

# Rede Neural

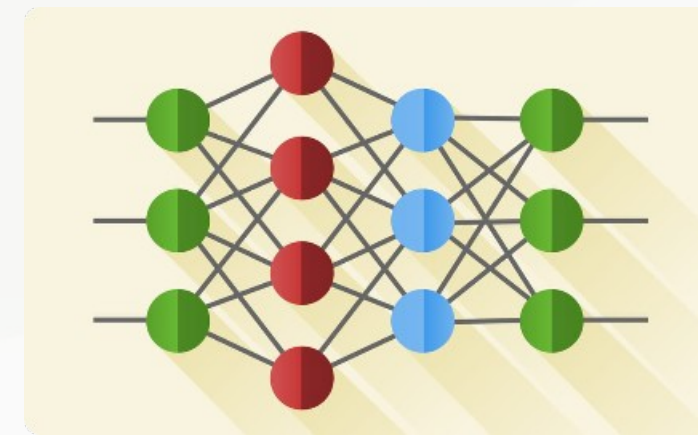
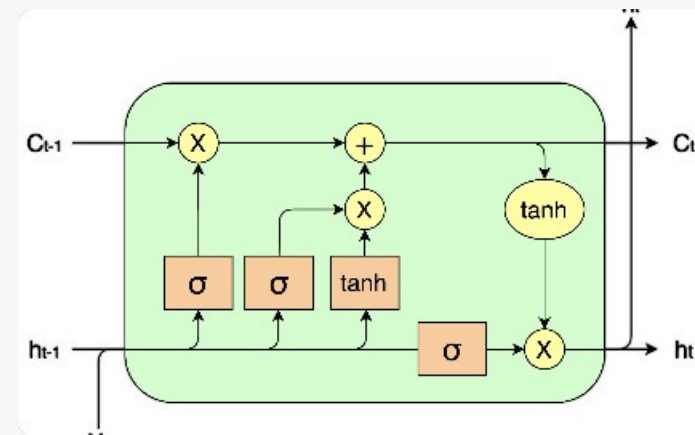
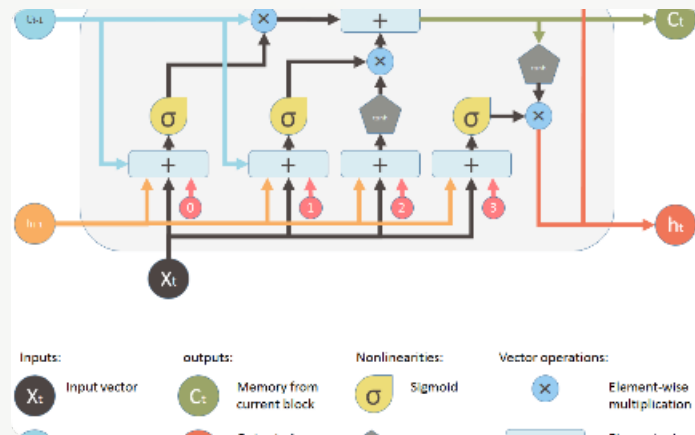
## Long Short-Term Memory LSTM

As redes neurais são modelos matemáticos inspirados no funcionamento do cérebro humano. Nesta apresentação, vamos explorar as redes neurais LSTM, um tipo especial de rede neural.

**Antonio Leal**

Setembro/2023

# O Que é LSTM?



## Definição

Uma rede neural recorrente (RNN) que pode lembrar informações anteriores por um período de tempo mais longo.

## Funcionamento

Usa três portões - entrada, saída e esquecimento - para processar informações, evitando problemas com gradiente.

## Aplicações

LSTM pode ser usada para modelar uma variedade de problemas, incluindo previsão de séries temporais e processamento de texto.

# Por que e quando surgiram?

1

## Motivação

As redes neurais recorrentes (RNN) simples, apresentavam o problema do desaparecimento/explosão do gradiente. Conforme número de camadas e iterações aumentavam, a rede mostrava-se ineficiente.

2

## Quando surgiram?

Foi proposta por Sepp Hochreiter e Jürgen Schmidhuber em um artigo publicado em 1997, intitulado "Long Short-Term Memory."

# Solução para vanishing gradient

Algoritmo Truncate BackPropagation Through Time (TBTT): com limite de etapas do backpropagation

LSTM: aplicada com sucesso na modelagem de longas sequências com camadas ocultas (células de memória interconectadas)

# Primeiros experimentos

REDES JORDAN (Michael Jordan-1986):

Primeira rede recorrente baseada na MLP.

Nesta arquitetura, além das 3 camadas usuais, é criada a camada de unidades de contexto.

Memórias ocultas são alimentadas pela camada de saída

REDES ELMAN (Jeffrey Elman-1990):

Diferente da feedforward tradicional (camada única) porque tem camadas de unidades de memória ou estados ocultos (hidden states).

As camadas de entrada e saída interagem com o ambiente externo, mas não as outras duas (oculta e de contexto);

# Estrutura das redes LSTM

## Unidade de memória

É a principal componente das redes LSTM e consiste em um conjunto de células e portas.

## Células de memória

Armazenam informações ao longo do tempo e são atualizadas através de portas específicas.

## Portas

Controlam o fluxo de informações dentro da rede LSTM, permitindo o acesso e a atualização das células de memória.

A principal idéia do LSTM é usar a mesma conexão de loops de feedbacks para eventos que aconteceram a muito tempo e ontem para fazer previsão para amanhã.



# Compreendendo os mecanismos de gates

## **Gate de entrada (input gate)**

Receptor de novo input.  
O gate de entrada decide quais informações novas serão adicionadas à memória da célula.

## **Gate de esquecimento (forget gate)**

O gate de esquecimento controla a quantidade de informação antiga que deve ser descartada da memória.

## **Gate de saída (output gate)**

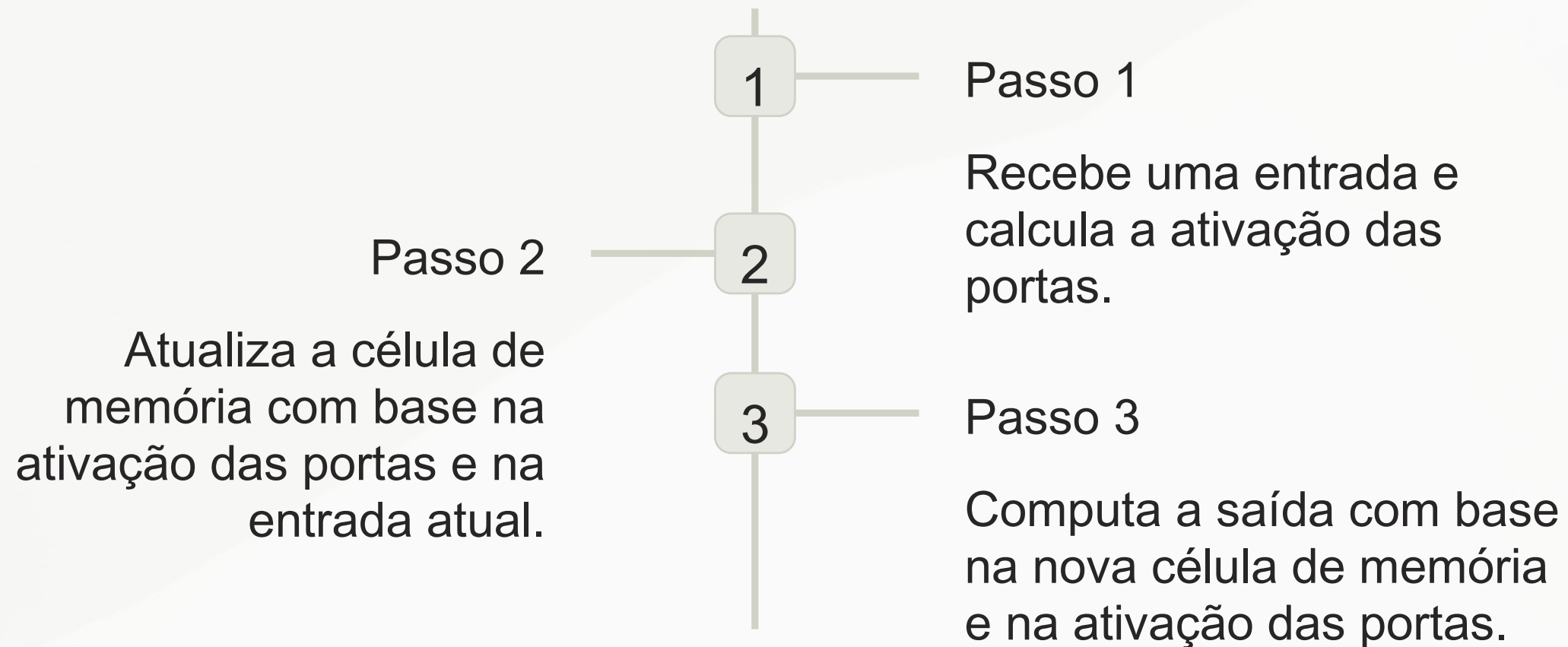
O gate de saída determina quais informações serão lidas da memória da célula e usadas na previsão ou na próxima etapa da sequência.

