

# APRENDIZADO DE MÁQUINA

REDES NEURAIS RECORRENTES

PROF. RONALDO CRISTIANO PRATI

[ronaldo.prati@ufabc.edu.br](mailto:ronaldo.prati@ufabc.edu.br)

Bloco A, sala 513-2

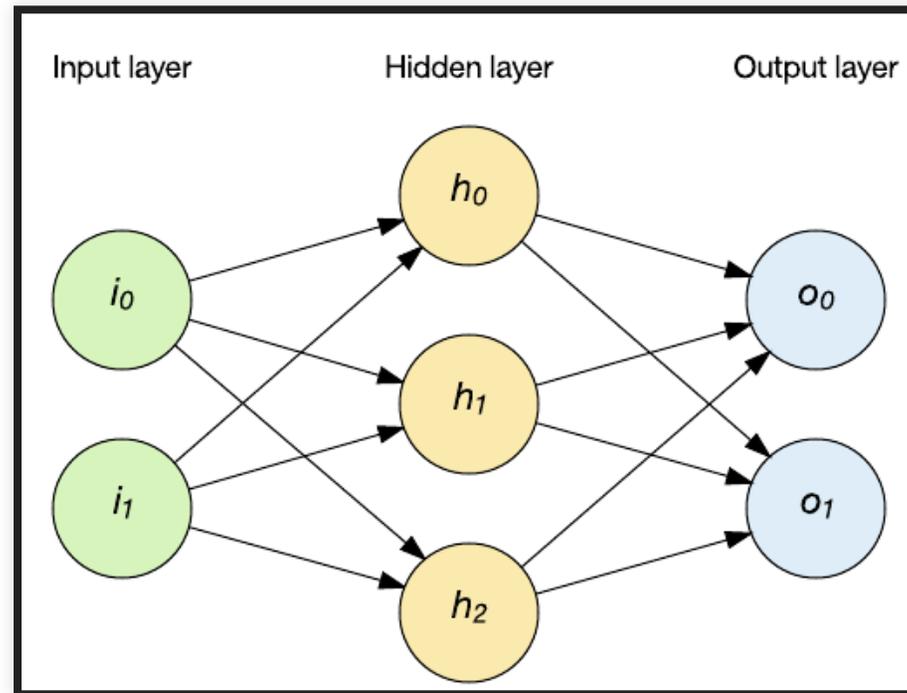
# REDES DE PROPAGAÇÃO FORWARD

- Redes de propagação forward consideram que OS
  - exemplos de treinamento são IID,
  - entrada e saída têm tamanho fixo.
- De que forma considerar dados de tamanhos variados em que há dependências de curto/longo prazo?

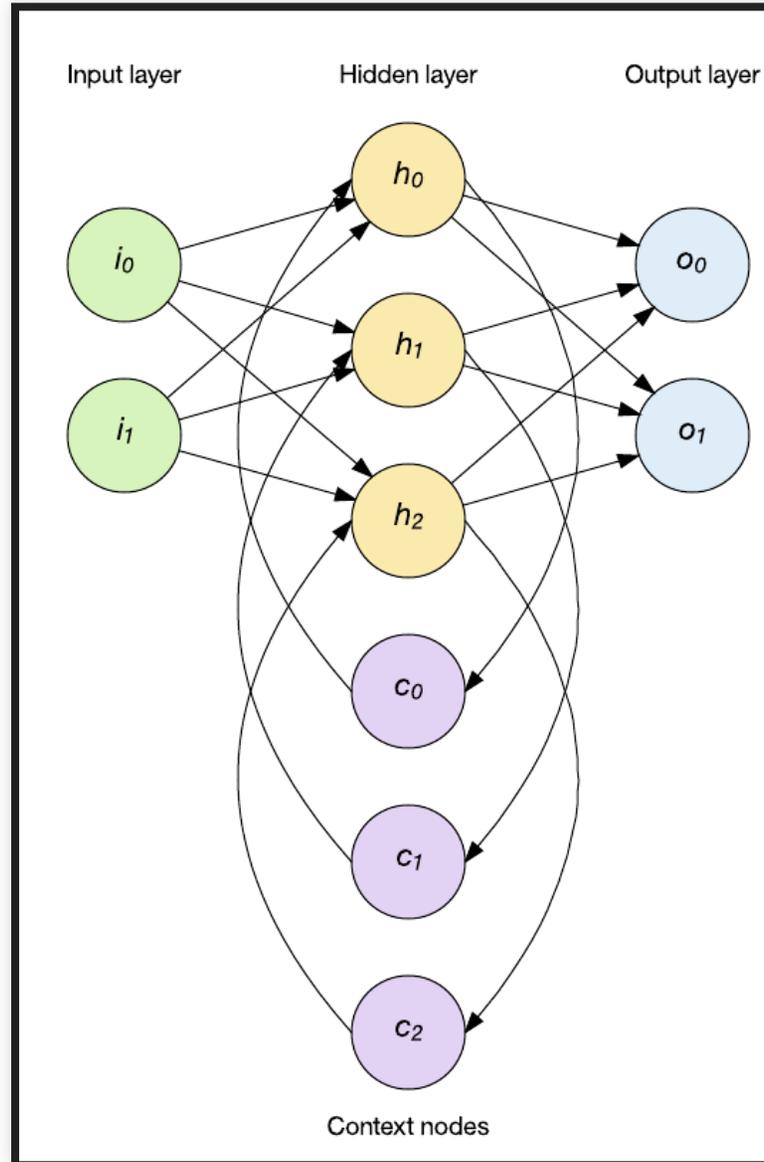
# REDES RECORRENTES

- Constituem uma ampla classe de redes cuja evolução do estado depende tanto da entrada corrente quanto do estado atual.
- São sistemas *Turing-completos* i.e., com tempo, dados e neurônios suficientes, RNNs podem aprender qualquer coisa que um computador pode fazer.
- Aplicáveis a dados sequenciais
  - Texto, vídeo, áudio, ...

# REDE FEED-FORWARD



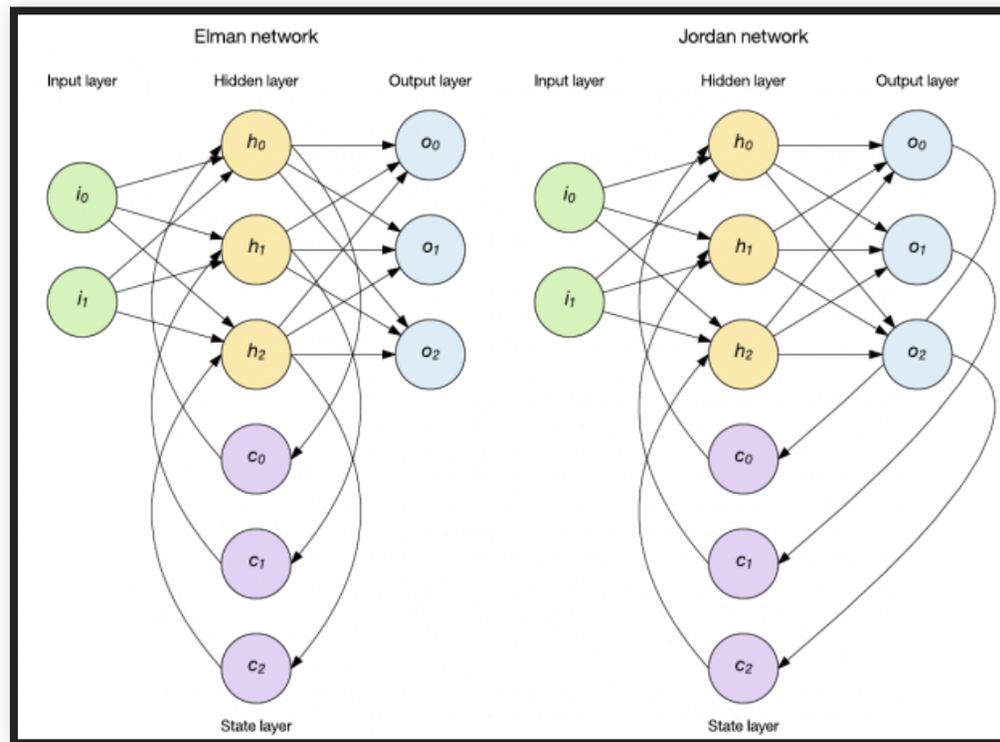
# REDES RECORRENTES



# REDES RECORRENTES SIMPLES

- No caso da rede **Elman**, a camada oculta alimenta uma camada de estado de nodes de contexto que retém memória de entradas passadas.
- As redes **Jordan** diferem em que, ao invés de manter o histórico da camada oculta, elas armazenam a camada de saída na camada de estado.

# REDES RECORRENTES



# REDES RECORRENTES

- Uma RNN mantém em sua(s) camada(s) oculta(s) uma **memória interna** dos dados que lhe foram apresentados.
- Essa memória é uma **combinação** dos dados de entrada e do estado anterior da camada oculta.

