



Universidade de Brasília

Departamento de Ciência da Computação



Long Short-Term Memory

Curso ENAP - Processamento de Linguagem Natural

Prof. Dr. Vinícius Ruela Pereira Borges

`viniciusrpb@unb.br`

Brasília-DF, 2024

- Esses slides foram redigidos e produzidos pelo Prof. Dr. Vinícius R. P. Borges;
- Material didático de referência:
 - Capítulo 9 do livro “Speech and Language Processing. Daniel Jurafsky & James H. Martin, 2021.”
 - Slides do curso “CS224n: Natural Language Processing with Deep Learning” - Stanford University

- Long Short-Term Memory
- Modelo de Linguagem LSTM
- Fluxo Bidirectional
- Montando Arquiteturas mais Complexas

Long Short-Term Memory

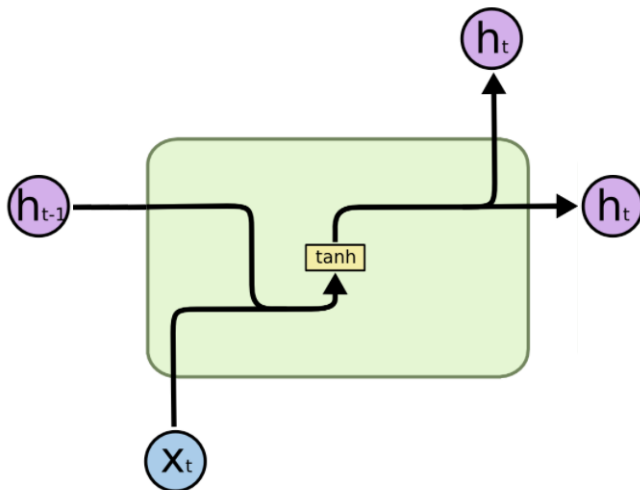


Long Short-Term Memory

- Long Short-Term Memory (LSTM) ¹ é um tipo de rede neural recorrente (RNN) modelada para lidar com o problema do “gradiente que encolhe” encontrado nas RNNs tradicionais;
- O modelo matemático que define a LSTM é mais complexo do que a RNN tradicional:
 - objetivo é permitir o fluxo de gradiente durante o *Backpropagation*;
 - mais parâmetros são necessários para regular o fluxo de informação que entra e sai da rede.

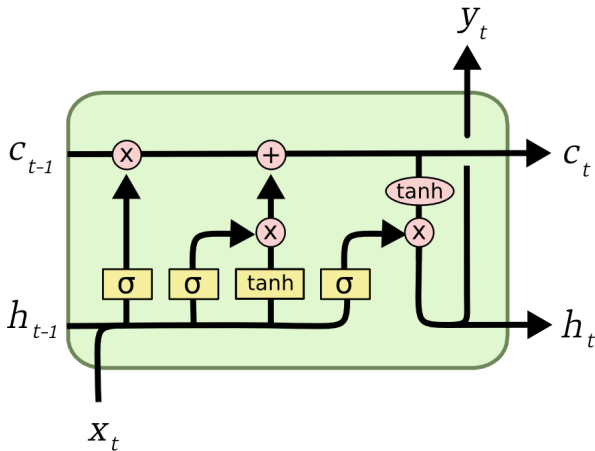
¹Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Relembrando a RNN



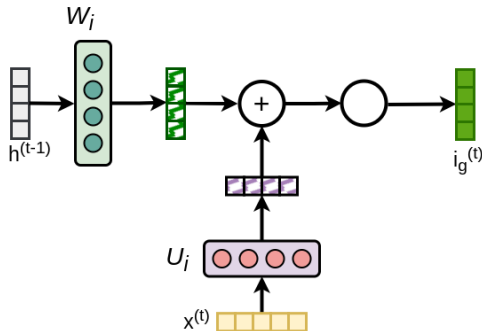
- Uma RNN de Elman com função de ativação “tangente hiperbólica”.