# Desafios na Utilização do Portão do Esquecimento

- Overfitting: alguns modelos podem memorizar em excesso os dados de treinamento, prejudicando a generalização para novas entradas.
- Hiperparâmetros: encontrar a combinação ideal dos hiperparâmetros da rede LSTM com Portão do Esquecimento pode ser desafiador.
- Grande quantidade de dados: o desempenho das redes LSTM com Portão do Esquecimento depende de um conjunto de dados amplo e diversificado.



# Funcionamento das redes LSTM



# Backpropagation Through Time - BPTT

## Definição

O algoritmo Backpropagation Through Time (BPTT) é um método para treinar redes neurais recorrentes.

## Gradientes

Os gradientes são propagados por cálculo retroativo em um desdobramento da rede ao longo do tempo, permitindo que os pesos sejam ajustados para minimizar o erro em todas as etapas.

## Atualização dos Pesos

Os pesos são atualizados usando um método de otimização, como Gradiente Descendente Estocástico.

# Real-Time Recurrent Learning -RTRL

## Definição

O algoritmo RLTL utiliza recompensas para otimizar o comportamento da rede neural LSTM, aprendendo a tomar melhores decisões ao longo do tempo e adaptando-se às mudanças do ambiente

## Gradientes

Os gradientes são calculados de forma recursiva para cada etapa de tempo individualmente, atualizando os pesos à medida que novos dados chegam. Ele não requer retropropagação ao longo de toda a sequência

## Atualização dos pesos

A ideia central do RTRL é calcular as derivadas locais das saídas da rede em relação aos pesos em cada passo de tempo e usar essas derivadas para atualizar os pesos.

# Eficiência computacional

## BPTT:

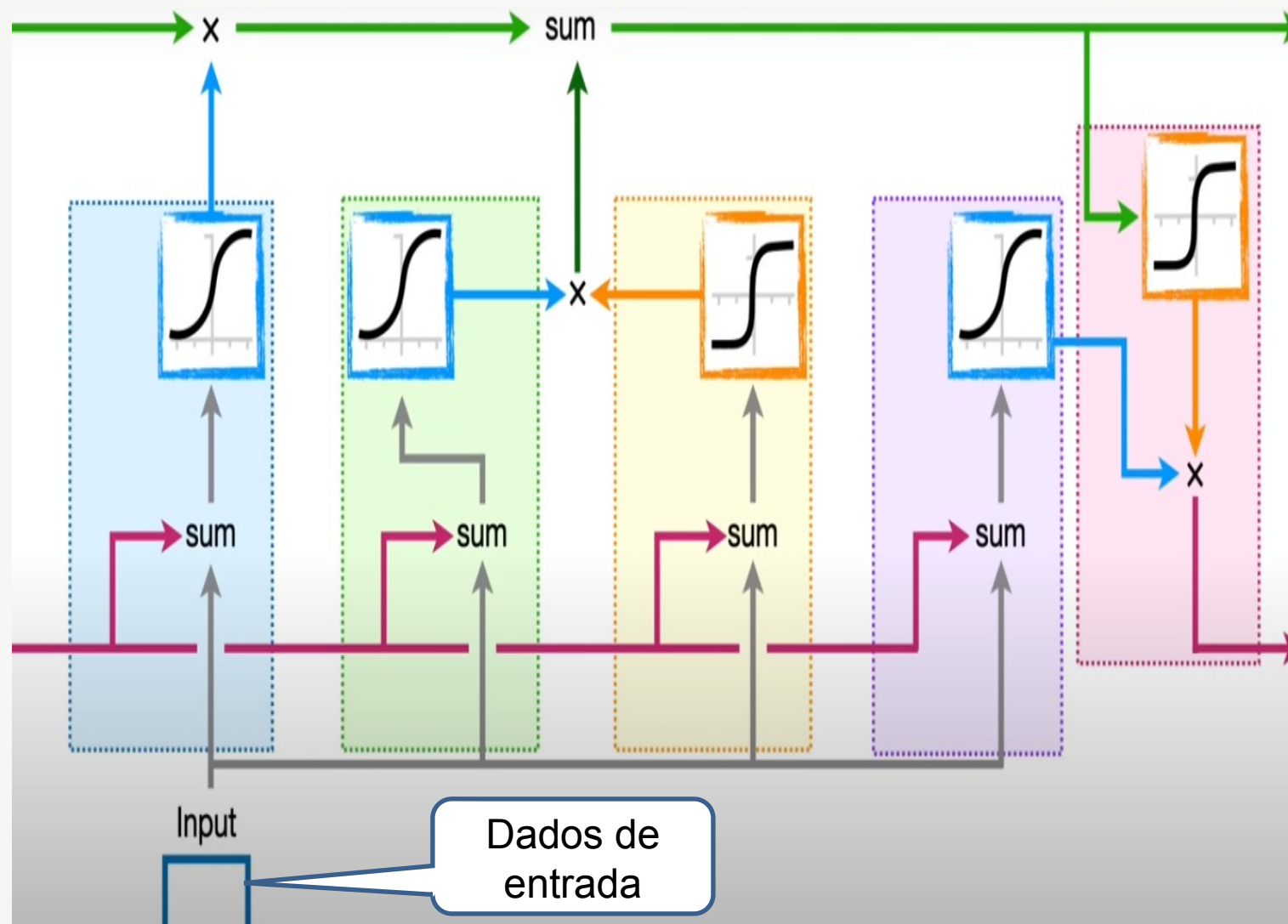
Pode ser mais eficiente em termos de memória em situações em que você tem lotes de dados disponíveis ao mesmo tempo e pode calcular gradientes em paralelo.

## RTRL:

Mais adequado para situações em que você precisa de atualizações em tempo real e não pode esperar por um lote completo de dados