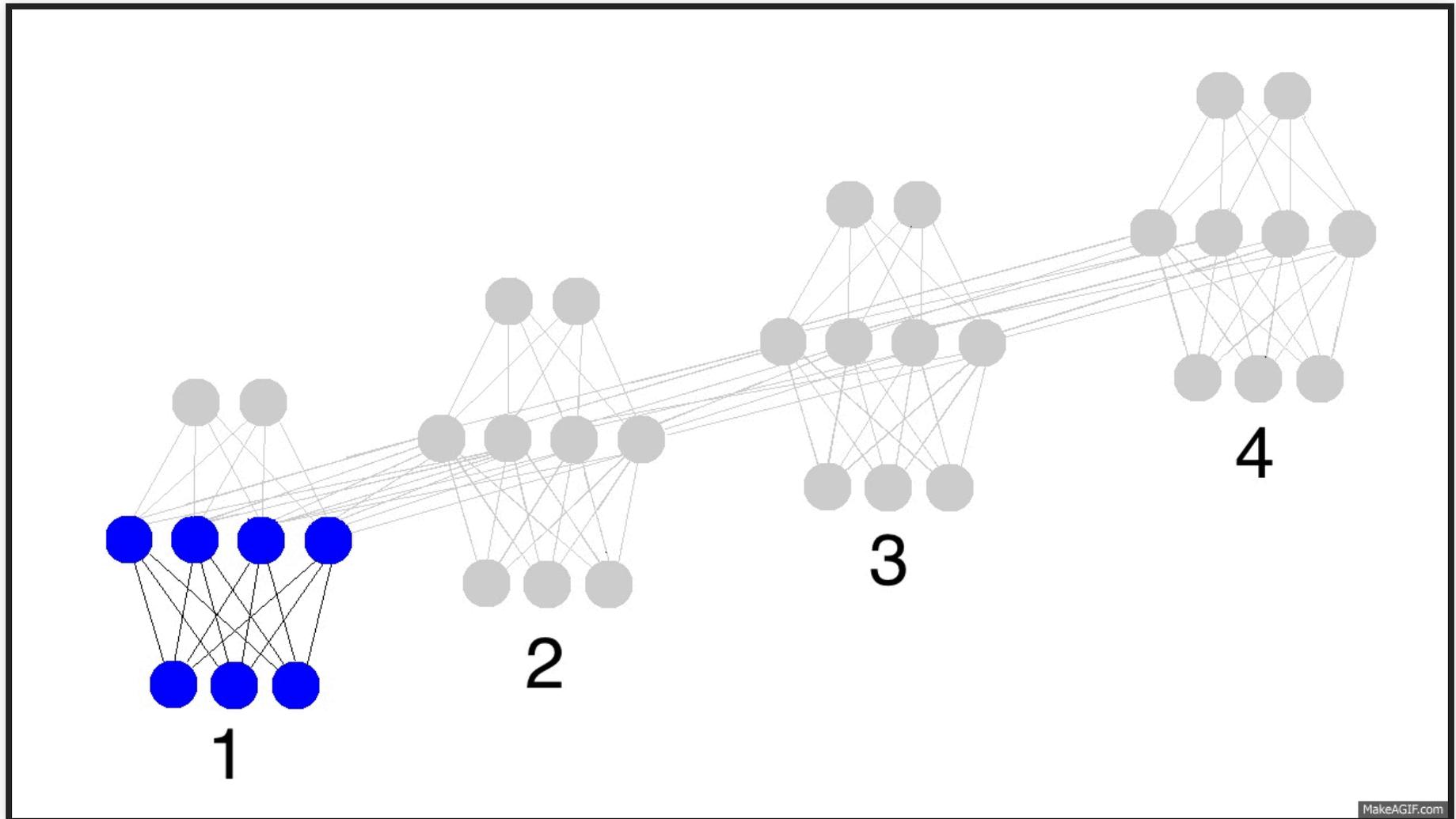
# REDES RECORRENTES

- A rede aprende o que "relembrar"

```
(input + empty_hidden) -> hidden -> output  
(input + prev_hidden) -> hidden -> output  
(input + prev_hidden) -> hidden -> output  
(input + prev_hidden) -> hidden -> output
```

# REDES RECORRENTES

- A rede aprende o que "relembrar"





# REDES RECORRENTES

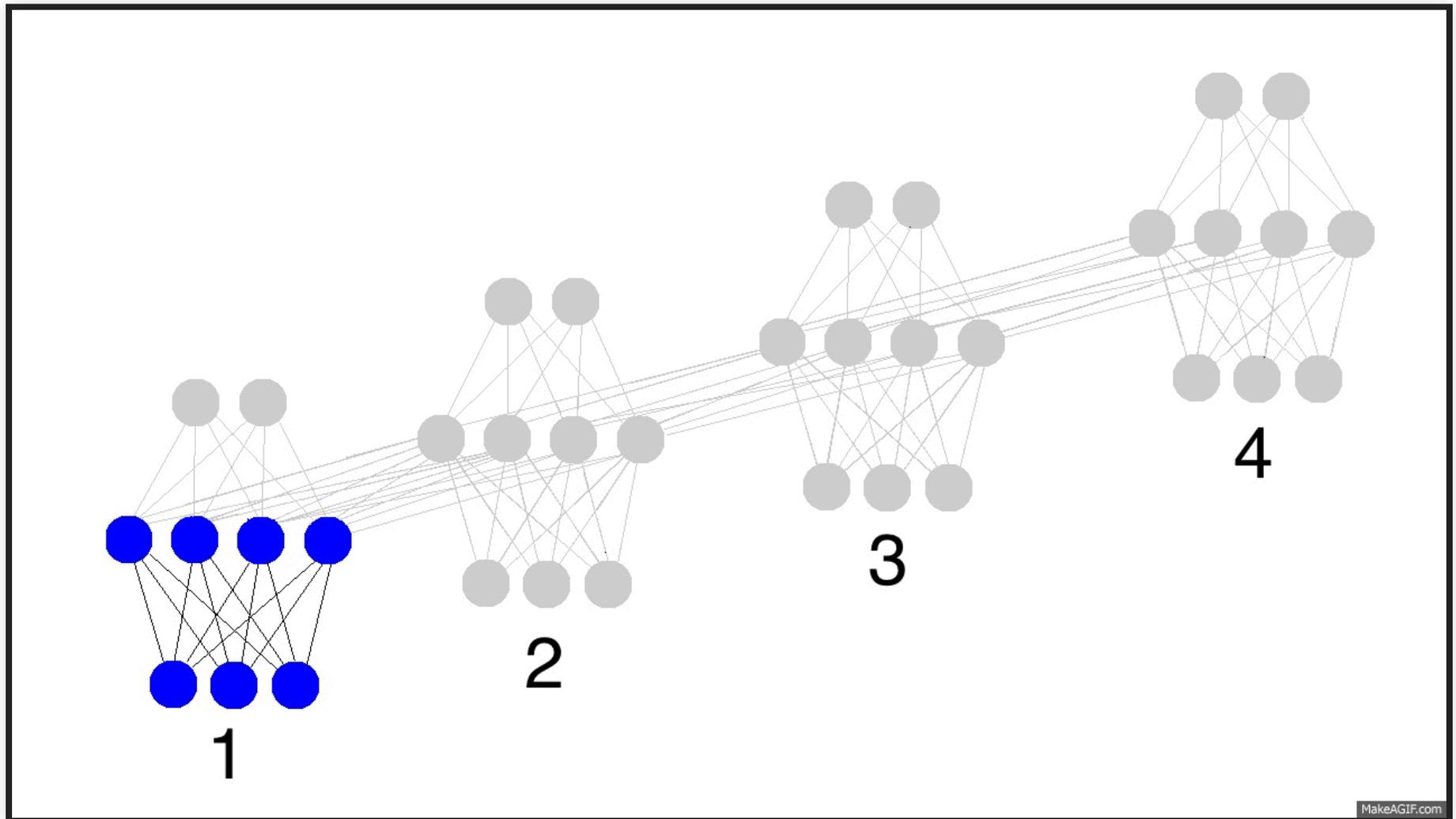
Also, this is when the network learns to forget irrelevant memories and remember important memories.

The input is just changing what's in the memory, and the output is exclusively based on the memory.

# DESDOBRAMENTO NO TEMPO

- RNNs são treinadas em uma sequência de sinais da entrada, um em cada passo de tempo.
- Uma RNN que recebe uma sequência de  $n$  vetores de entrada pode ser vista como uma rede alimentada adiante de  $n$  camadas.
- Matrizes de pesos são reusadas a cada passo de tempo (*shared weights*).

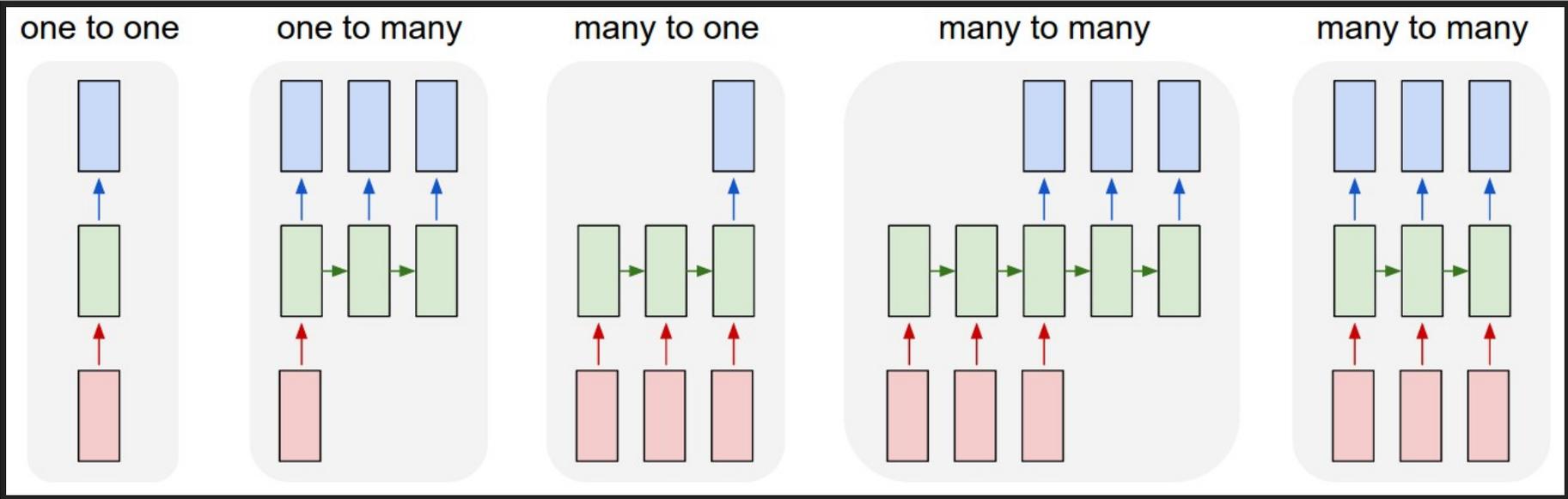
# DESDOBRAMENTO NO TEMPO



# DESDOBRAMENTO NO TEMPO

- Desdobramento no tempo permite o aprendizado de mapeamentos entre sequências de entrada e de saída com diferentes tamanhos:
  - um-para-muitos,
  - muitos-para-um,
  - muitos-para-muitos.