Long Short-Term Memory



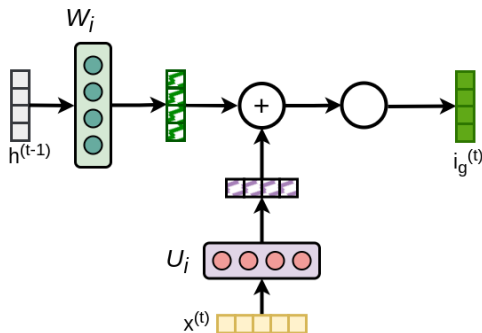
- A LSTM com várias... vamos detalhar a seguir!

LSTM: Input Gate



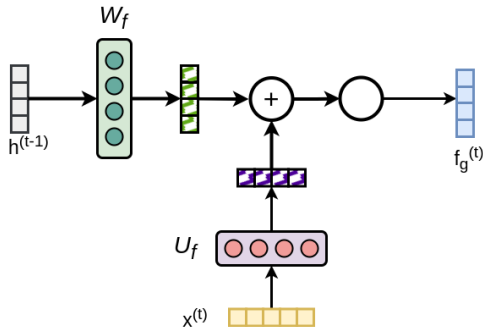
- Objetivo: determina as informações que serão adicionadas ao estado da célula $c^{(t)}$

LSTM: Input Gate



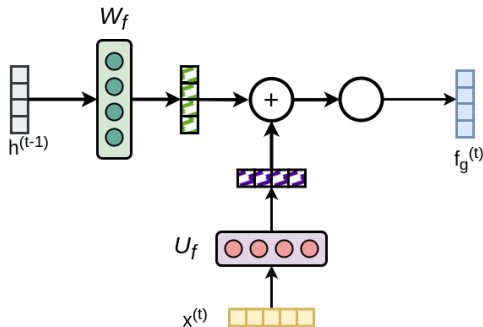
$$i_t = \sigma(U_i x^{(t)} + W_i h^{(t-1)} + b_i) \quad (1)$$

LSTM: Forget Gate



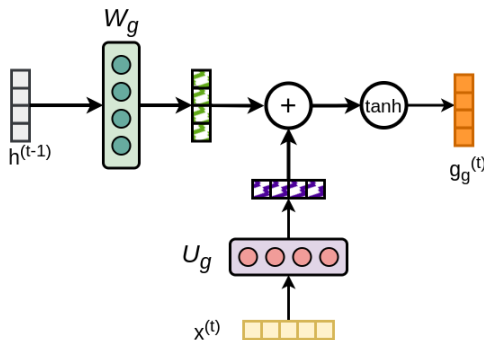
- Objetivo: remoção de informações armazenadas na célula em relação às entradas $h^{(t-1)}$ e $x^{(t)}$;

LSTM: Forget Gate



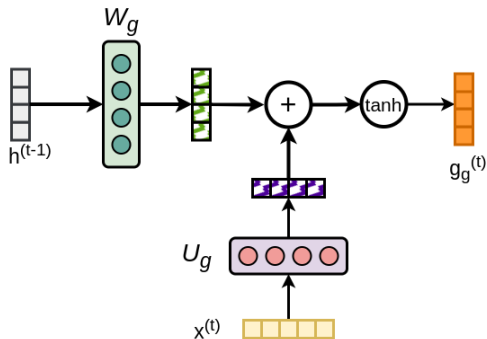
$$f_t = \sigma(U_f x^{(t)} + W_f h^{(t-1)} + b_f) \quad (2)$$

LSTM: Atualização de Célula



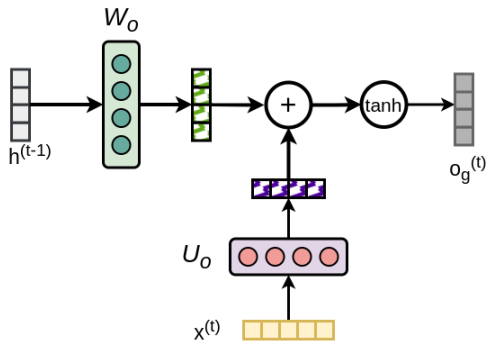
- Objetivo: trata-se de uma estimativa das informações que devem ser adicionadas ao estado da célula no tempo $c^{(t)}$