A formulação original da LSTM usou uma combinação de BPTT e RTRL

# LSTM desdobrada ao longo do tempo

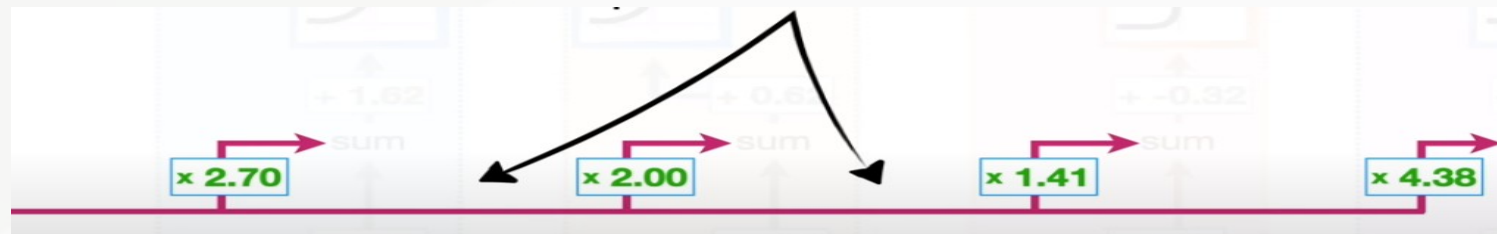


Caminho para  
memória de longo  
prazo

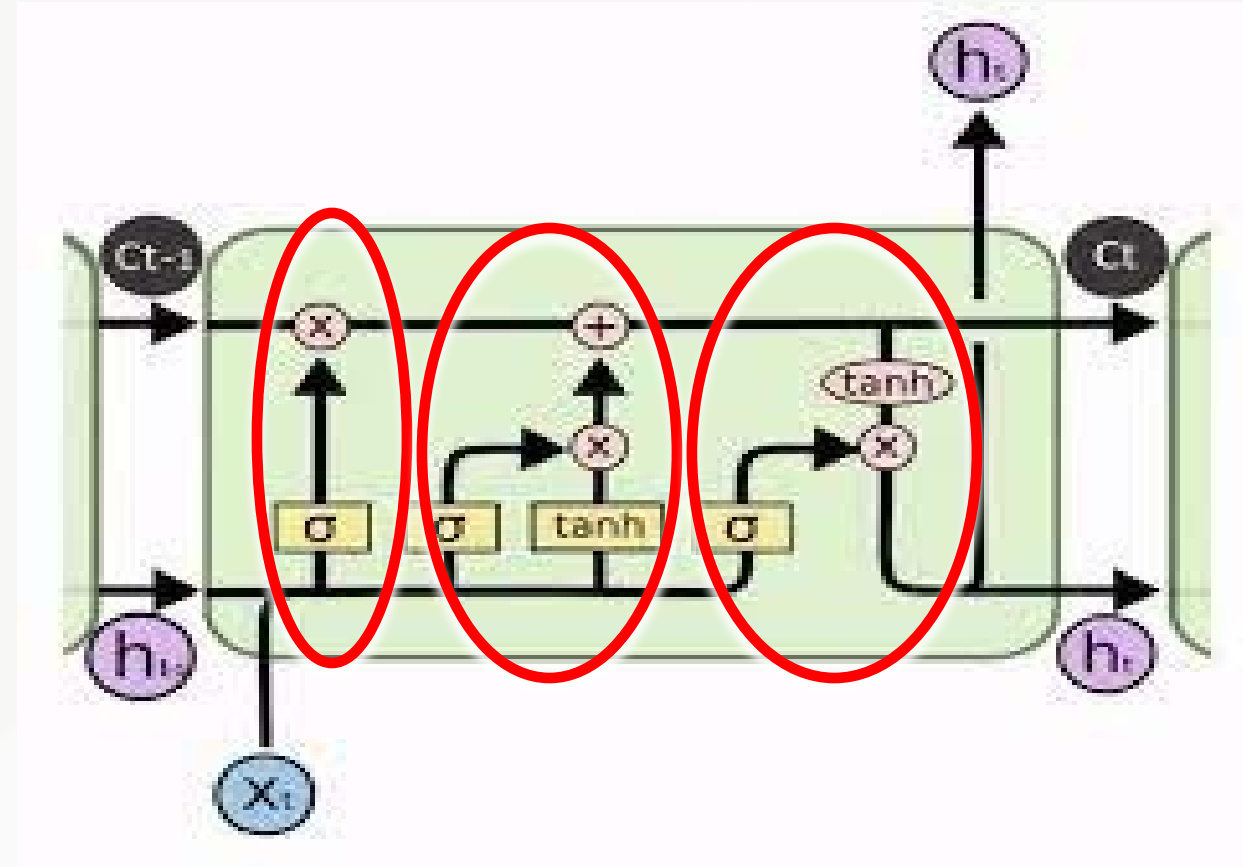
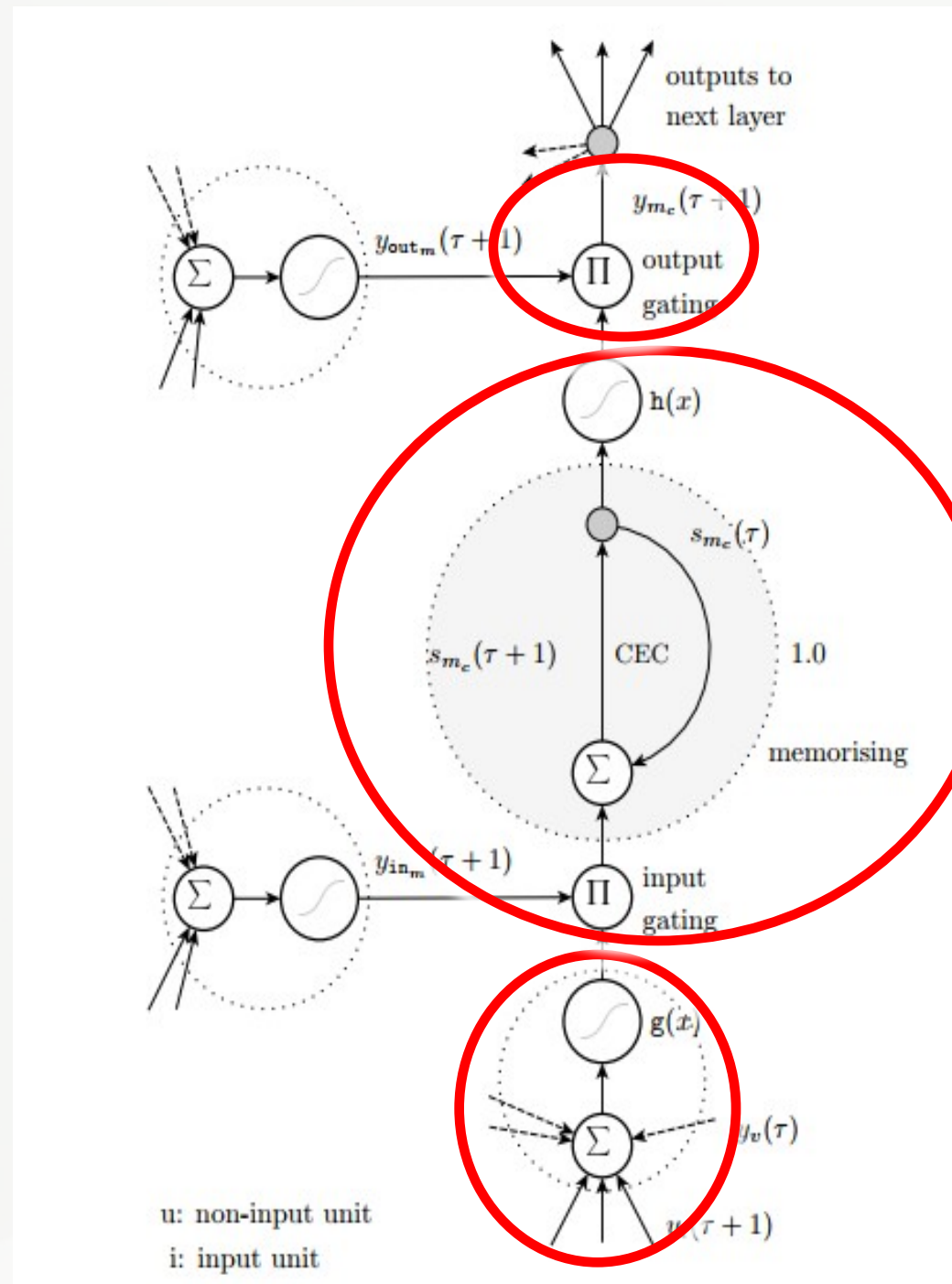
a linha verde é chamada de estado de célula, representa a memória de longo prazo, embora possa ser modificada pelos processos de multiplicação e adição, não há pesos ou vieses que possam modificá-la diretamente, isso evita a explosão ou desaparecimento do gradiente.

Caminho para  
memória de curto  
prazo

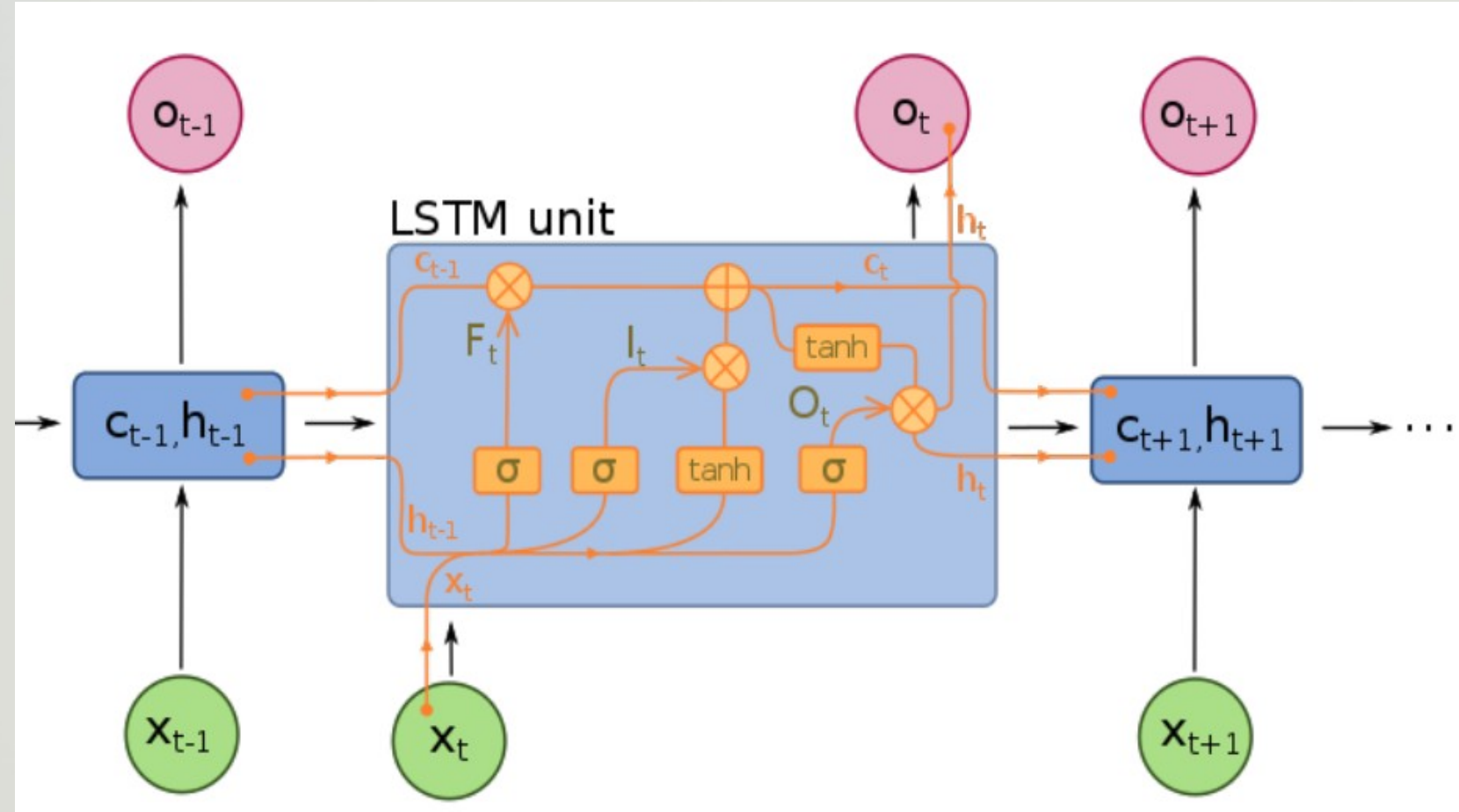
a linha rosa chamada de estado oculto representa as memórias de curto prazo, como podemos ver estão diretamente relacionados a pesos que podem modificá-las