# DESDOBRAMENTO NO TEMPO



# DESDOBRAMENTO NO TEMPO

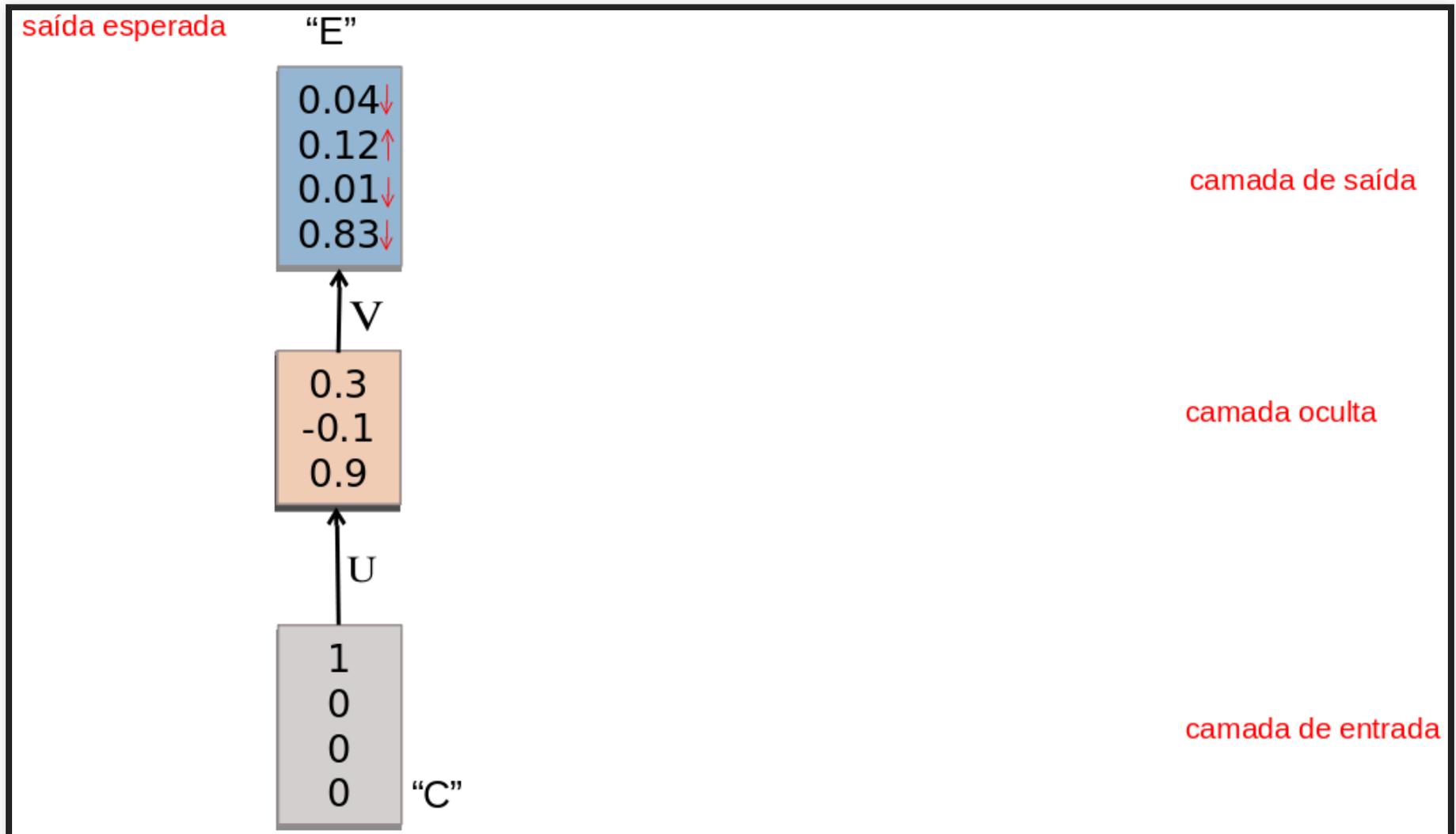
- um-para-muitos
  - prever a legenda correspondente a uma imagem.
- muitos-para-um
  - prever a carga de sentimento (positiva ou negativa) de uma frase de entrada.
- muitos-para-muitos
  - traduzir uma frase do inglês para o português
  - rotular os quadros (frames) de um vídeo
  - TTS (text-to-speech)

# EXEMPLO: MODELO DE LINGUAGEM

- Considere uma RNN que, a cada passo de tempo, prediz o próximo caractere de uma sequência, dada uma sequência de caracteres anteriores.
- Essa rede pode ser usada para gerar novos textos, um caractere por vez (character-level language model).
- Considere também:
  - representação one-hot-encoding para vetores de entrada,
  - alfabeto de apenas 4 caracteres: {C, E, F, T},
  - uso da função softmax na saída.