LSTM: Atualização de Célula



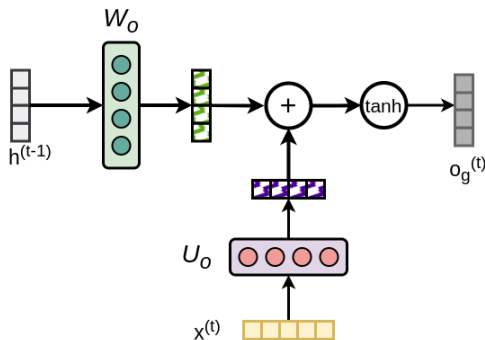
$$g_t = \tanh(U_g x^{(t)} + W_g h^{(t-1)} + b_g) \quad (3)$$

LSTM: Output Gate



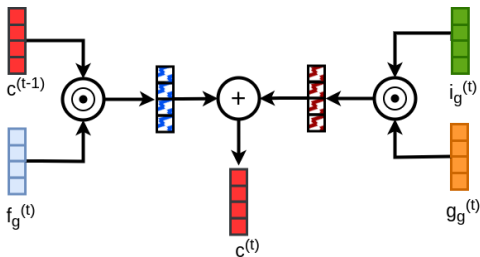
$$o_t = \sigma(U_o x^{(t)} + W_o h^{(t-1)} + b_o) \quad (4)$$

LSTM: Output Gate



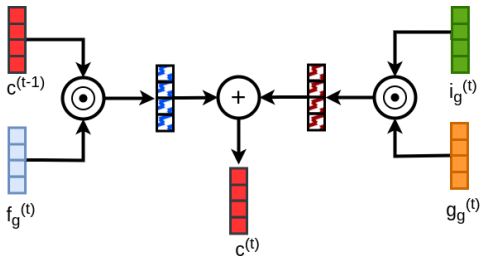
- Objetivo: determina a quantidade de informação útil que está célula que deve ser gerada na saída $h^{(t)}$;

LSTM: Estado da Célula



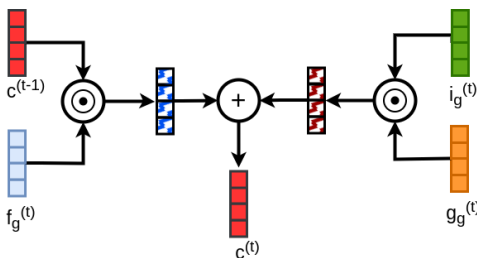
$$c^{(t)} = f^{(t)} \odot c^{(t-1)} + i^{(t)} \odot g^{(t)} \quad (5)$$

LSTM: Estado da Célula



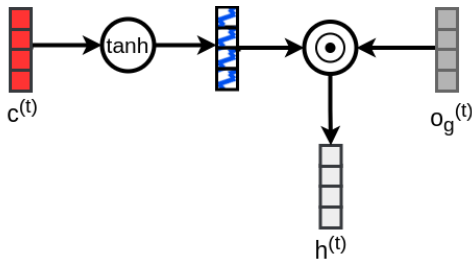
- Objetivo: representa a memória a longo prazo da LSTM;

LSTM: Estado da Célula



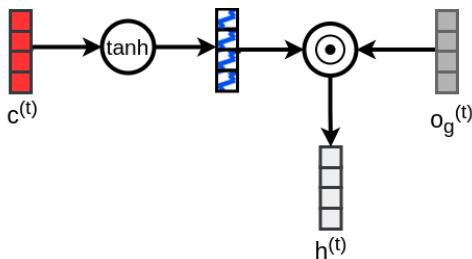
- Repare que seu valor é o quanto se quer esquecer do estado anterior e o quanto que se quer aprender da informação que chega a LSTM.