# Estrutura Interna LSTM







## Equações que definem a estrutura interna de uma LSTM

$$I_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

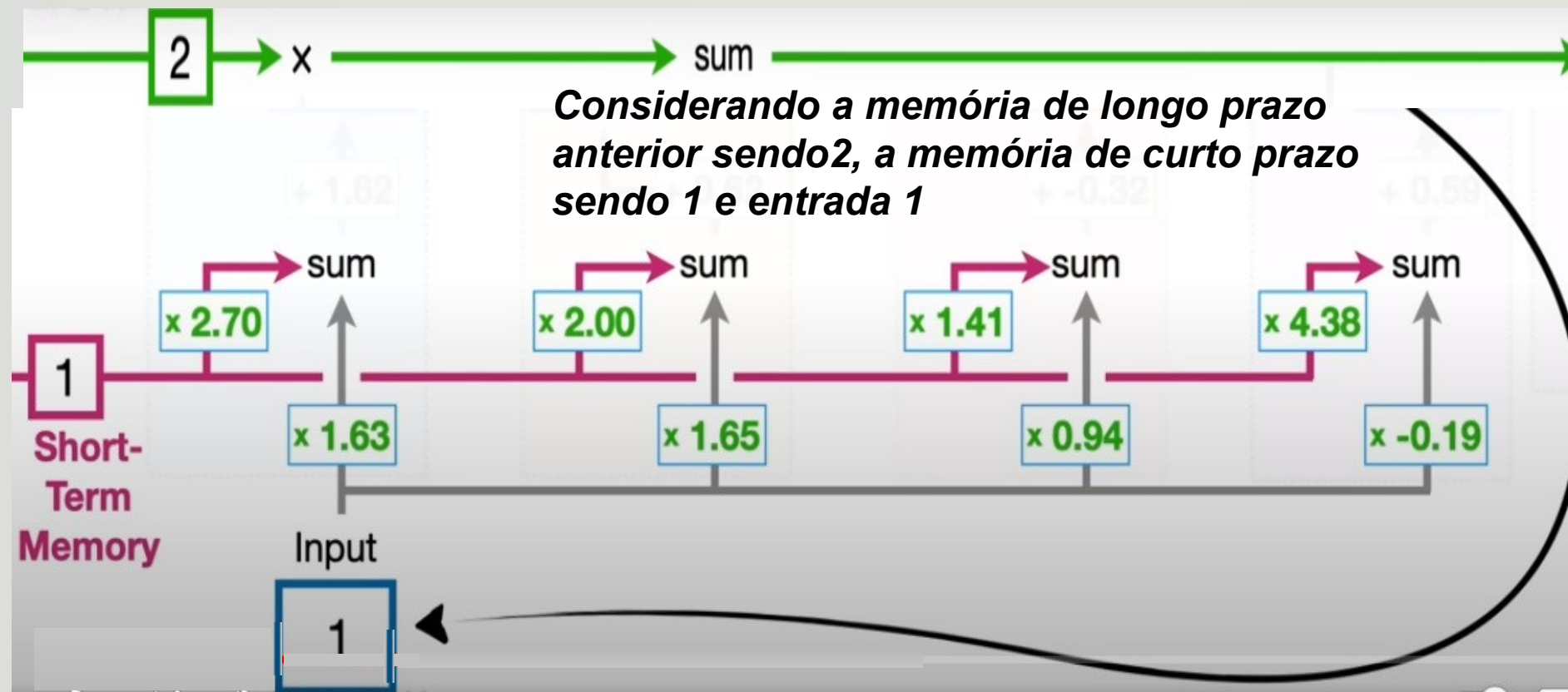
$$F_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

$$c_t = F_t \times c_{t-1} + I_t \times \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c)$$

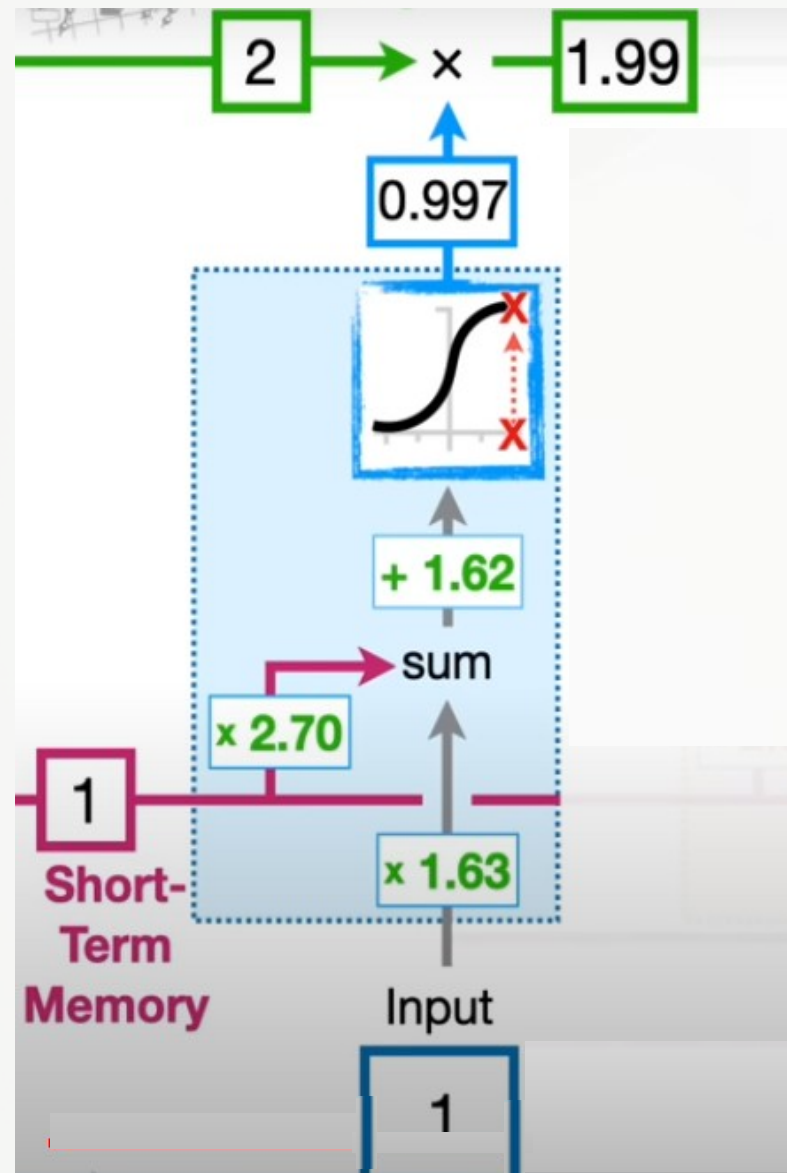
$$O_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$$

