

REDES NEURAIS

A máquina restrita de Boltzmann (RBM)



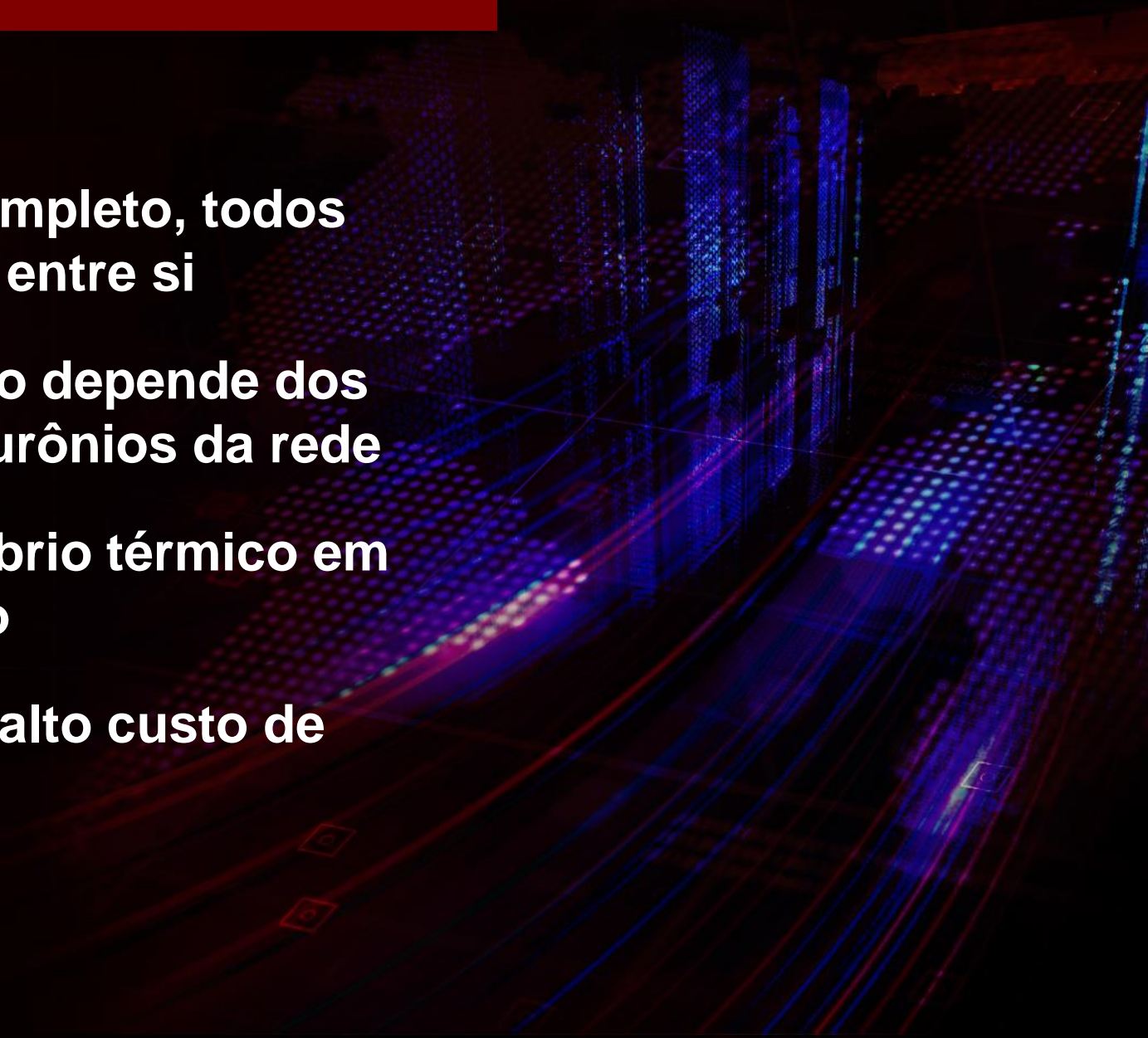
TÓPICOS

1. Os problemas da máquina de Boltzmann
2. A máquina restrita
3. Treinamento
4. Algoritmo de treinamento otimizado:
Divergência Contrastiva