LSTM: Estado Interno da Célula



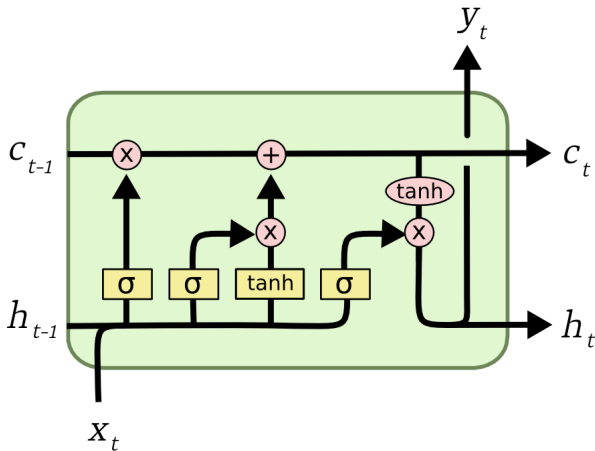
$$h^{(t)} = o^{(t)} \odot \tanh(c^{(t)}) \quad (6)$$

LSTM: Estado Interno da Célula



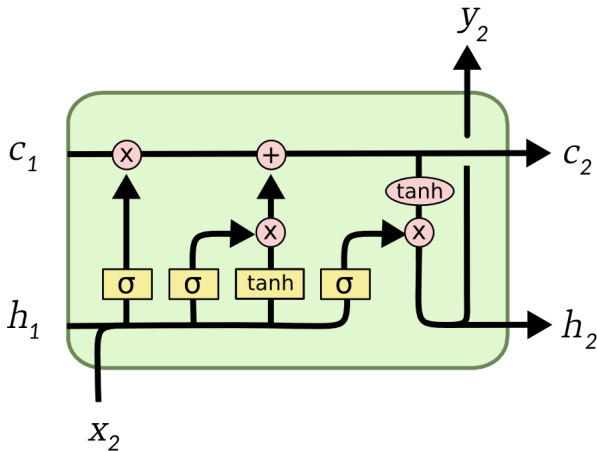
- Objetivo: gerar a saída da célula, que compreende uma sumarização toda a entrada processada até o instante de tempo t .