$$h_t = O_t \times \tanh(c_t)$$

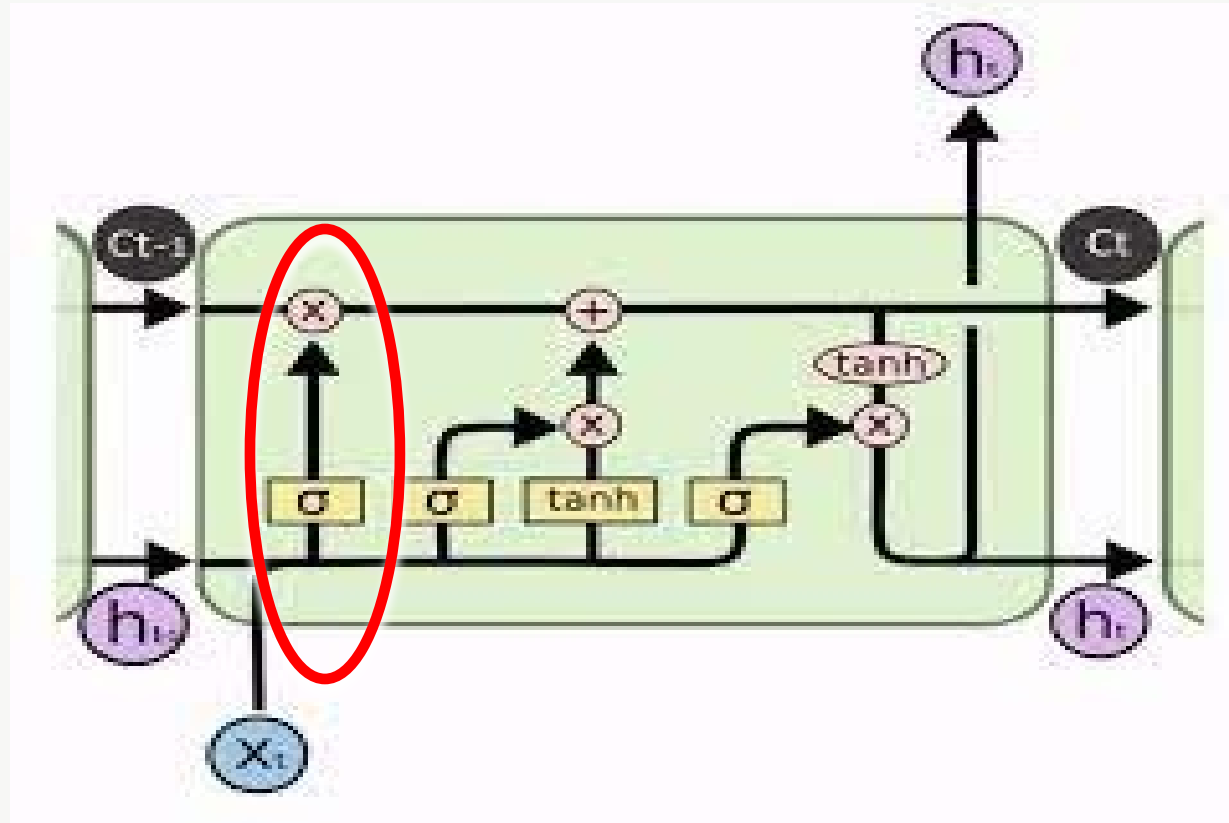
# Executando uma LSTM



# 1º. estágio da unidade de memória



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

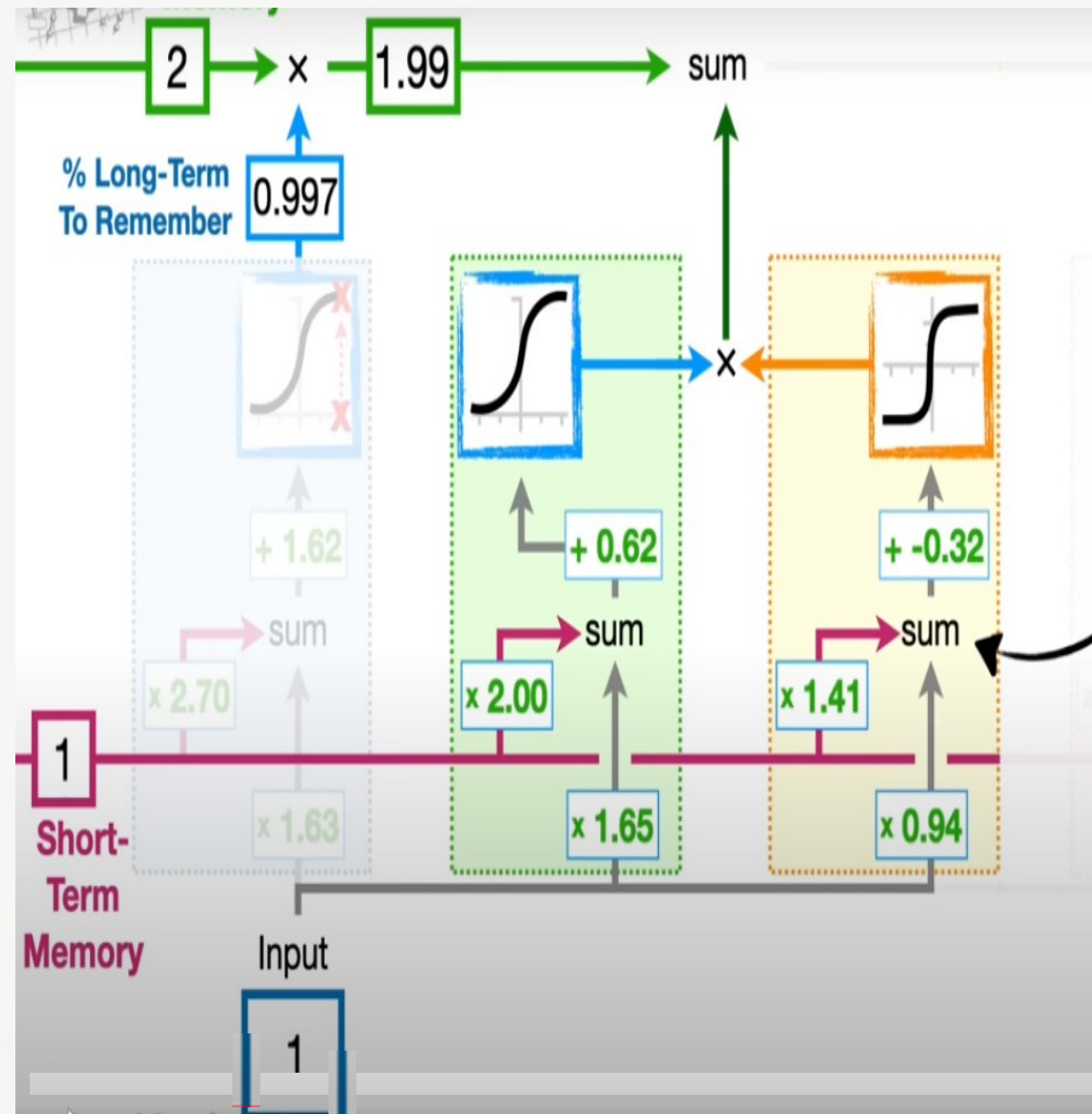


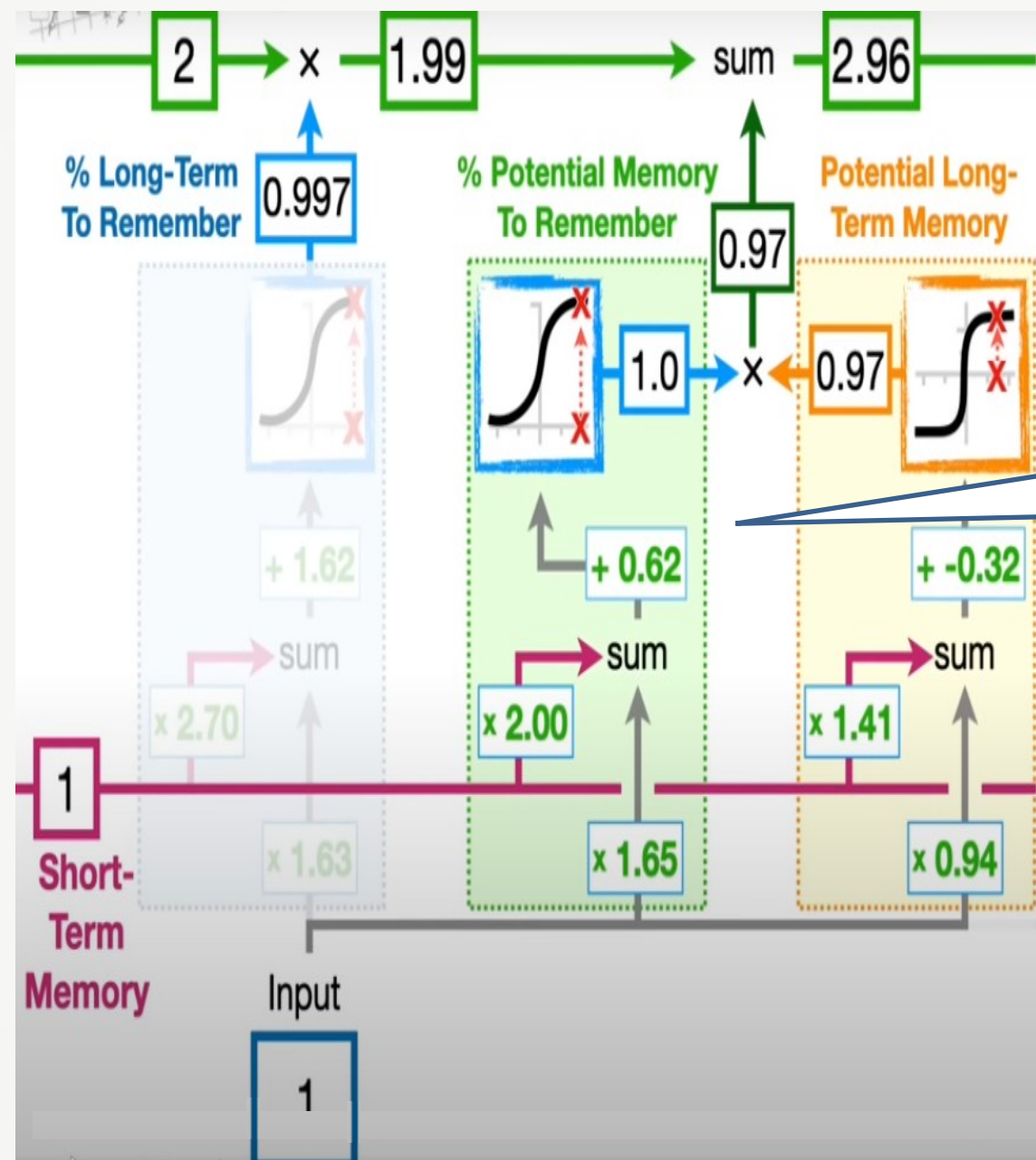
o resultado de 1.99 é o primeiro estágio, e reduz um pouco a memória de longo prazo. saída (0,997) determina a porcentagem da memória de longo prazo é lembrada.