A MÁQUINA DE BOLTZMANN

- É representada por um grafo completo, todos os neurônios estão conectados entre si
- Logo, o estado de cada neurônio depende dos estados de todos os demais neurônios da rede
- O processo para atingir o equilíbrio térmico em ambas as fases é muito custoso
- Aplicações limitadas devido ao alto custo de treinamento



A MÁQUINA DE BOLTZMANN

- A aprendizagem consiste em reduzir a energia dos estados que representam os padrões

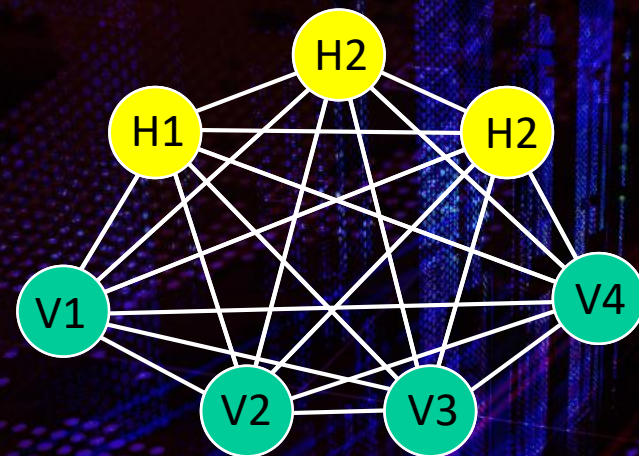
$$E(v, h) = - \sum_{i < j} v_i v_j w_{ij} - \sum_{i, k} v_i h_k w_{ik} - \sum_{k < l} h_k h_l w_{kl}$$

- Duas fases são consideradas na atualização

$$\Delta w_{ij} = \eta \left[(y_i y_j)^+ - (y_i y_j)^- \right]$$

A MÁQUINA RESTRITA

- Podemos restringir a conectividade da rede para melhorar o processo de inferência e o aprendizado da rede
 - As duas camadas são mantidas: oculta e visível
 - Restringe-se as conexões: exclusão das conexões intra-camada
 - Não há ligações entre neurônios ocultos
 - Não há ligações entre neurônios visíveis
- Nessa nova máquina, com apenas um passo conseguimos atingir o equilíbrio térmico na fase positiva



MÁQUINA RESTRITA DE BOLTZMANN

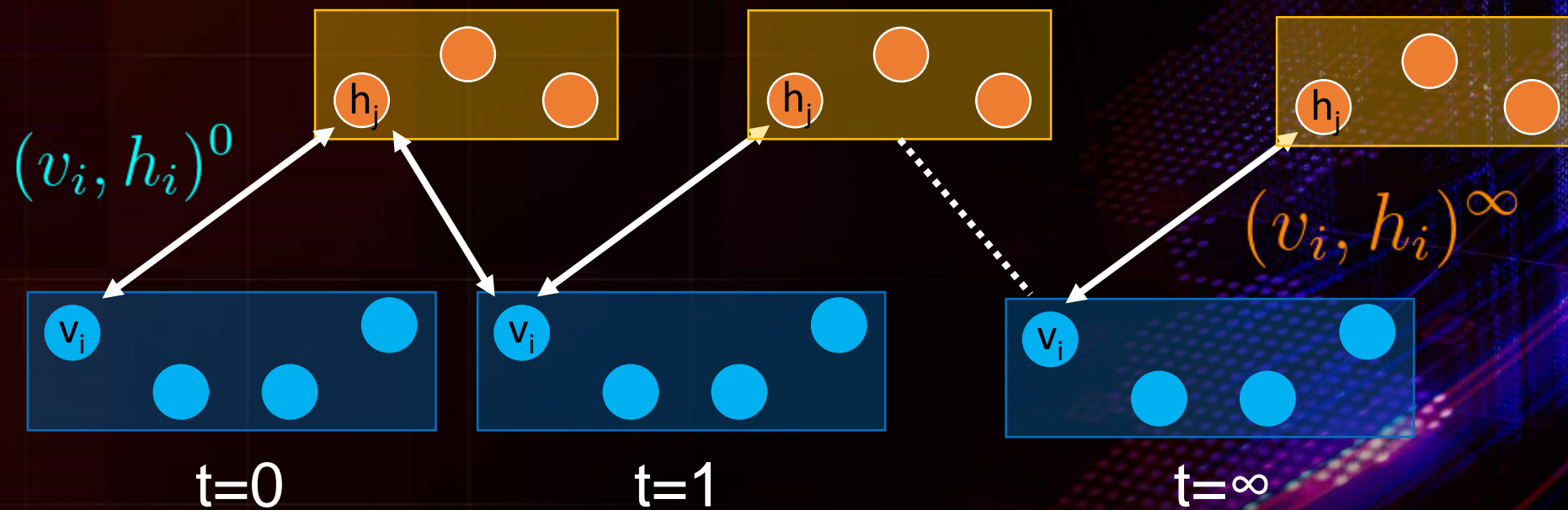
FASE POSITIVA NA RBM:

- Fixa-se um padrão na camada visível e obtém-se as correlações para todos os pares (v_i, h_i) + para todos os exemplos i do batch

FASE NEGATIVA NA RBM:

- A partir de um conjunto de estados iniciais representativos, evoluir a máquina até o equilíbrio
- Calcular as correlações para todos os pares (v_i, h_i) - para todos os estados iniciais

ILUSTRAÇÃO TREINO



$$\Delta w_{ij} = \eta \left[(y_i y_j)^0 - (y_i y_j)^\infty \right]$$

CONSIDERAÇÕES

1. A evolução é mais simples $v \rightarrow h$, $h \rightarrow v$
2. O custo da fase positiva é baixo (1 passo)
3. Porém, o custo da fase negativa é muito alto, limitando a aplicação do modelo em cenários reais

$$\Delta w_{ij} = \eta \left[(y_i y_j)^0 - (y_i y_j)^\infty \right]$$

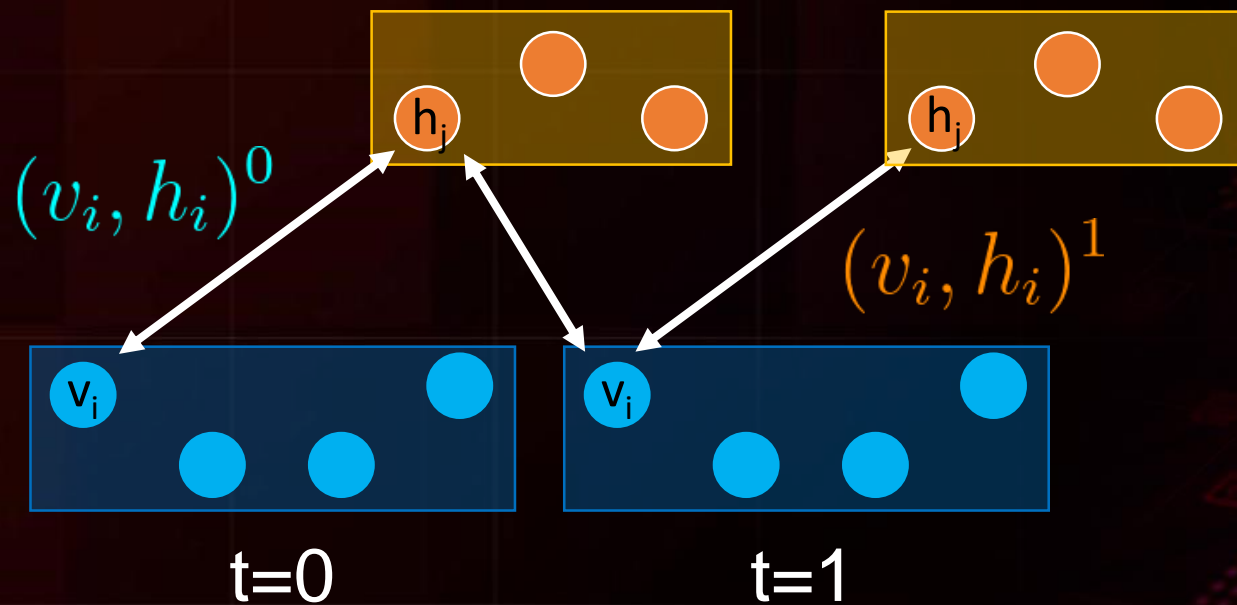
DIVERGÊNCIA CONTRASTIVA

QUATRO PASSOS:

1. A camada visível é fixada com um padrão da base de treino
2. Atualiza os estados das unidades ocultas e calcula as correlações do passo 0 (positivo)
3. Atualiza os estados das unidades visíveis (reconstrução).
4. Atualiza os estados das unidades ocultas novamente e calcula as correlações do passo 1 (estimativa da fase negativa)

$$\Delta w_{ij} = \eta \left[(y_i y_j)^0 - (y_i y_j)^1 \right]$$

DIVERGÊNCIA CONTRASTIVA



$$\Delta w_{ij} = \eta \left[(y_i y_j)^0 - (y_i y_j)^1 \right]$$

DIVERGÊNCIA CONTRASTIVA

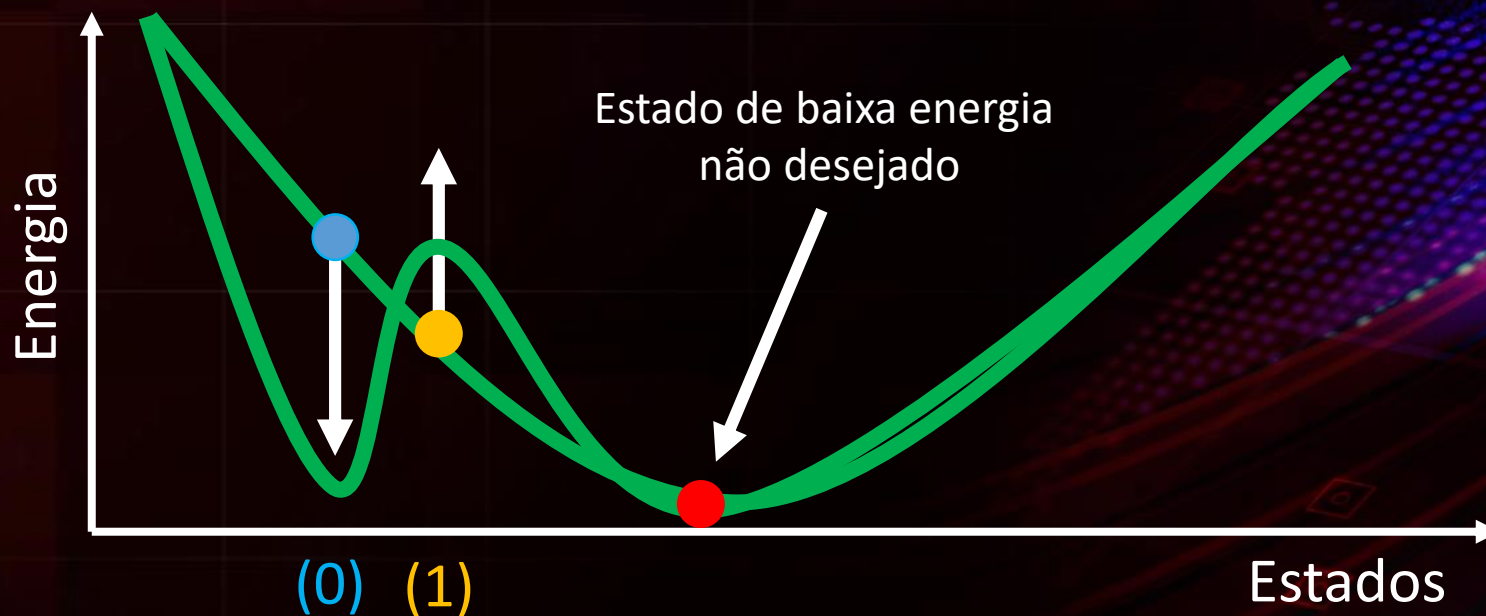
POR QUE FUNCIONA?