# EXEMPLO: MODELO DE LINGUAGEM

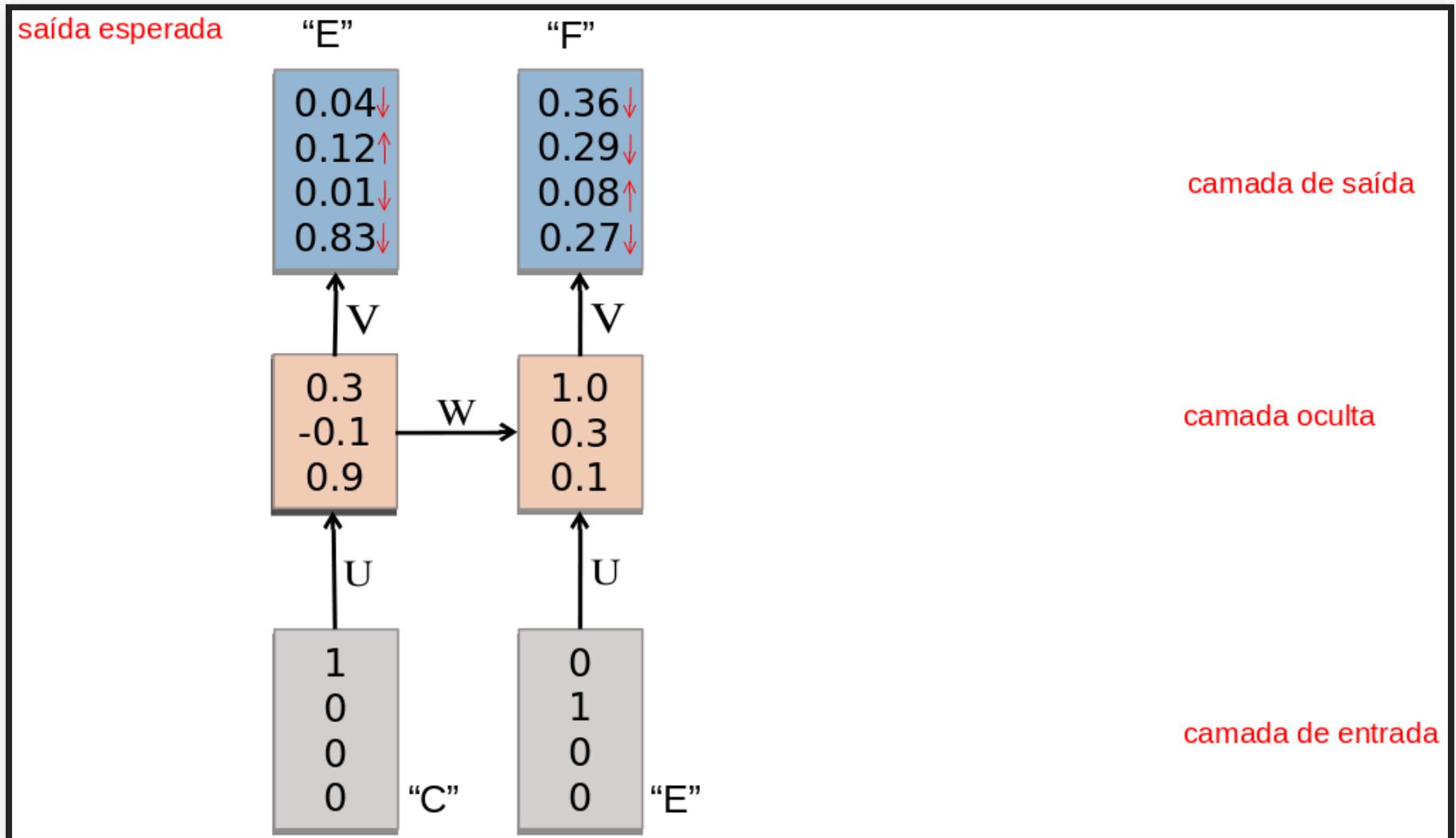
- Um para um





# EXEMPLO: MODELO DE LINGUAGEM

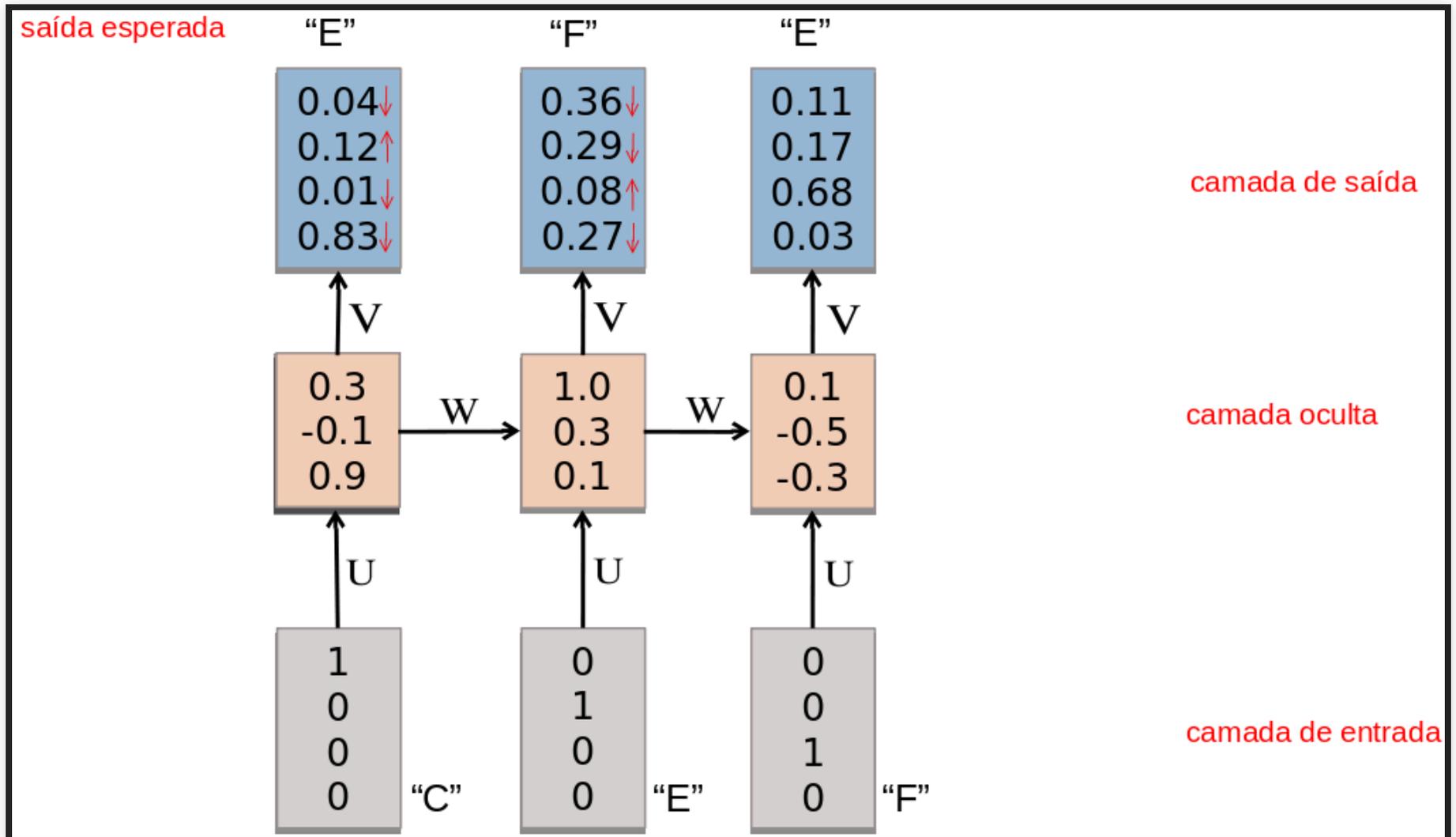
- Um para um





# EXEMPLO: MODELO DE LINGUAGEM

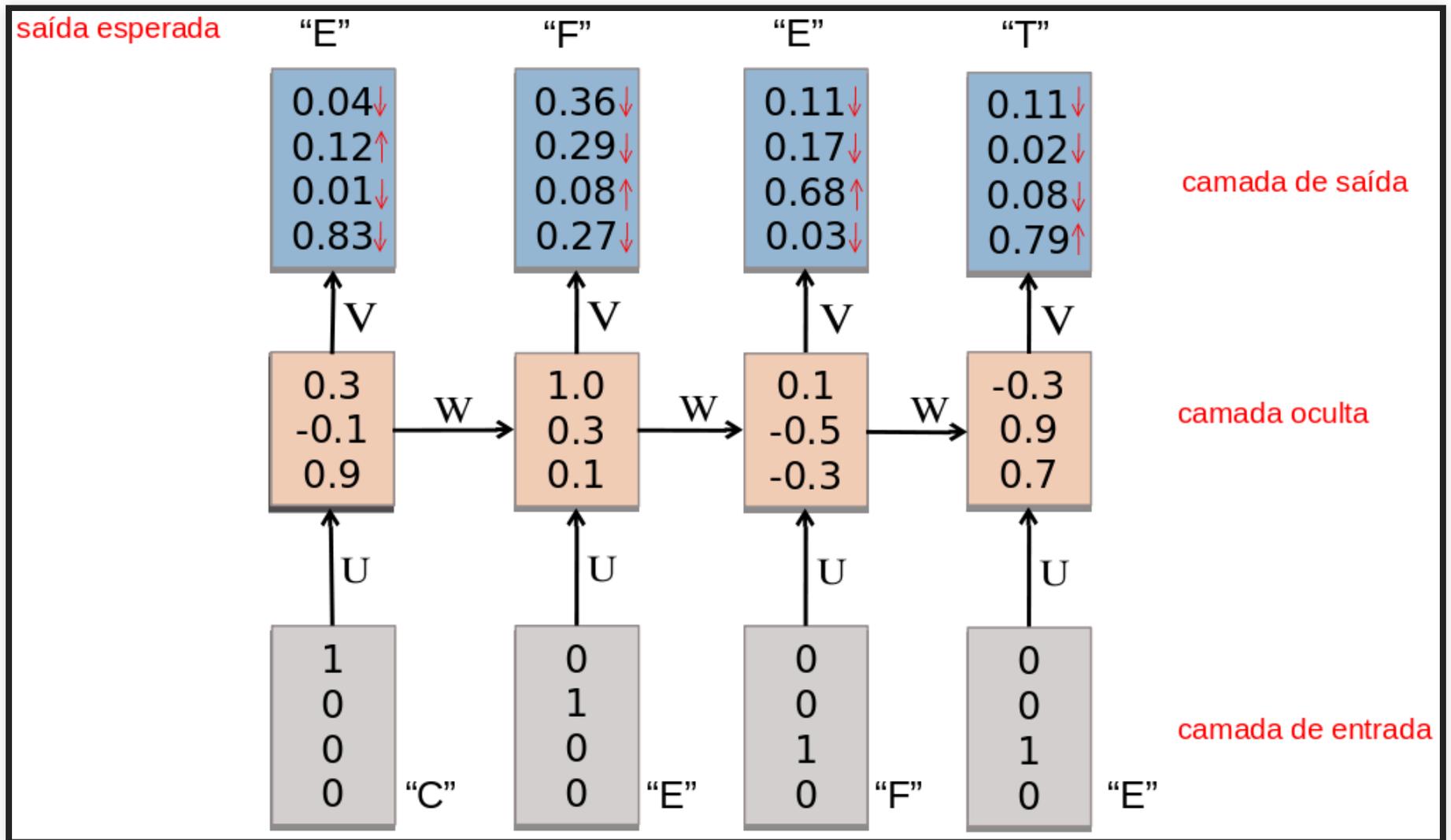
- Um para um





# EXEMPLO: MODELO DE LINGUAGEM

- Um para um



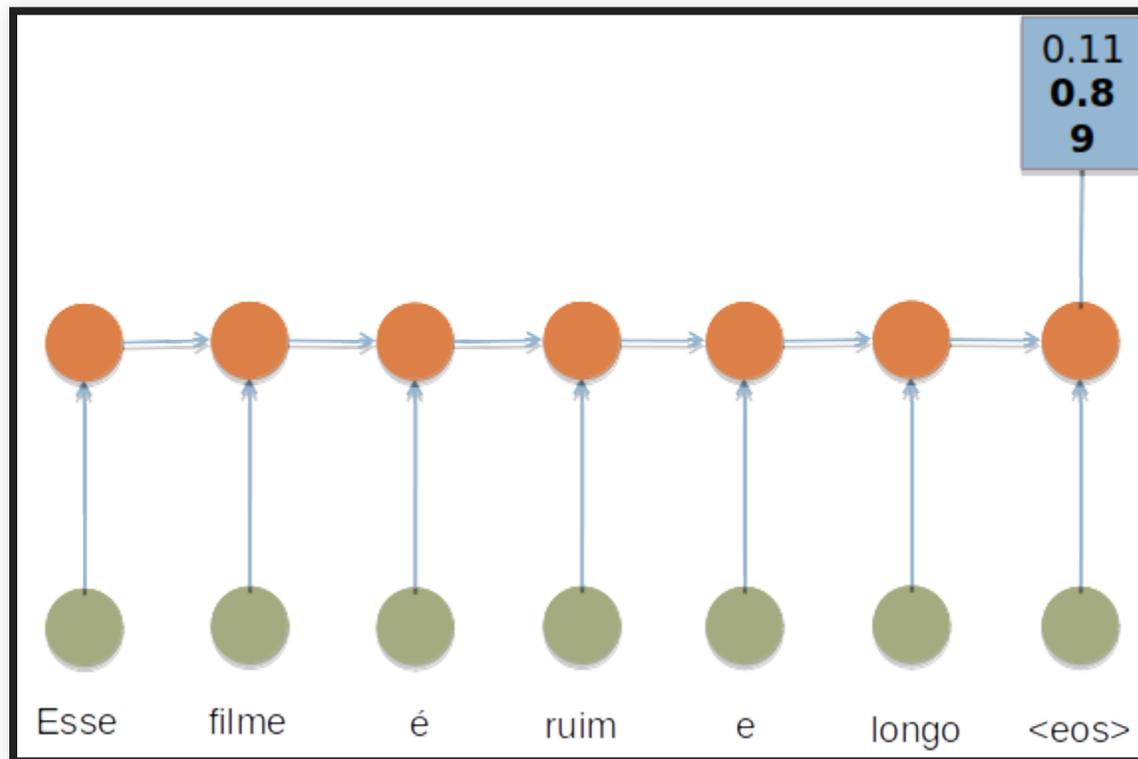


# EXEMPLO: ANÁLISE DE SENTIMENTOS

- Considere um RNN que, a cada passo de tempo, é alimentada com uma palavra.
- Ao final de uma sentença, prever qual é a carga de sentimento daquela sentença.

# EXEMPLO: ANÁLISE DE SENTIMENTOS

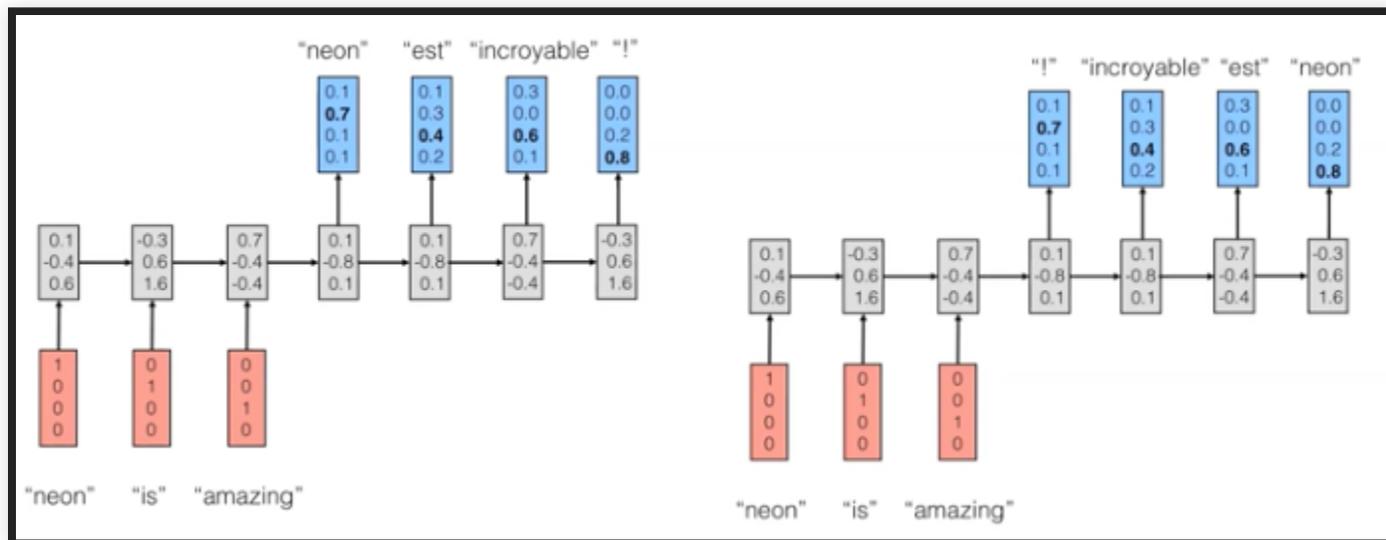
- muitos para um



# EXEMPLO: TRADUÇÃO

- Considere um RNN que, a cada passo de tempo, é alimentada com uma sentença.
- O objetivo é tentar prever a tradução da sentença em uma outra língua.