Long Short-Term Memory



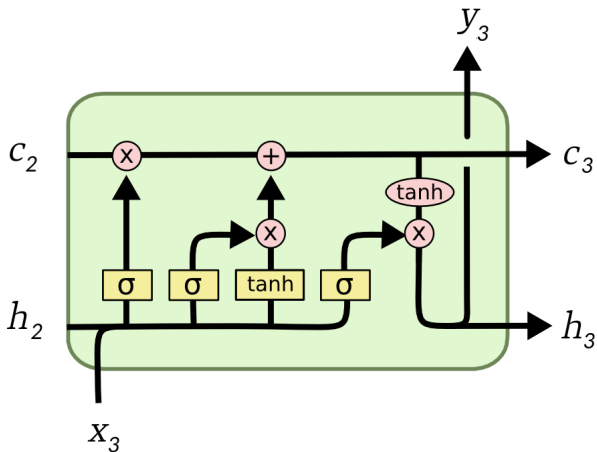
Eu não acredito em duendes

Long Short-Term Memory



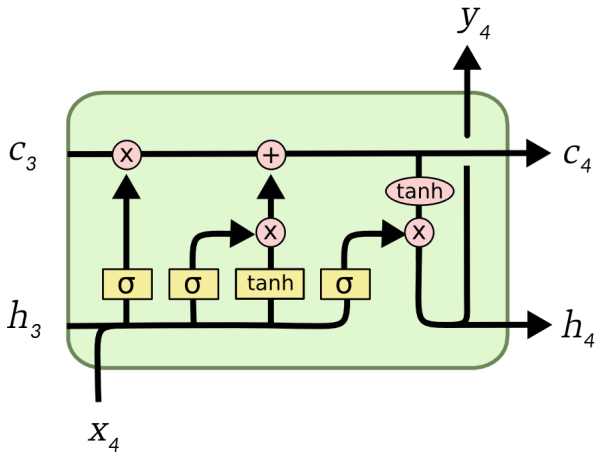
Eu não acredito em duendes

Long Short-Term Memory



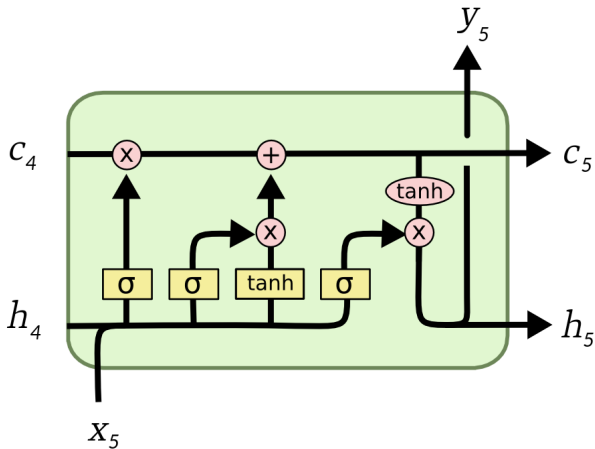
ão acredito em duendes

Long Short-Term Memory



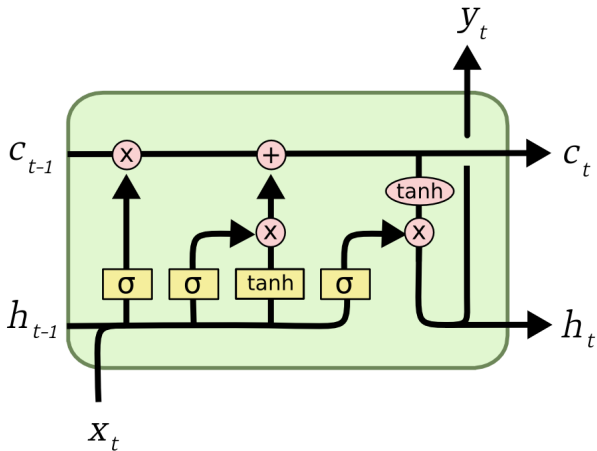
redito em duendes

Long Short-Term Memory



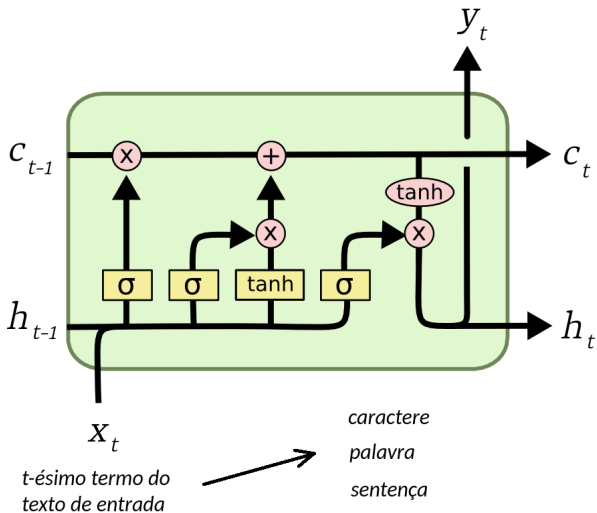
em duendes

Long Short-Term Memory

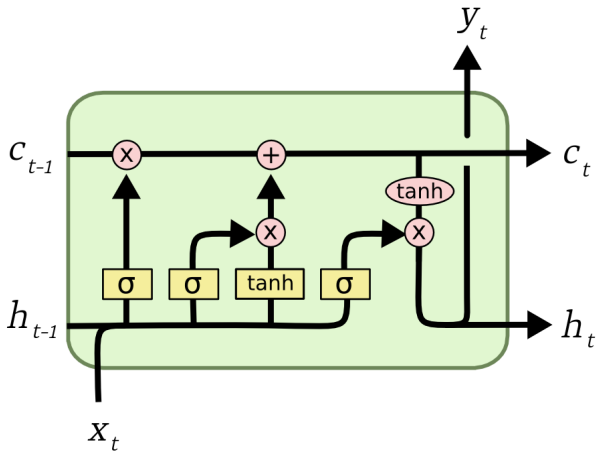