- A partir do dado inicial, o modelo evolui em direção ao mínimo de energia
- Ao observar essa direção, podemos evitá-la
- Isso pode ser feito com poucos passos, pois, sabendo que o modelo irá divergir do estado positivo, não há necessidade de atingir o equilíbrio térmico

$$\Delta w_{ij} = \eta \left[(y_i y_j)^0 - (y_i y_j)^1 \right]$$

DIVERGÊNCIA CONTRASTIVA

Para isso, precisamos aumentar a probabilidade dos estados desejados e reduzir a probabilidade dos estados de equilíbrio não desejados



DIVERGÊNCIA CONTRASTIVA

- Para resolver essa limitação, podemos usar mais passos ao invés de apenas um
- Estratégia:
 1. Iniciamos com um passo do algoritmo de Divergência Contrastiva (CD1)
 2. Após estabilização dos pesos, ampliamos para três passos (CD3)
 3. Por fim, após nova estabilização, seguimos com o CD10 para ajuste fino
- Com isso conseguimos um bom balanceamento entre custo computacional e eficiência do modelo

ILUSTRAÇÃO DÍGITO 2

O que os neurônios ocultos representam?
Exemplo do Prof. Geoffrey Hinton utilizando
padrões de “2” manuscritos

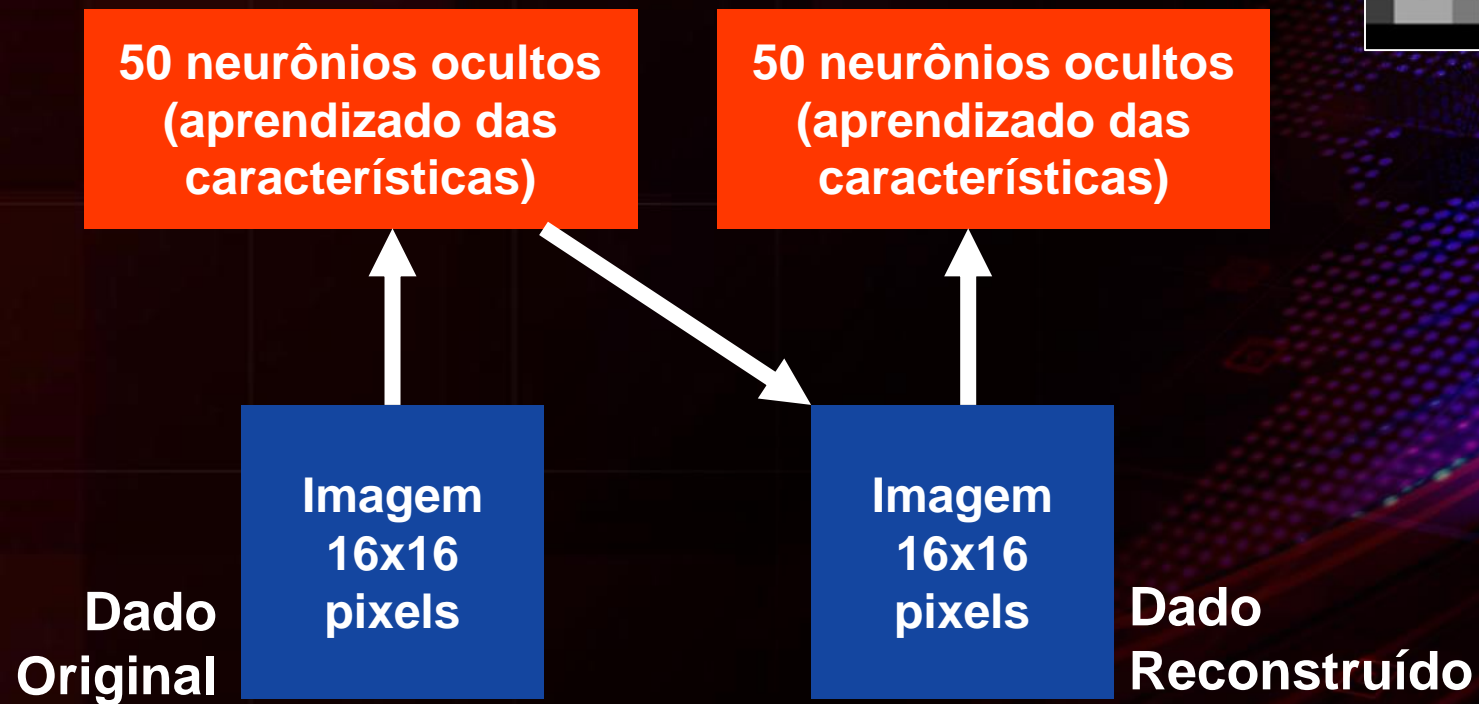
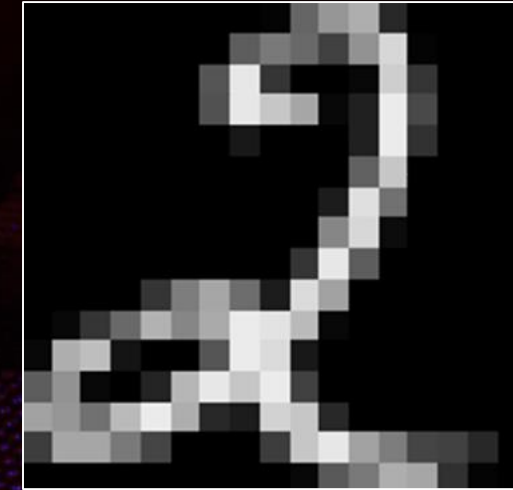
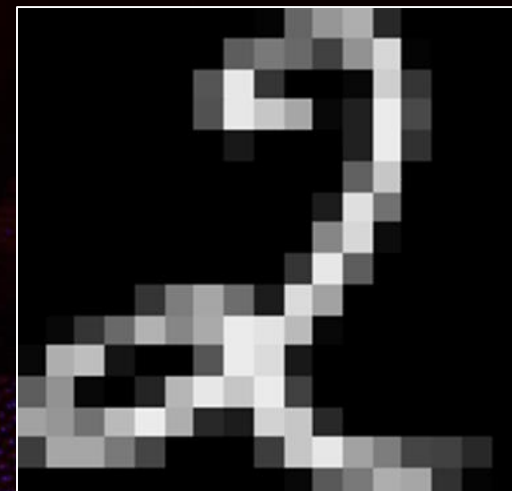
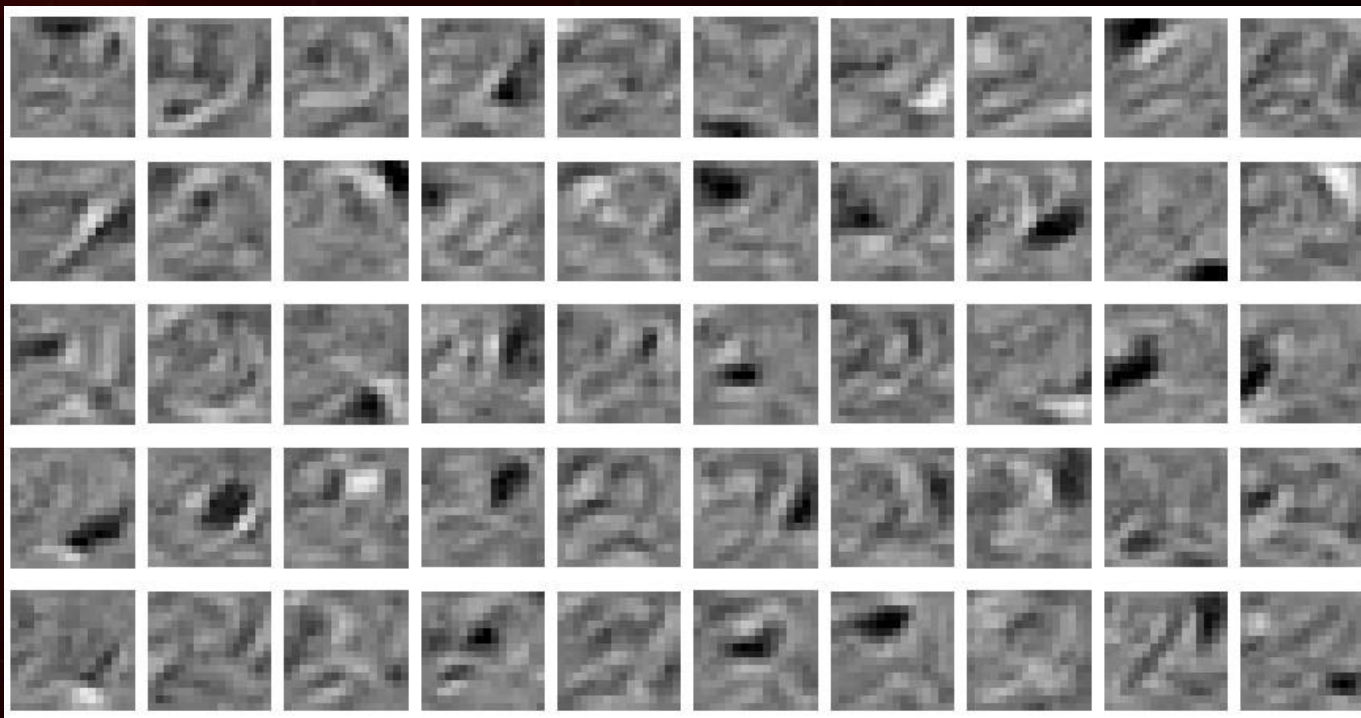
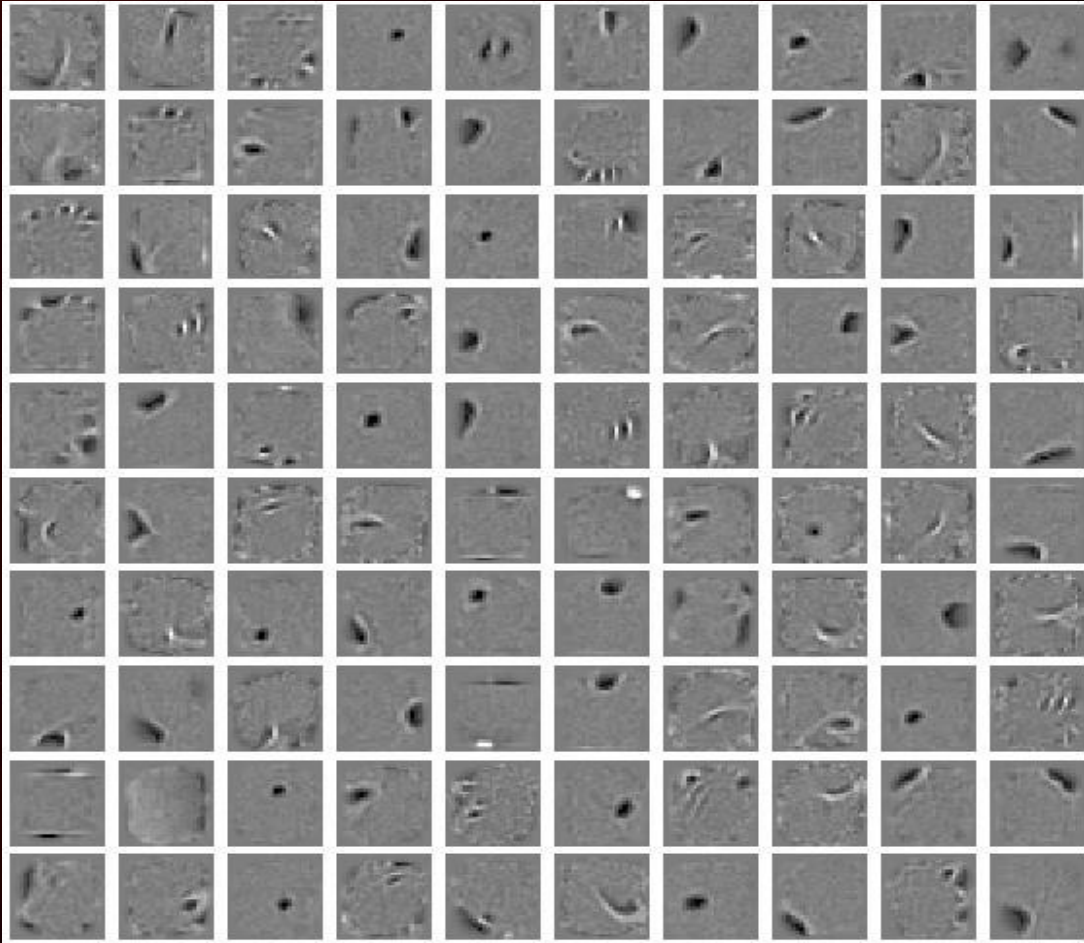