# EXEMPLO: TRADUÇÃO



# BACKPROPAGATION THROUGH TIME (BPTT)

- Variante do backprop para RNNs.
- Pode ser entendido ao interpretarmos uma RNN como a uma rede alimentada adiante na qual:
  - há uma camada para cada passo de tempo;
  - a matriz de pesos de cada camada é a mesma (i.e., há o compartilhamento de pesos).

# BACKPROPAGATION THROUGH TIME (BPTT)

- Por exemplo, considere que  $w_1$  e  $w_2$  são pesos de conexões correspondentes na camada oculta em instantes de tempo distintos.
- o BPTT calcula os gradientes relativos a  $\frac{\partial J}{\partial w_1}$  e  $\frac{\partial J}{\partial w_2}$  e usa a média dessas quantidades para atualizar  $w_1$  e  $w_2$ .

# DISSIPACÃO/EXPLOÇÃO DOS GRADIENTES

- Dois problemas que podem ocorrer durante o treinamento de RNNs:
  - Vanishing gradients
  - Exploding gradients
- Uma das causas: o fluxo de informação (propagação dos gradientes) em uma RNN ocorre por meio de sequências de multiplicações.

# DISSIPACÃO/EXPLOÇÃO DOS GRADIENTES

- **Vanishing:** Os gradientes “desaparecem” (i.e., tendem a zero) quando são continuamente multiplicados por números inferiores a um.
- **Exploding:** Os gradientes se tornam exponencialmente grandes quando são multiplicados por números maiores que 1.

# EXPLOÇÃO DOS GRADIENTES

- Algumas técnicas para atenuar o problema da explosão dos gradientes:
  - Truncar o BPTT (truncated BPTT)
  - Truncar os gradientes (gradient clipping)
  - Usar otimizadores que ajustem adaptativamente a taxa de aprendizado (os pacotes com keras, tensorflow, etc, implementam vários).

# BPTT TRUNCADO

- Corresponde a limitar a quantidade de passos de tempo no passo *backward* a um valor máximo preestabelecido.
  - e.g., aplicar backprop a cada 10 passos.
- Desvantagem: perda do contexto temporal.

# GRADIENTE TRUNCADO

- Se o comprimento do vetor gradiente ultrapassar um valor máximo preestabelecido, truncar.
- Várias possibilidades...uma delas:
  - $\text{new\_gradients} = \text{gradients} * \text{threshold} / \text{l2\_norm}(\text{gradients})$
- Técnica simples e efetiva!

# DISSIPACÃO DOS GRADIENTES

- Muito mais difícil de detectar!
- Alternativas de solução:
  - Usar boas estratégias de inicialização dos pesos
  - Usar otimizadores que adaptam a taxa de aprendizado (os pacotes como keras, tensorflow, etc, implementam vários).
  - Usar LSTMs ou GRUs

# REDES LSTMS

- Tipo de RNN que possui uma dinâmica de propagação de gradientes mais rebuscada.
- Adequada quando há intervalos muito longos de tamanho desconhecido entre eventos importantes.
- Durante o treinamento, uma rede LSTM pode aprender quando deve manter ou descartar informações.

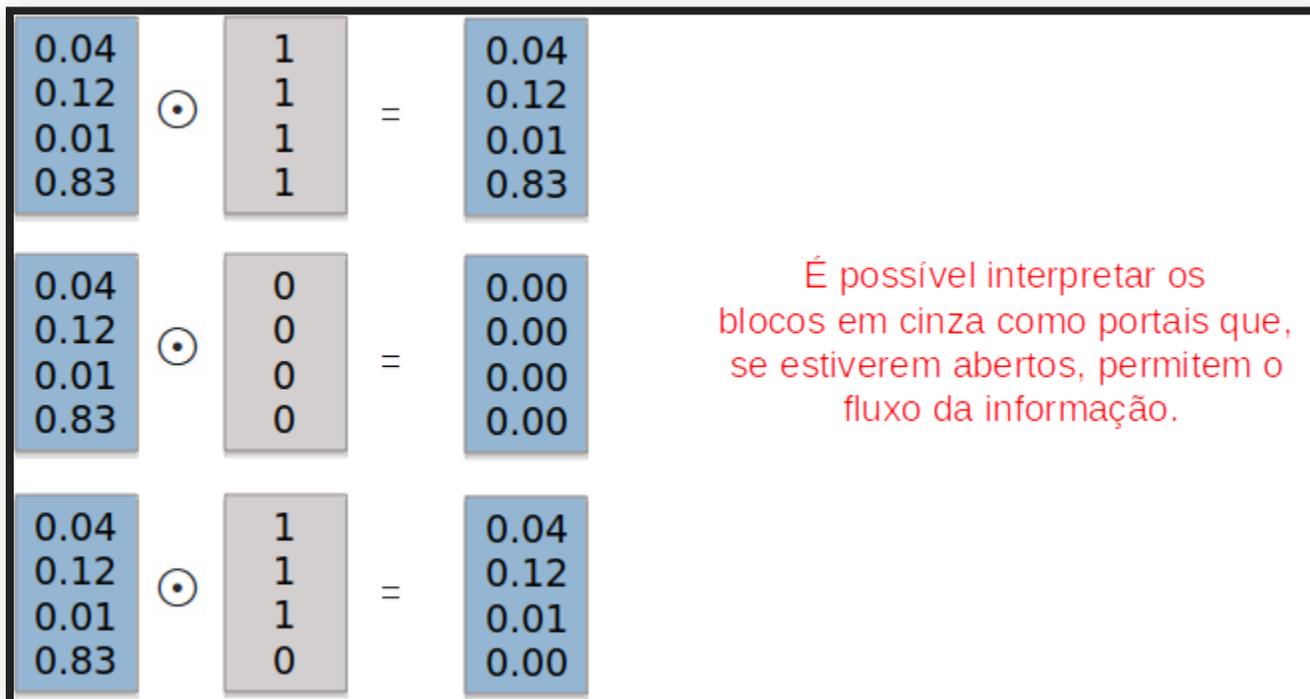
# INTUIÇÃO:

- Em um vídeo (filme):
  - quando termina uma cena, podemos esquecer o dia em que ela ocorreu;
  - mas se um personagem morre na cena, devemos lembrar disso para entender as próximas.
- Em um texto, quando termina uma frase, um tradutor pode esquecer o gênero do sujeito.

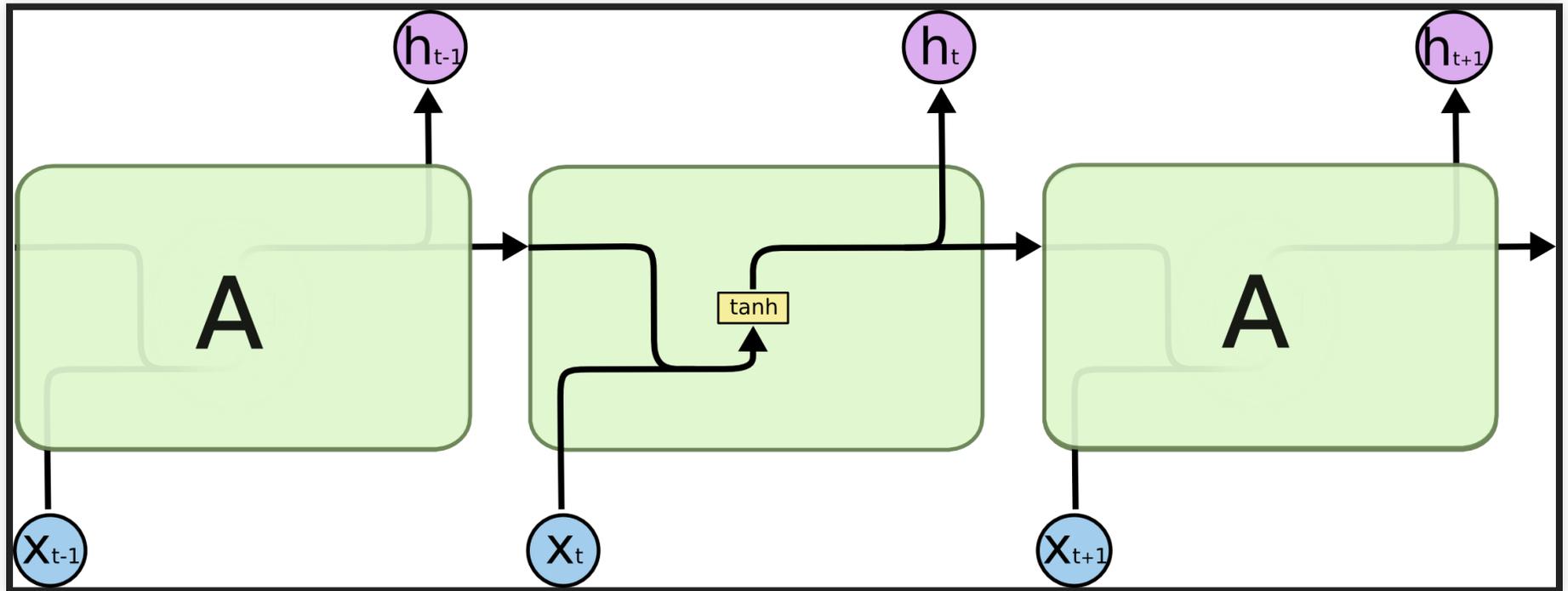
# INTUIÇÃO:

0.04 0.12 0.01 0.83	⊙	1 1 1 1	=	0.04 0.12 0.01 0.83	Guardei toda a informação!
0.04 0.12 0.01 0.83	⊙	0 0 0 0	=	0.00 0.00 0.00 0.00	Esqueci toda a informação!
0.04 0.12 0.01 0.83	⊙	1 1 1 0	=	0.04 0.12 0.01 0.00	Esqueci uma parte!