Essa parte da unidade de memória é chamada de

**"Portão do esquecimento"**

## 2º. estágio da unidade de memória





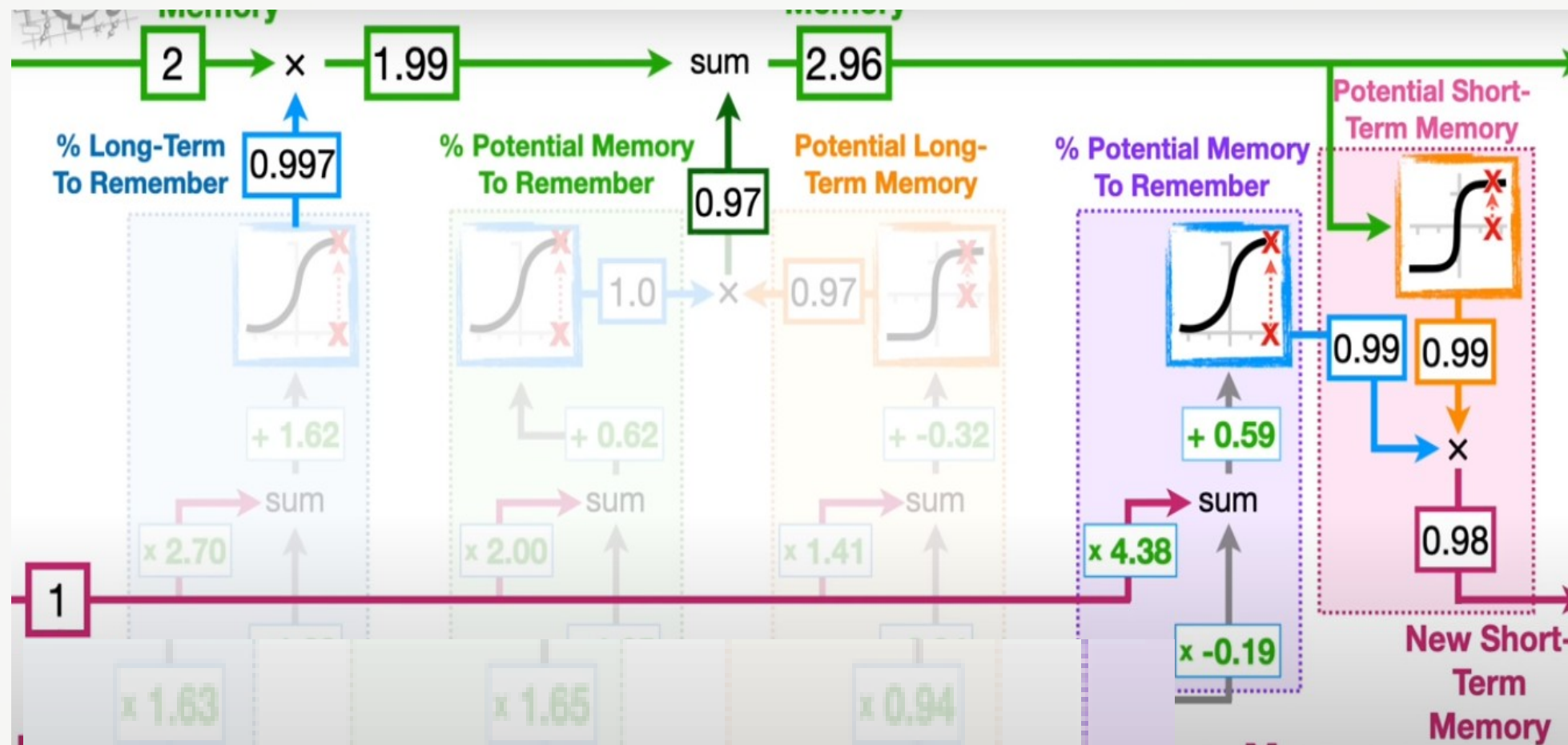
combina memória de curto prazo e a entrada para criar uma possível memória de longo prazo.

determina qual porcentagem da memória candidata deve ser adicionada à memória de longo prazo

Essa parte da unidade de memória é chamada de “**portão de entrada**”.



# Último estágio da unidade de memória

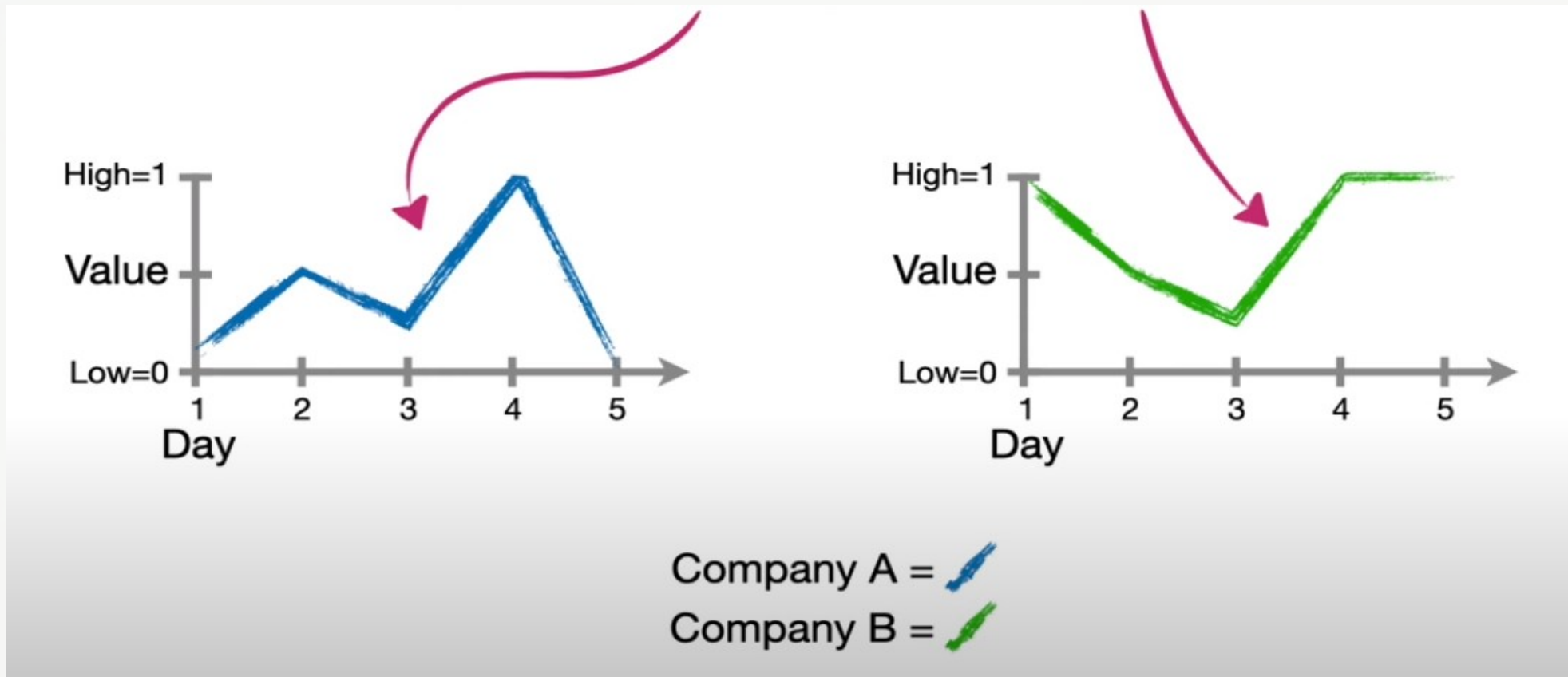


a nova memória de longo prazo (2,96) é a entrada para função tangente hiperbólica., o resultado dá 0,99. Esse valor representa a potencial memória de curto prazo.