x_t
t-ésimo termo do
texto de entrada

Long Short-Term Memory

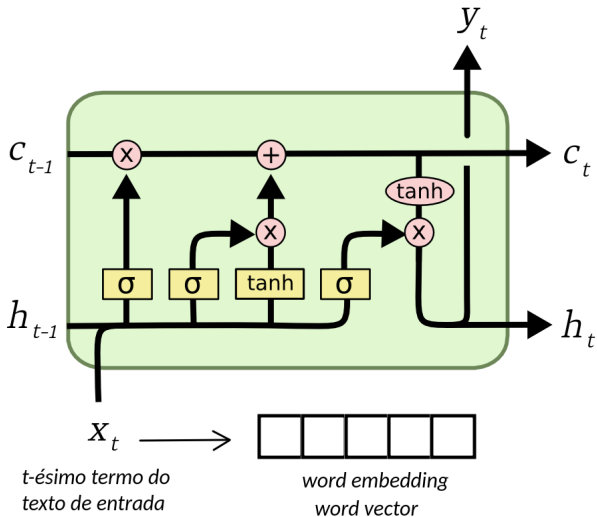


Long Short-Term Memory

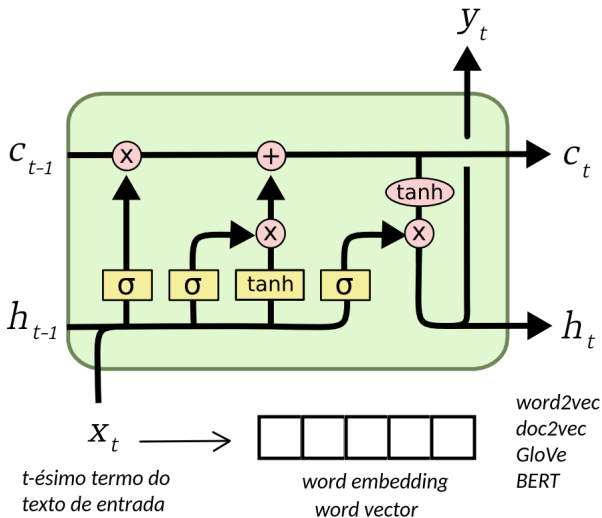


x_t
t-ésimo termo do
texto de entrada

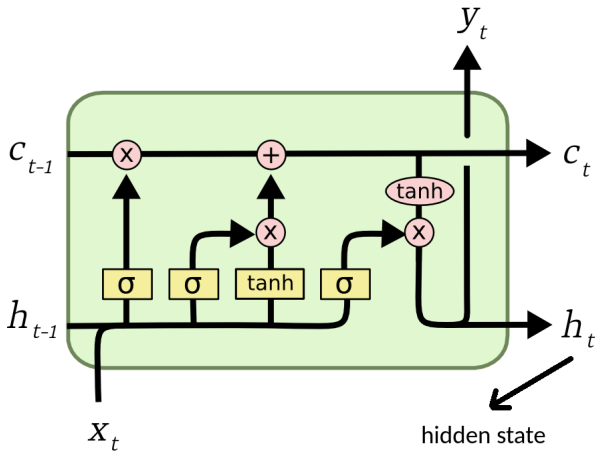
Long Short-Term Memory



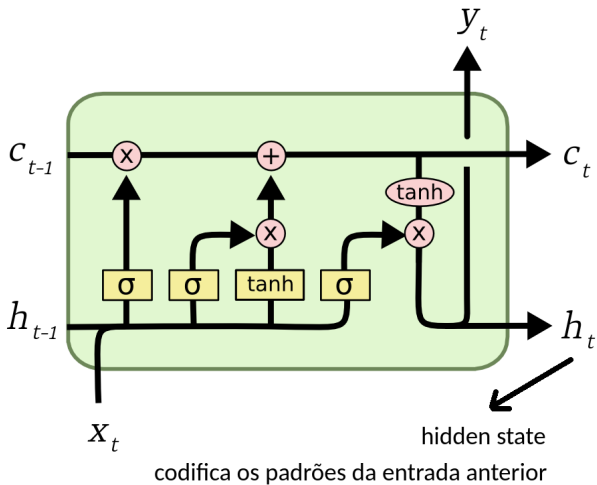
Long Short-Term Memory



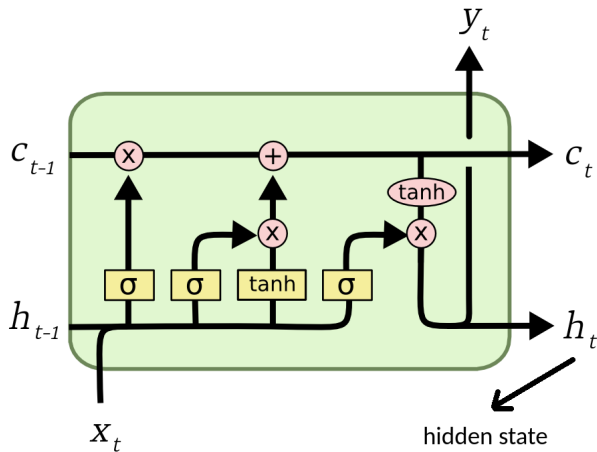
Long Short-Term Memory



Long Short-Term Memory

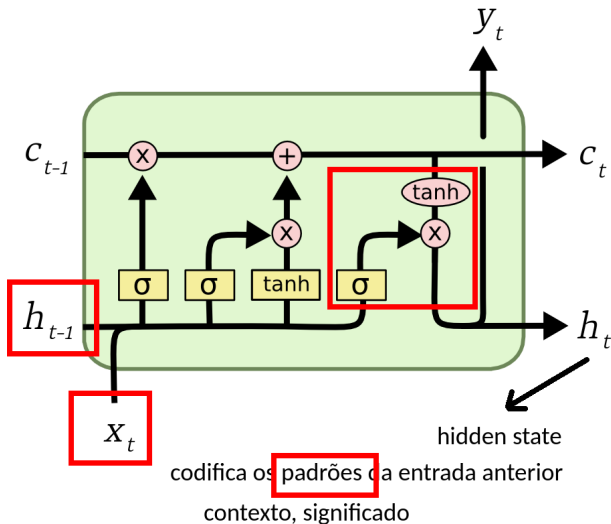