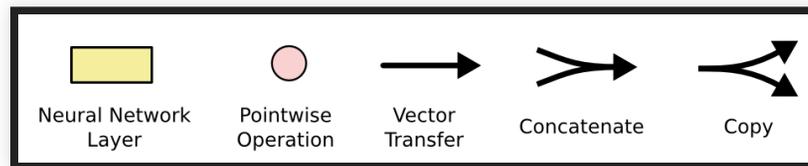
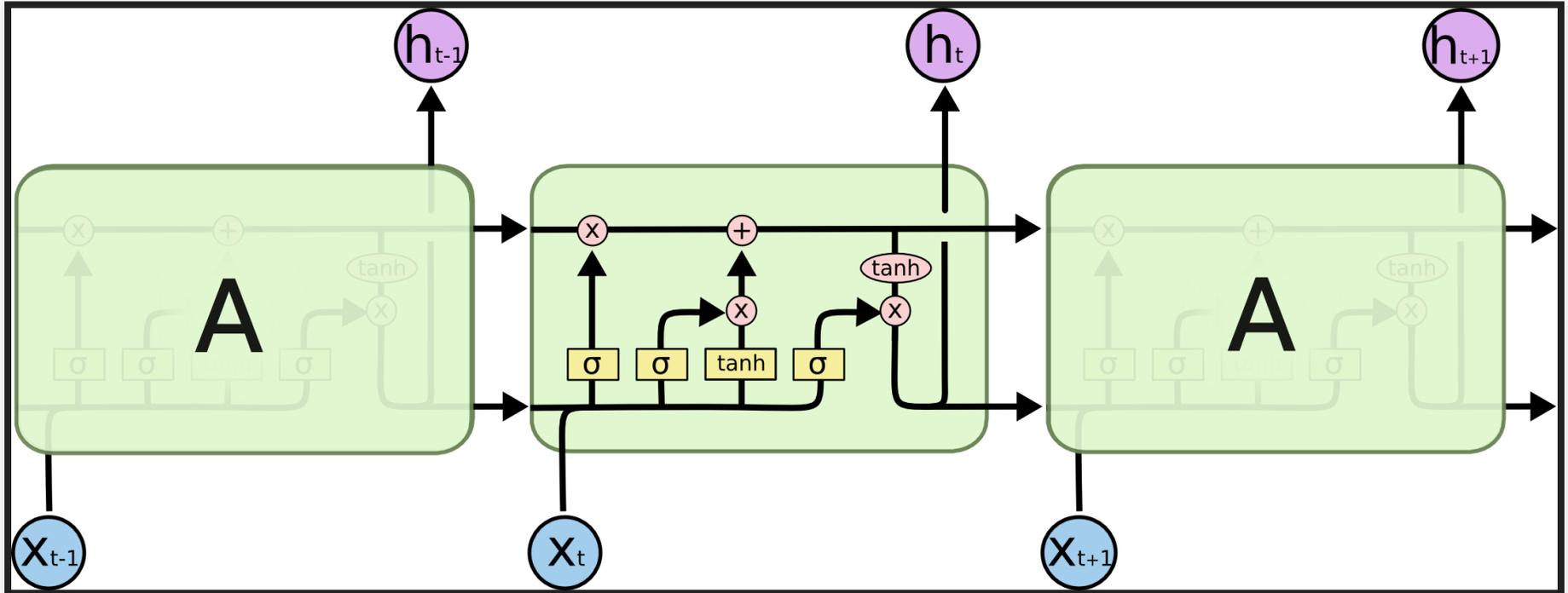
# INTUIÇÃO:



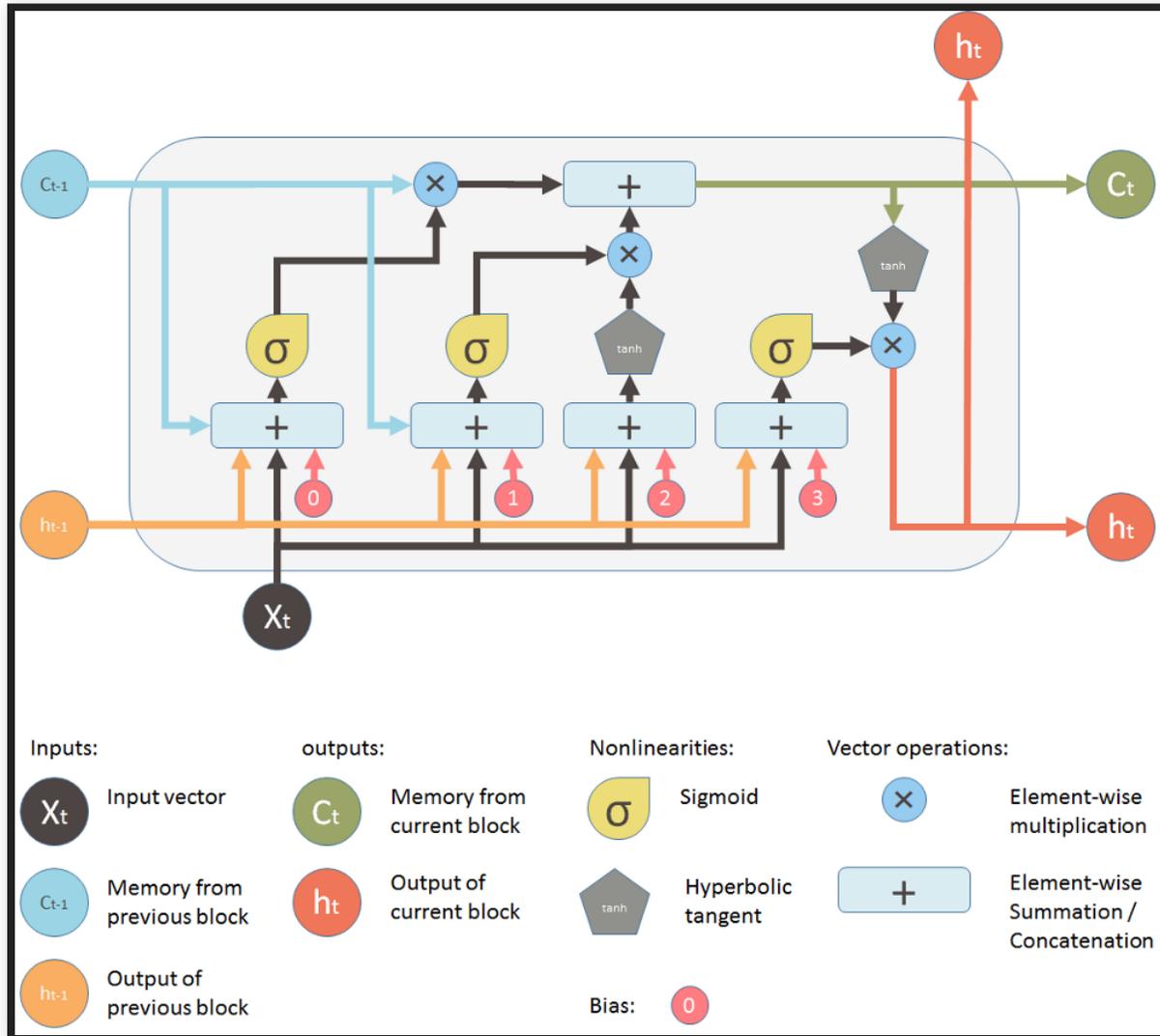
# RNNS SIMPLES



# LSTM



# LSTMS

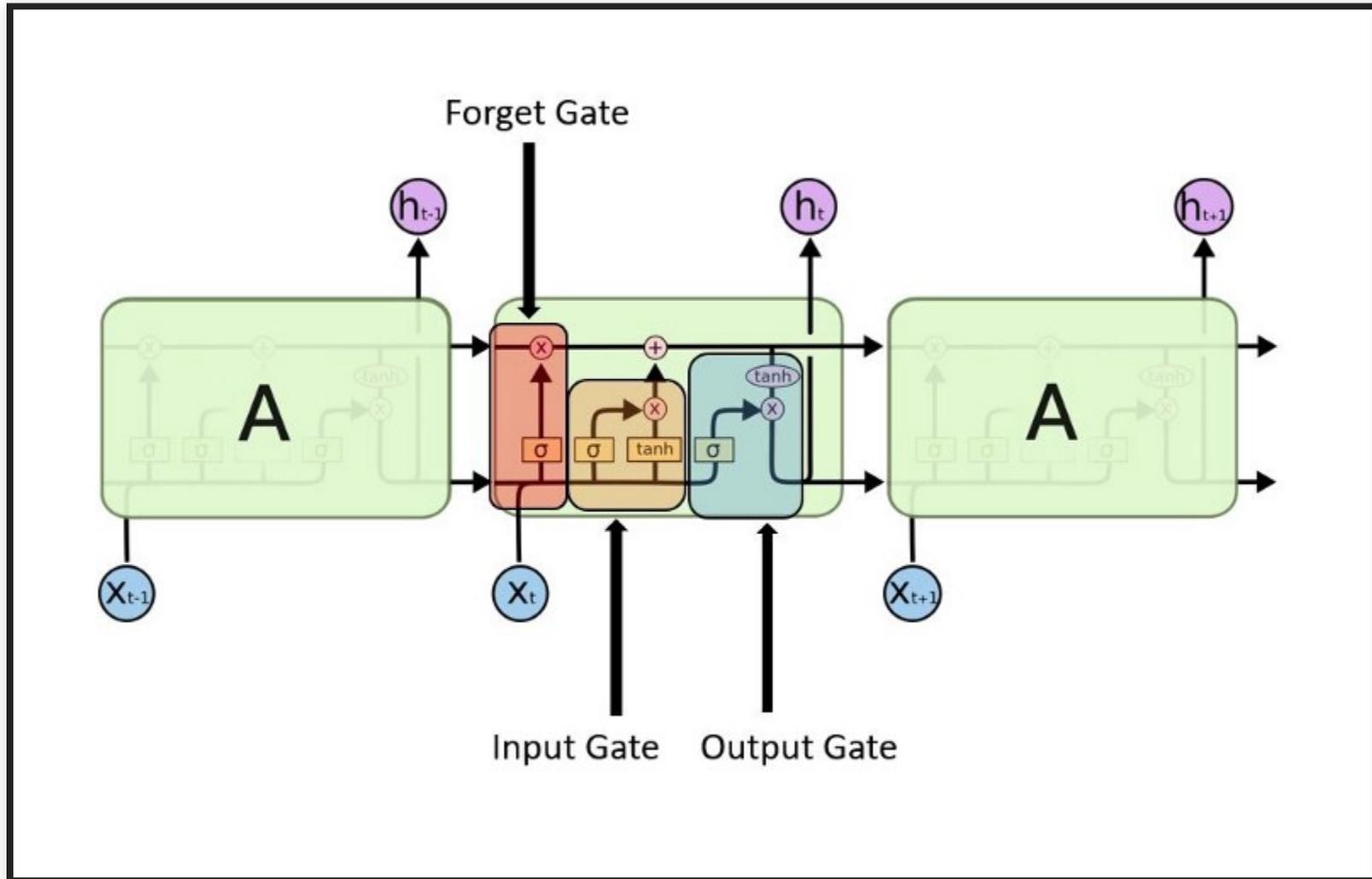


# LSTMS

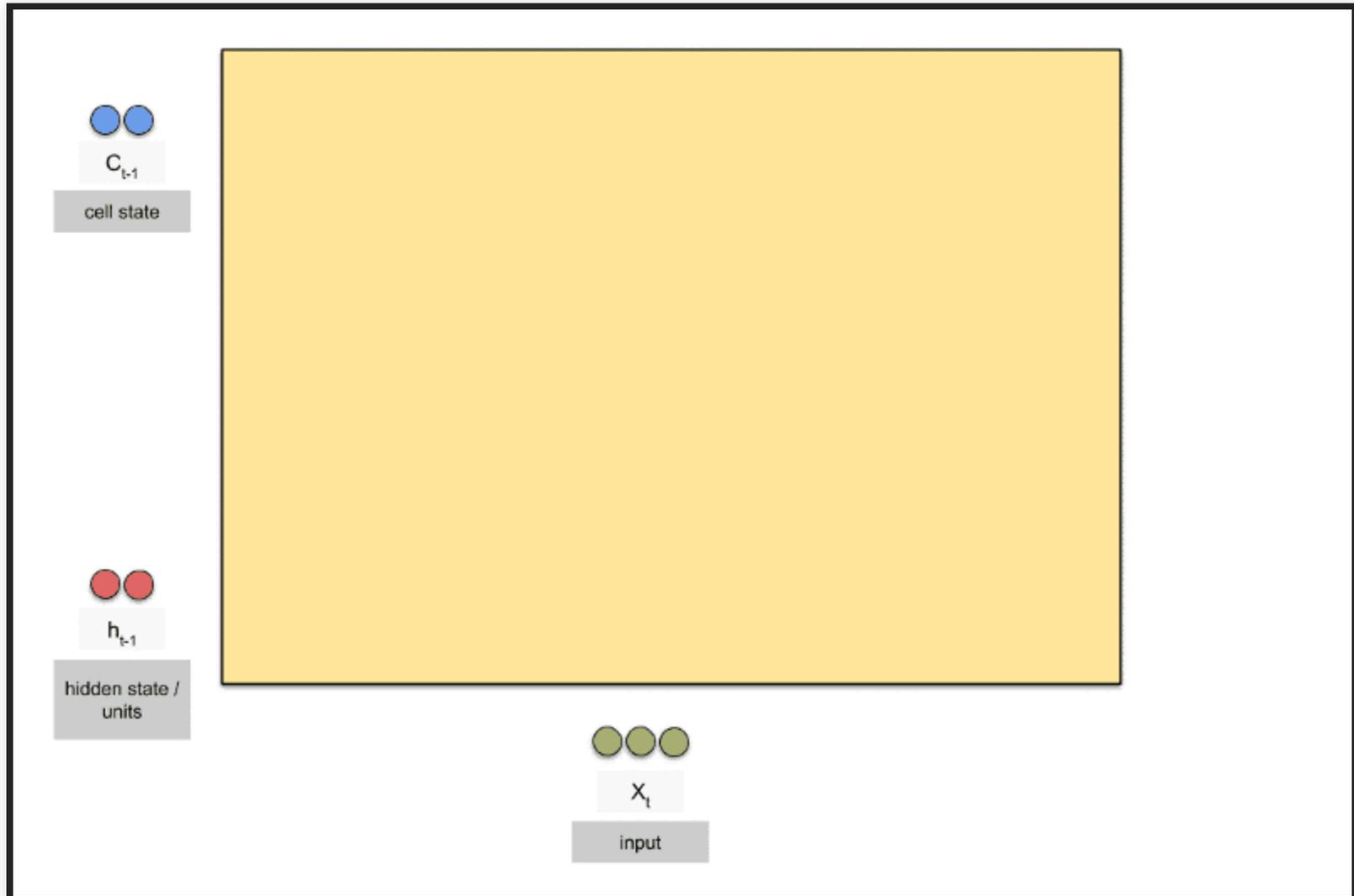
Há três portais em um bloco LSTM:

- forget gate
- memory gate (input gate)
- output gate

# LSTMS



# LSTMS

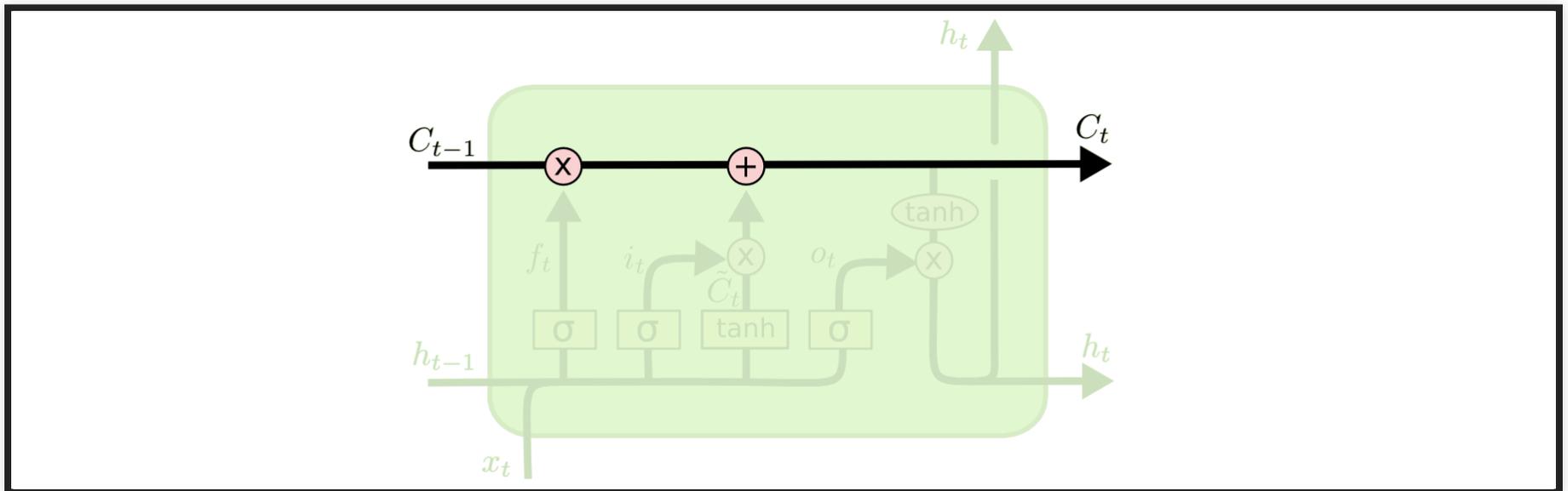


# LSTMS

- O estado interno ( $h$ ) carrega a informação sobre o que uma célula ( $C$ ) leu ao longo do tempo.
- Essa informação pode ser usada no tempo presente, de modo que uma função de custo não dependa apenas dos dados que está vendo neste instante, mas também dados vistos anteriormente.

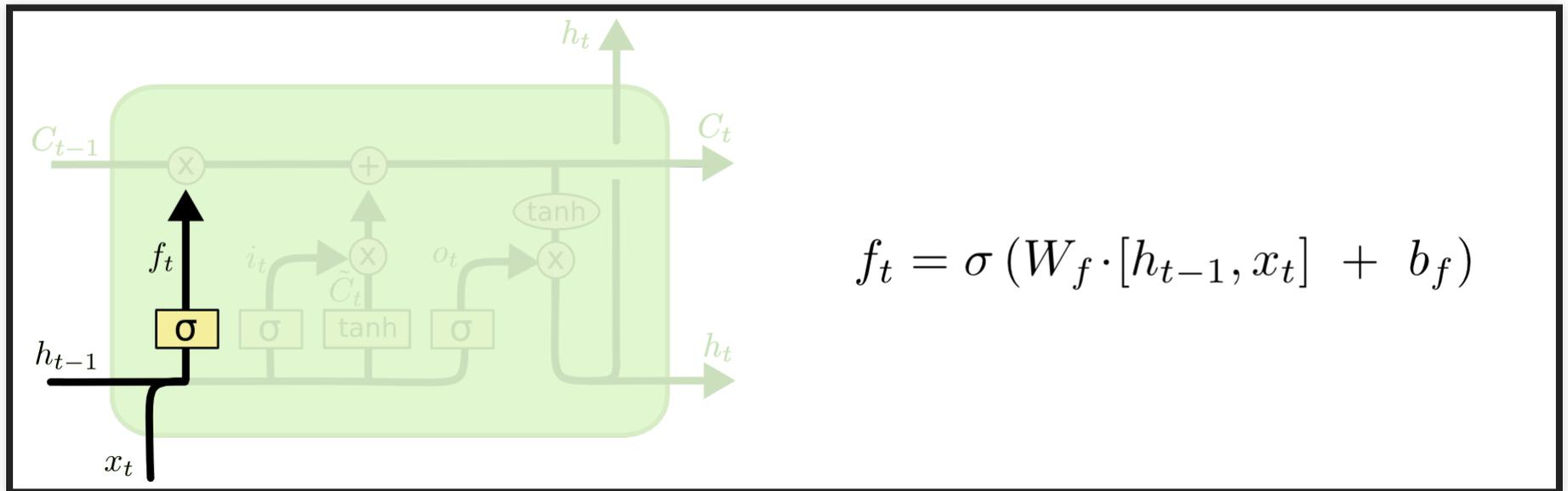
# LSTMS

- O "tubo" superior é o tubo de memória, cuja entrada é a memória antiga  $C_{t-1}$  (um vetor).



