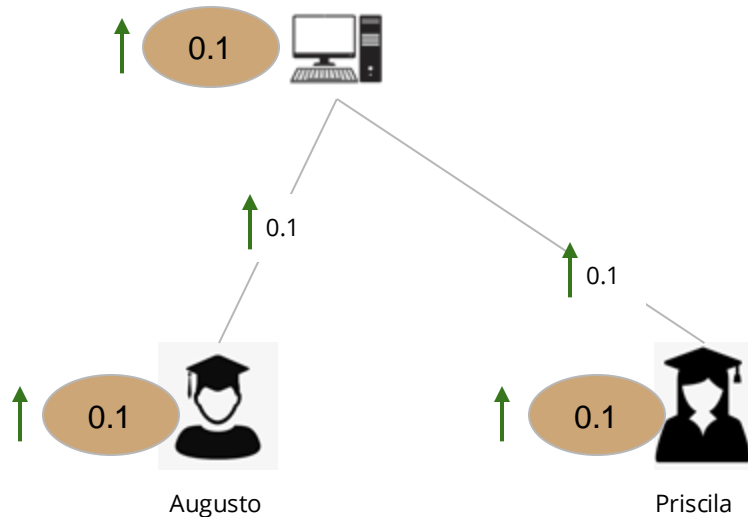
- Para a amostra ao lado, selecionamos o cenário possível:

AP p (Augusto e Priscila compareceram a uma aula prática)



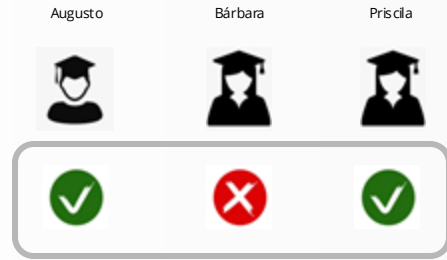
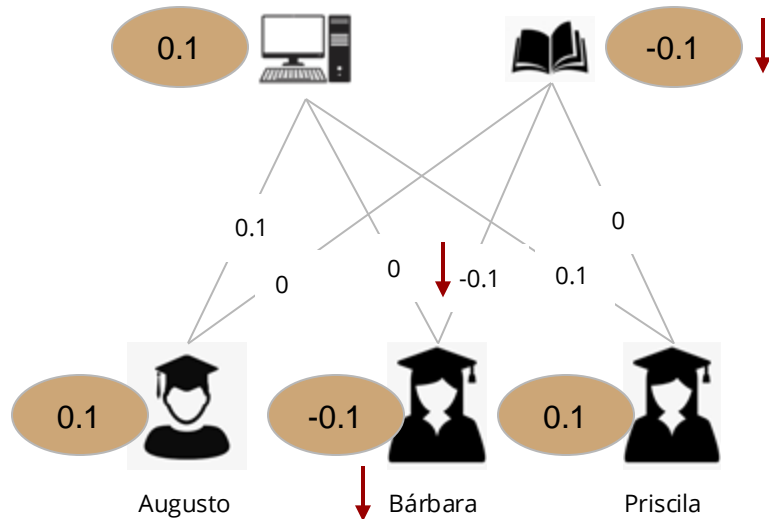
# Atualizando os pesos da rede

- Para atualização dos pesos, vamos considerar somente os nós e conexões que atendem à este cenário:



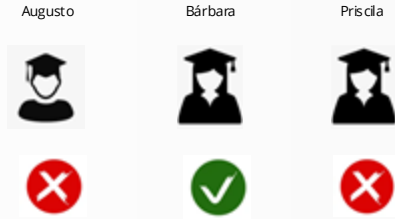
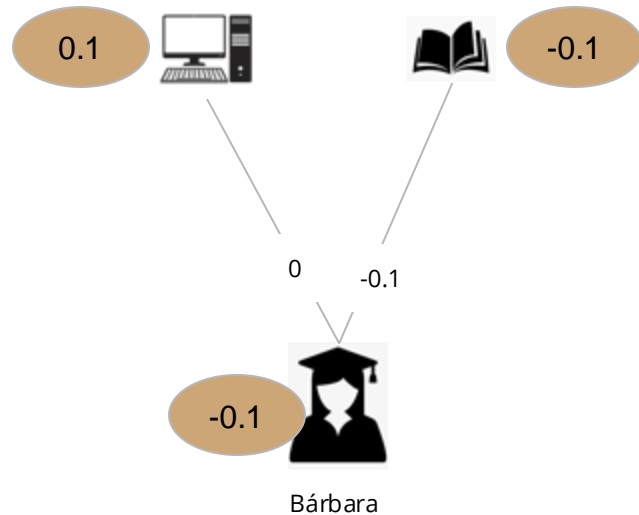
# Atualizando os pesos da rede

- Fazemos o mesmo decrementando o valor de um cenário aleatório (B t):



# Atualizando os pesos da rede

- Para a seleção das amostras, verificamos os pesos atuais para que a seleção do cenário respeite a probabilidade atual:
  - $B_p = 0.5$
  - $B_t = 0.42$



# Aplicações

# Aplicações

- Embora o uso das RBMs tenha diminuído com o advento de arquiteturas como Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs), elas ainda são relevantes em alguns contextos



# Aplicações

- **Cenários com escassez de dados rotulados:** Sua capacidade de aprender distribuições complexas sem supervisão ainda é um diferencial importante





# Aplicações

- **Sistemas de Recomendação (filtragem colaborativa):**
  - RBMs são usadas para prever as preferências dos usuários com base em seus comportamentos passados ou no comportamento de usuários semelhantes. Isso é feito modelando a matriz de interações usuário-item, onde cada linha representa um usuário e cada coluna representa um item. As RBMs aprendem a capturar características ocultas que refletem as preferências dos usuários e os atributos dos itens



# Aplicações

- Existem algumas outras aplicações que podem ser desenvolvidas com RBM:
  - **Detecção de anomalias:** A habilidade das RBMs em modelar dados complexos as torna adequadas para identificar padrões anômalos
  - Redução de dimensionalidade
  - etc



# Aplicações

- **Eficiência de Treinamento:** A inicialização com RBMs pode levar a uma convergência mais rápida e estável durante o treinamento da rede neural completa. Isso ocorre porque os pesos iniciais estão mais próximos de uma solução útil, reduzindo a quantidade de ajuste fino necessário.



# Resumo da Aula de Hoje

- Máquina Boltzmann
- Máquina Restrita de Boltzmann
- Treinamento destas redes



Dúvidas?



Obrigada!