ILUSTRAÇÃO DÍGITO 2

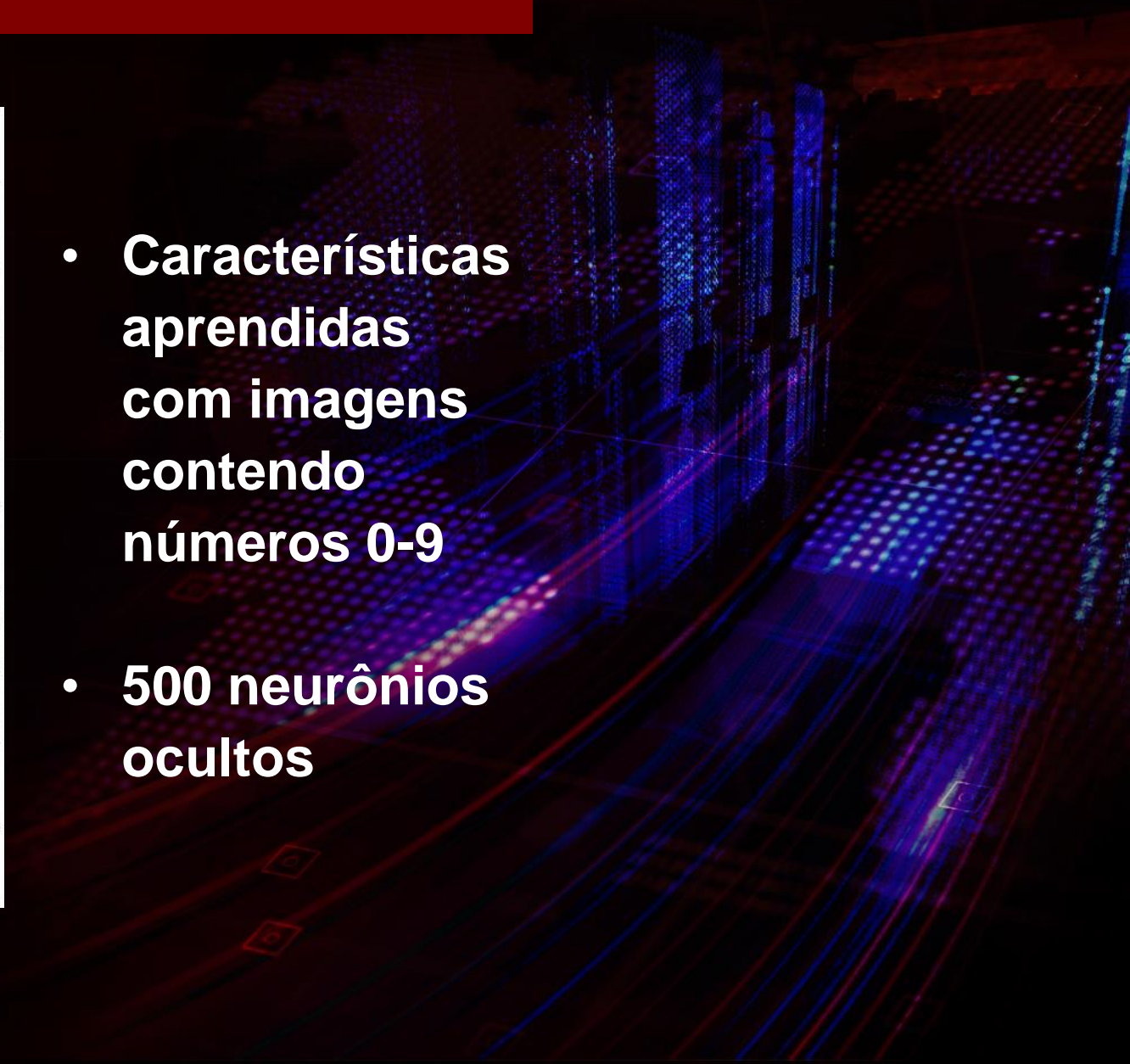


Fim do ciclo de treinamento

ILUSTRAÇÃO: 10 DÍGITOS



- Características aprendidas com imagens contendo números 0-9
- 500 neurônios ocultos



EXTENSÕES

- Tratamento de sinais contínuos
- Múltiplas camadas via empilhamento de RBMs, denominadas (deep) Belief Networks (DBN)

O QUE VIMOS?

- Conhecemos a máquina restrita de Boltzmann e o seu algoritmo de aprendizagem: divergência contrastiva



PRÓXIMA AULA

➤ Na próxima semana aprenderemos sobre outras redes recorrentes não baseadas em energia

- RNNs
- GRU
- LSTM

ATÉ A PRÓXIMA VIDEOAULA