A nova memória de curto prazo é a saída da LSTM esse estágio é chamado de "**Porta de Saida**"

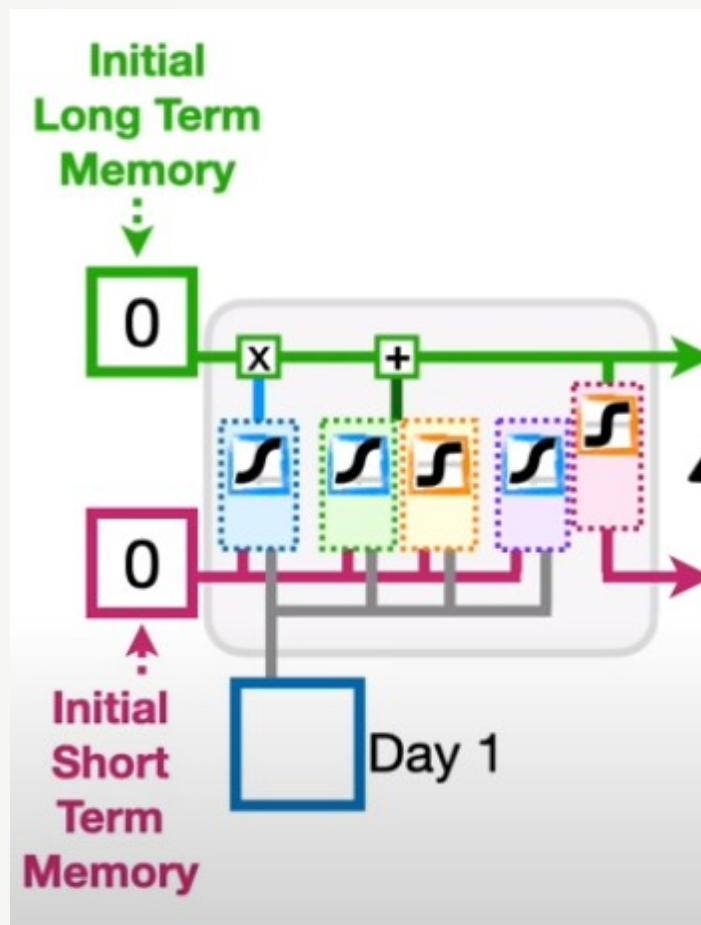


# Executando com dados reais

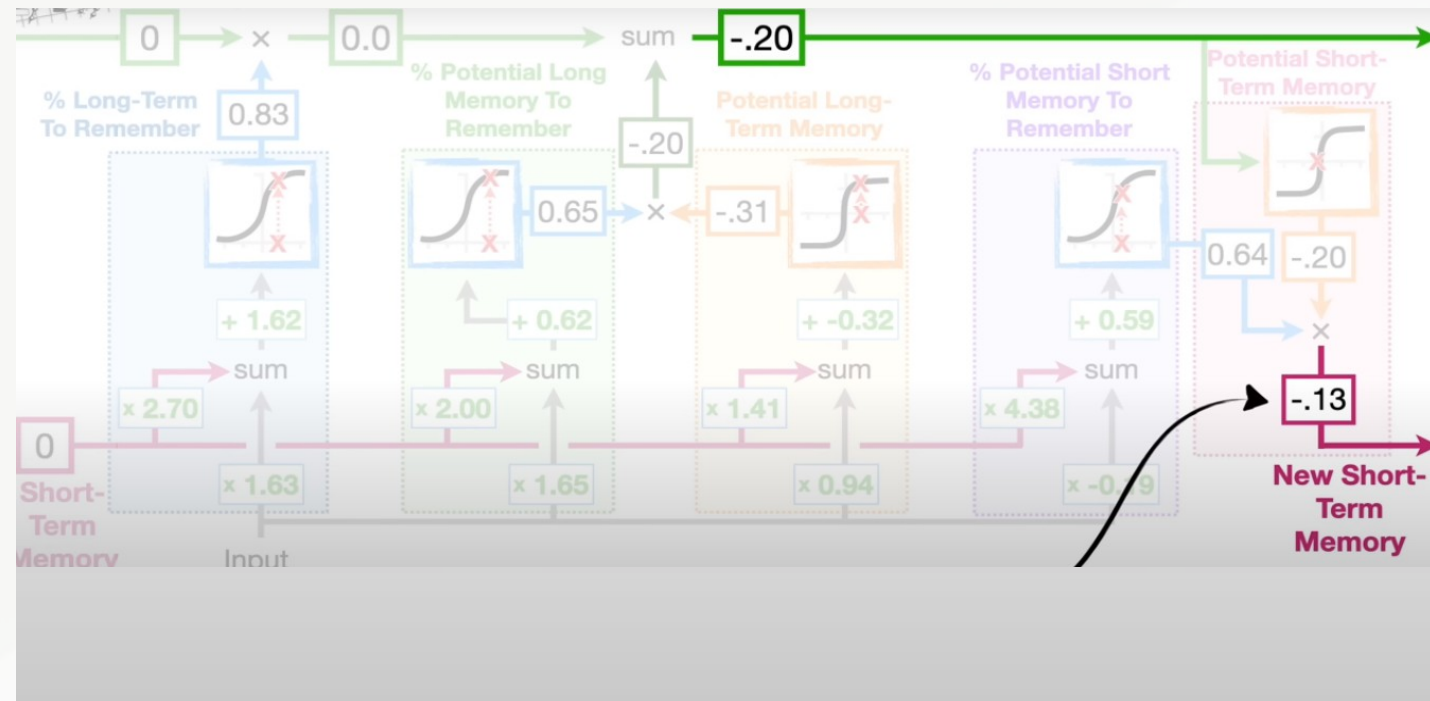
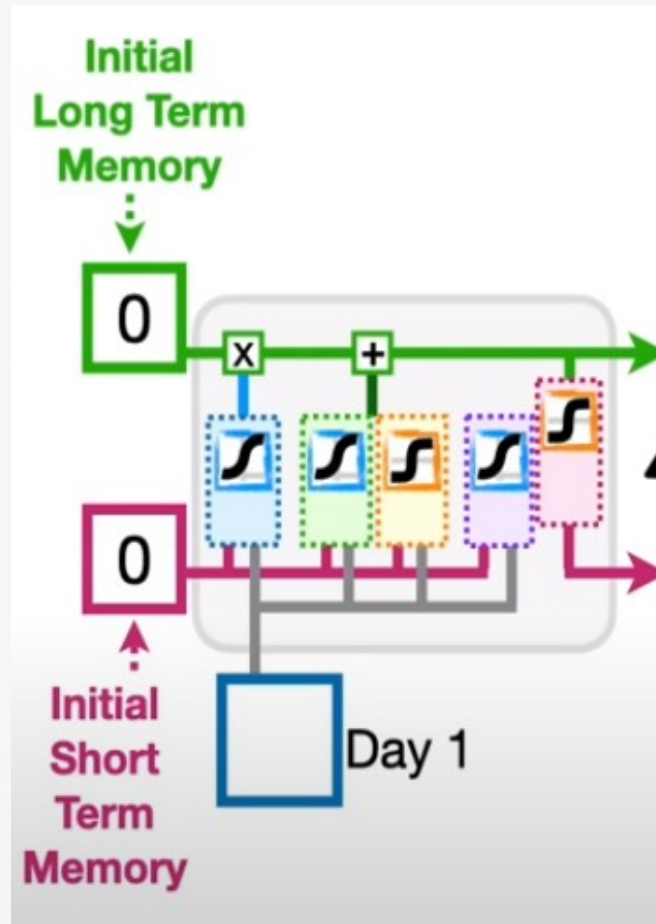


# Desdobrando a rede LSTM

Vamos executar sequencialmente os valores dos dias 1 a 4.  
Começando com valor de entrada=1