Long Short-Term Memory

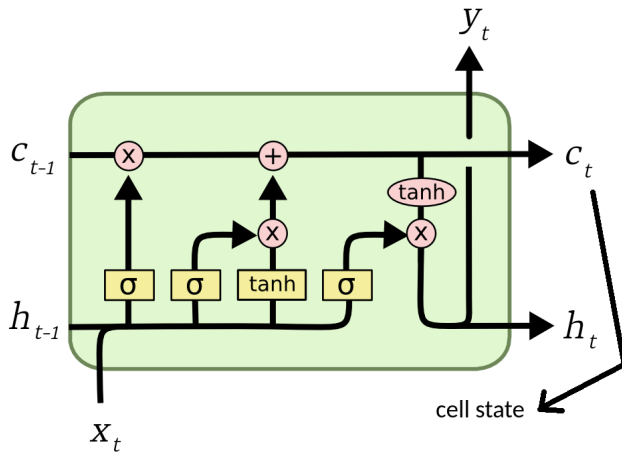


hidden state
codifica os padrões da entrada anterior
contexto, significado

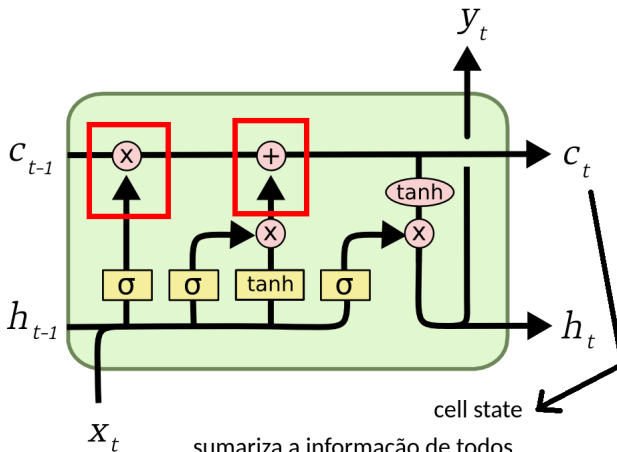
Long Short-Term Memory



Long Short-Term Memory

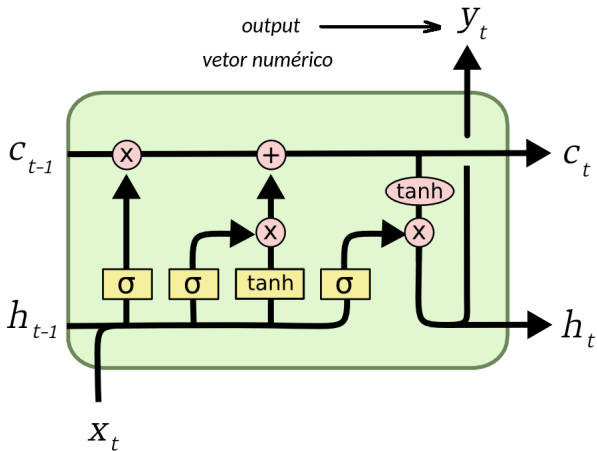


Long Short-Term Memory

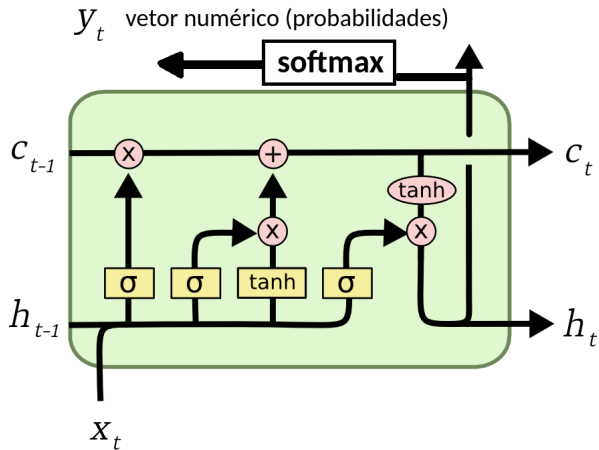


sumariza a informação de todos
os termos processados anteriormente
da entrada

Long Short-Term Memory



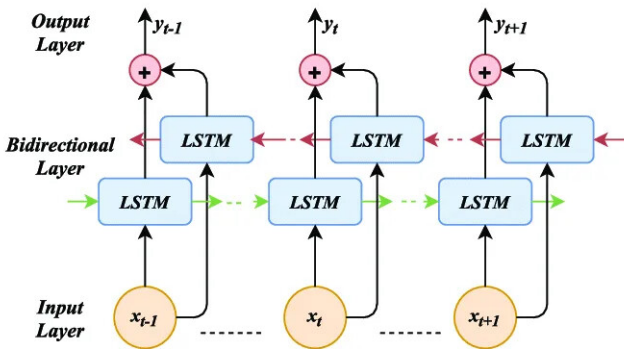
Long Short-Term Memory



Long Short-Term Memory

- Treinamento ocorre por meio do algoritmo Backpropagation through time (BTT);
- Deve-se otimizar os hiperparâmetros:
 - taxa de aprendizado;
 - quantidade de neurônios artificiais interno à célula.

Bidirectional Long Short-Term Memory (BiLSTM)





Universidade de Brasília

Departamento de Ciência da Computação



Redes Neurais Recorrentes

Curso ENAP - Processamento de Linguagem Natural

Prof. Dr. Vinícius Ruela Pereira Borges

`viniciusrpb@unb.br`

Brasília-DF, 2024