# LSTMS - FORGET GATE

- A conexão que se liga por baixo é  $f_t$ , o vetor resultante do portal esquecer (forget gate).

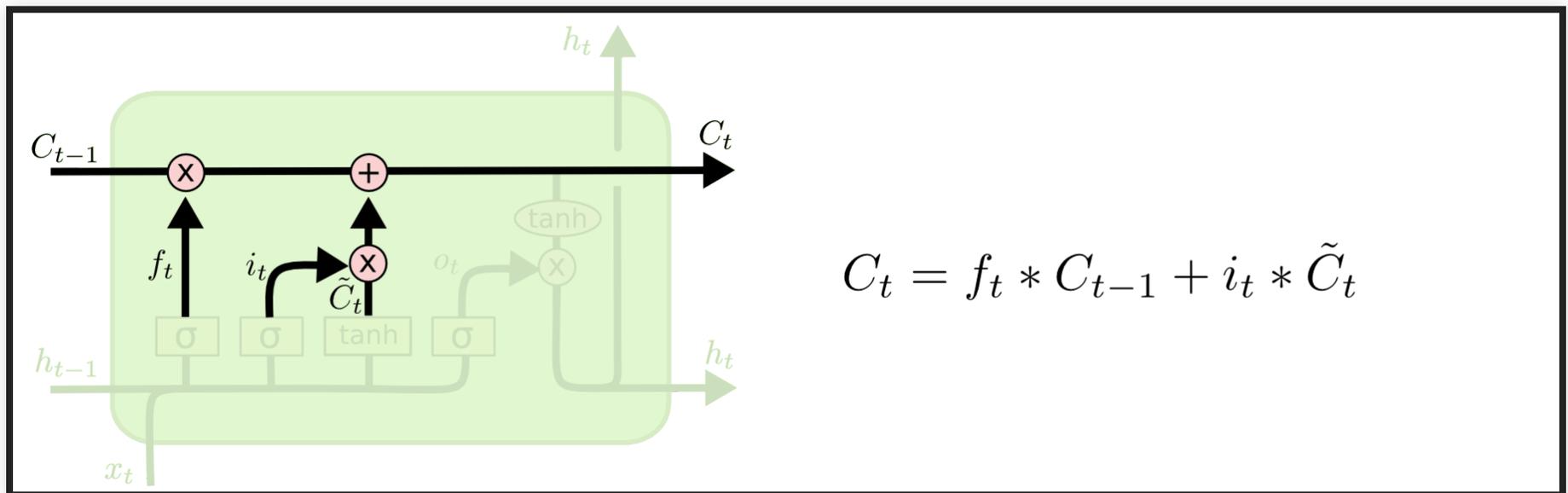


# LSTMS - FORGET GATE

- Se  $C_{t-1}$  for multiplicado por um vetor  $f_t$  cujas componentes são próximas de
  - 0, isso significa que a RNN quer esquecer a maior parte da memória antiga.
  - 1, isso significa que a RNN deixa a memória antiga passar.

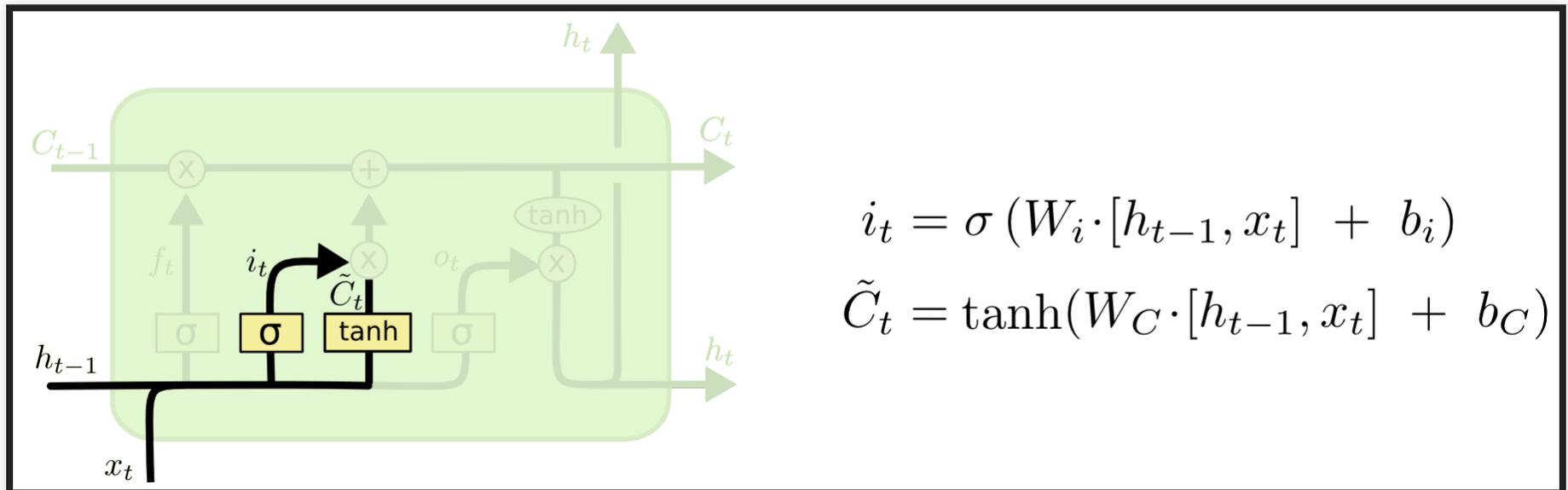
# LSTMS - MEMORY GATE

- A segunda operação aplicada ao fluxo de memória é uma soma.
- A soma mescla as memórias antiga e nova.



# LSTMS - MEMORY GATE

- A válvula abaixo da soma controla o quanto a nova memória deve ser mesclada com a antiga.



$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

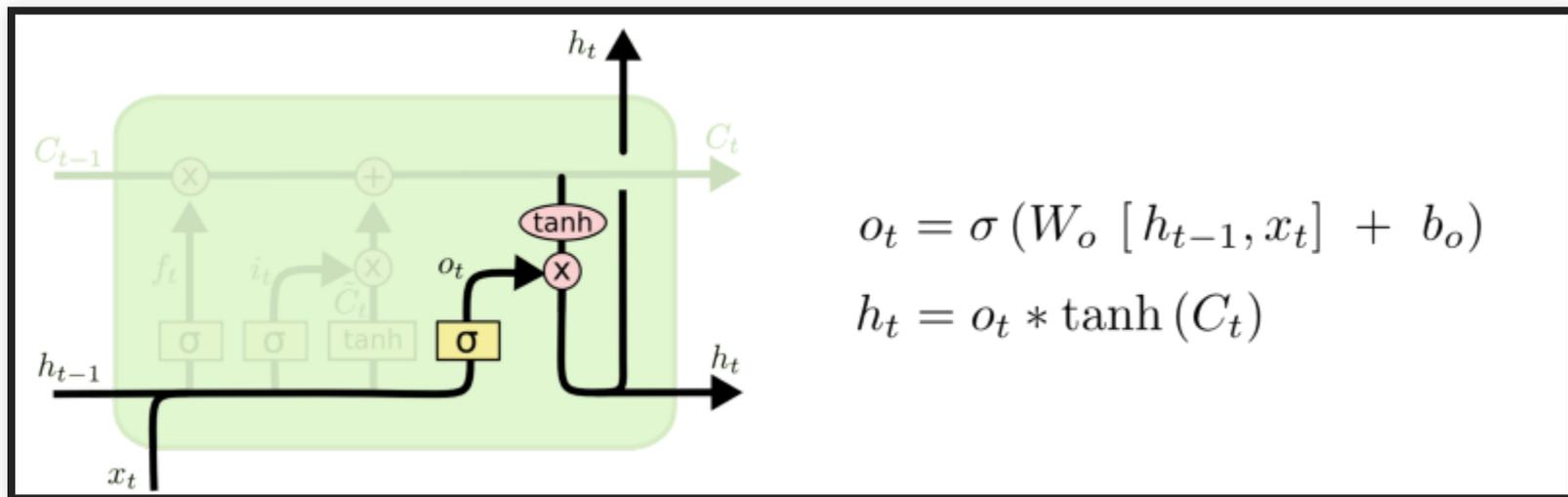
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

# LSTMS - MEMORY GATE

- Esse portal controla o quanto a nova memória influencia a memória anterior.
- Permite adicionar informação à memória da célula.
- A RNA desse portal recebe as mesmas entradas que as do forget gate.

# LSTMS - OUTPUT GATE

- A última porta  $h_t$ , a saída para o bloco LSTM.



# LSTMS - OUTPUT GATE

- Este passo possui uma válvula de saída que é controlada pelos seguintes componentes:
  - nova memória,
  - saída anterior  $h_{t-1}$ ,
  - entrada  $X_t$ ,
  - vetor bias.

# CNN VS. LSTMS

- RNNs clássicas sobrescrevem o estado interno.
- LSTMs adicionam ao estado interno.
- Unidades LSTM proveem à RNN células de memória, que possuem operações para manipular a memória:
  - leitura,
  - gravação e
  - reinicialização.

# LSTMS

- Geff Hinton sobre o artigo que propôs as LSTMs:

*if you can understand the paper, you are better than many people in machine learning. It took 10 years until people understand what they were talkink about”.*

# LSTMS

- O interesse vem aumentando consideravelmente nos últimos anos
- Citações do Google acadêmico



# APLICAÇÕES

Geração de textos:

<https://reiinakano.github.io/tfjs-lstm-text-generation/>

# APLICAÇÕES

- Geração de rótulos para imagens



A close up of a hot dog on a bun.



A bath room with a toilet and a bath tub.



A vase filled with flower sitting on a table.

# APLICAÇÕES

- Reconhecimento de desenhos  
<https://quickdraw.withgoogle.com/#>