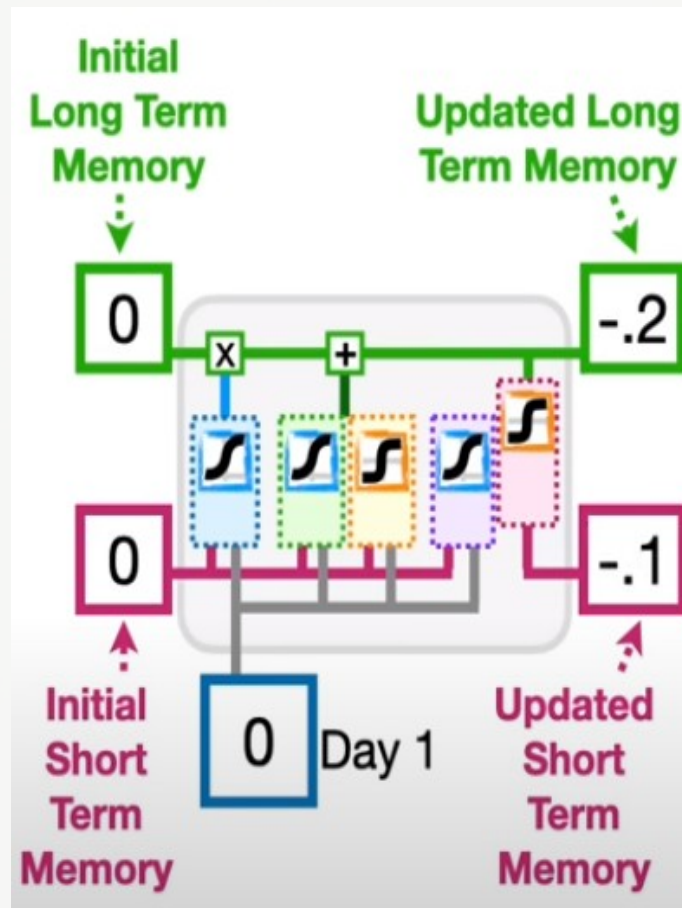


Executando sequencialmente dos dias 1 a 4  
Começando com valor de entrada=1

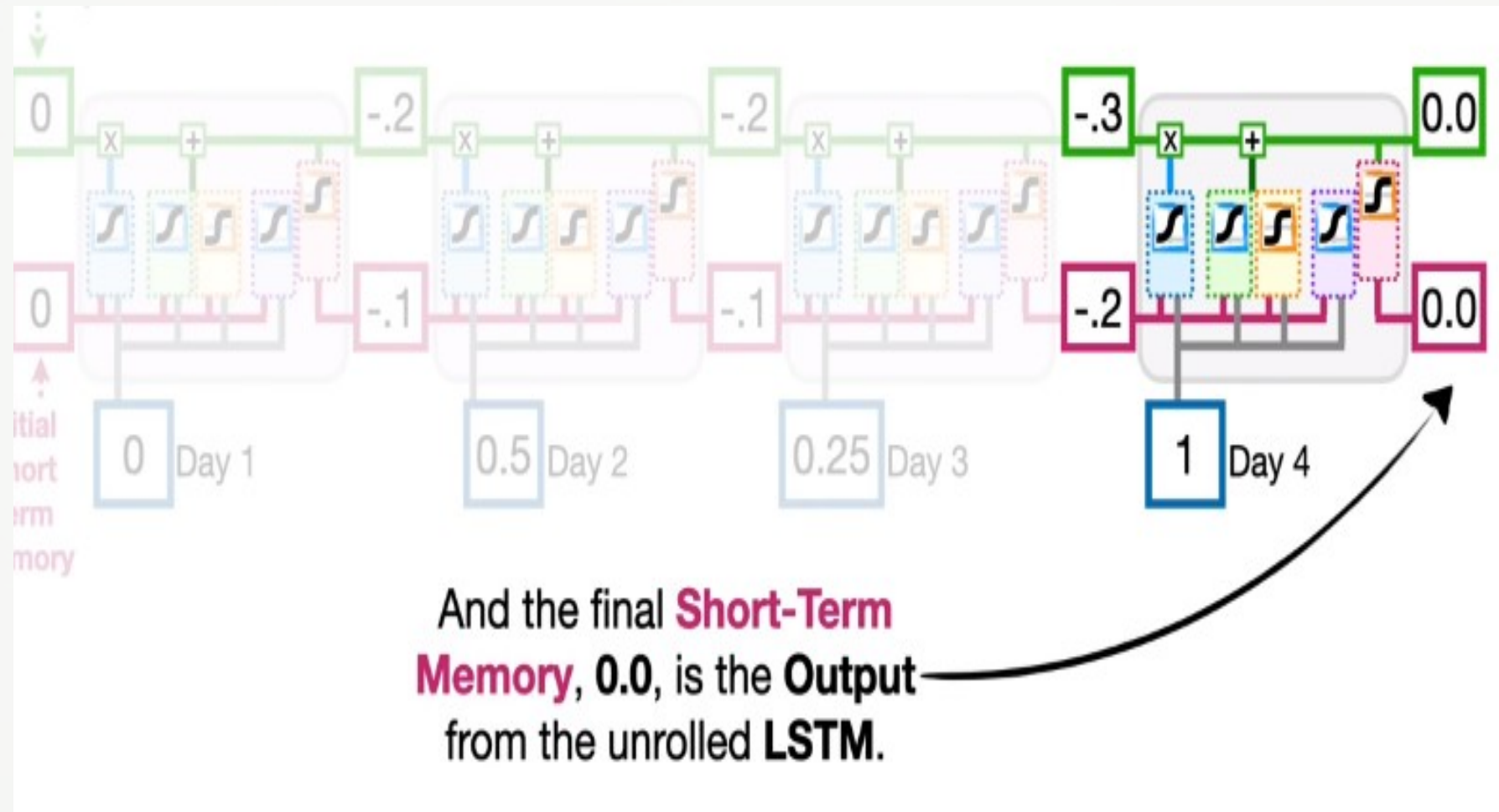


Depois das contas, a nova memória de longo prazo atualizada é de -0,20.  
e a memória de curto prazo atualizada é -0,13

Então conectamos -0,2 para memória de longo prazo atualizada e -0,1 (arredondada) para memória de curto prazo atualizada. Depois desenrolamos o LSTM usando essas memórias atualizadas



Saída do dia 4 é o dado previsto para o dia 5 da Companhia A.



# Complexidade Treinamento LSTM

A complexidade computacional do LSTM por etapa e peso é  $O(1)$ , ou seja, complexidade constante, ou tempo de execução constante, independentemente do tamanho da entrada



# Aplicações das redes LSTM



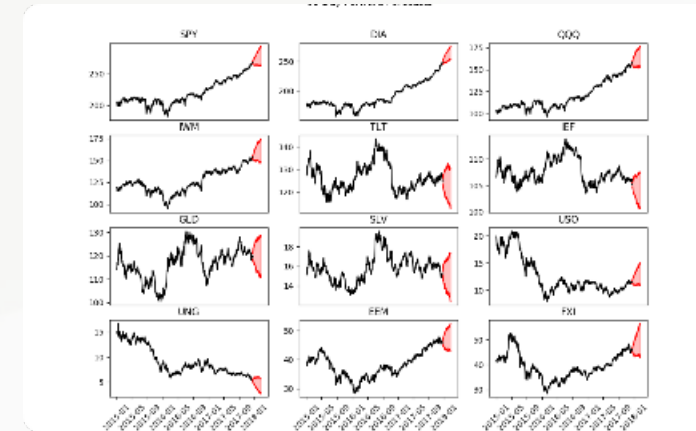
## Tradução automática

Redes LSTM têm sido amplamente utilizadas em sistemas de tradução automática, produzindo resultados mais precisos.



## Reconhecimento de voz

Graças à capacidade de processar sequências temporais, as redes LSTM têm sido aplicadas com sucesso no reconhecimento de fala.



## Previsão de mercado de ações

As redes LSTM podem capturar padrões complexos em séries temporais e, por isso, são frequentemente usadas para prever o comportamento do mercado de ações.

# Vantagens e desvantagens das redes LSTM

## Vantagens

Capacidade de lidar com longas sequências temporais, preservando informações relevantes ao longo do tempo.

Descarta informações irrelevantes

Podem esquecer completamente seu estado anterior, evitando vieses na previsão

## Desvantagens

Maior complexidade computacional e maior consumo de memória em comparação com outros modelos de redes neurais recorrentes.

Blocos de memória não mudam dinamicamente, assim a memória acaba sendo limitada.

É improvável superar essa limitação, a modularização promove o aprendizado eficaz, no entanto esse processo não está claro de modo geral

# Outras Topologias

***Bidirecional:*** significa que a entrada é apresentada para frente e para trás duas redes separadas, ambas conectadas a mesma camada de saída (usado para classificar fonemas)

**Grid LSTM (N-LSTM. grade N-Dimensional),** com células ao longo e entre algumas (ou todas) das dimensões, permitindo a comunicação entre as camadas consecutivas.

**LSTM Empilhado:** empilha camadas LSTM, para aumentar capacidade.

**LSTM Multidimensional:** as entradas são estruturadas em uma grade N-dimensional em vez de serem sequência de valores.

**Blocos LSTM de grade:** calcula N transformações LSTM, uma para cada dimensão

**GATED RECURRENT UNIT - UNIDADE RECORRENTE FECHADA (GRU);** não tem célula de memória, tem unidades de controle (porta de redefinição e porta de atualização)

# Aplicações

**Reconhecimento de fala:** sistemas LSTM padrão (2003), depois baseados em modelo de Markov Oculto (HMM), BLSTM e BLSTM-CTC e depois BLSTM-CTC-empilhada (2013), depois ampliada, obtendo resultados excepcionais. Em 2015 introduziu arquitetura listener (ouvinte) e attend and spell (atender e soletrar) - aprendizado baseado em atenção

**Escrita à mão.** 2007 - BLSTM-CTC (reconhecimento on-line), superando posteriormente o HMM. Combinou o BLSTM-CTC com LSTM multidimensional para reconhecimento off-line, também superando HMM. Em 2013 passou a aplicar o método de regularização dropout.

**Reconhecimento de emoções, e modelagem de conversação.**

**Tradução automática:** 2014 - rede neural codificador-decodificador (aprendizado sequencia-a-sequencia), sendo aprimorada para traduzir palavras raras (fora do vocabulário), e posteriormente tradução de frases longas (atenção do decodificador)

**Processamento de Imagens:** 2012- detecção de palavras-chave e detecção de modo. Banco de dados IMAGENET. 2015 (sequencia-a-sequencia). Depois combinaram LSTMs com extrator de recursos visuais hierárquicos profundos para interpretar e classificar imagens (reconhecimento de atividades e descrição de imagens/vídeos)

# Outras Tarefas

- ☐ Estrutura secundária de proteínas
- ☐ Geração de músicas
- ☐ Segurança de redes
- ☐ Avaliação de programas de computador cursos usando o sequênci-a-sequência - 2014