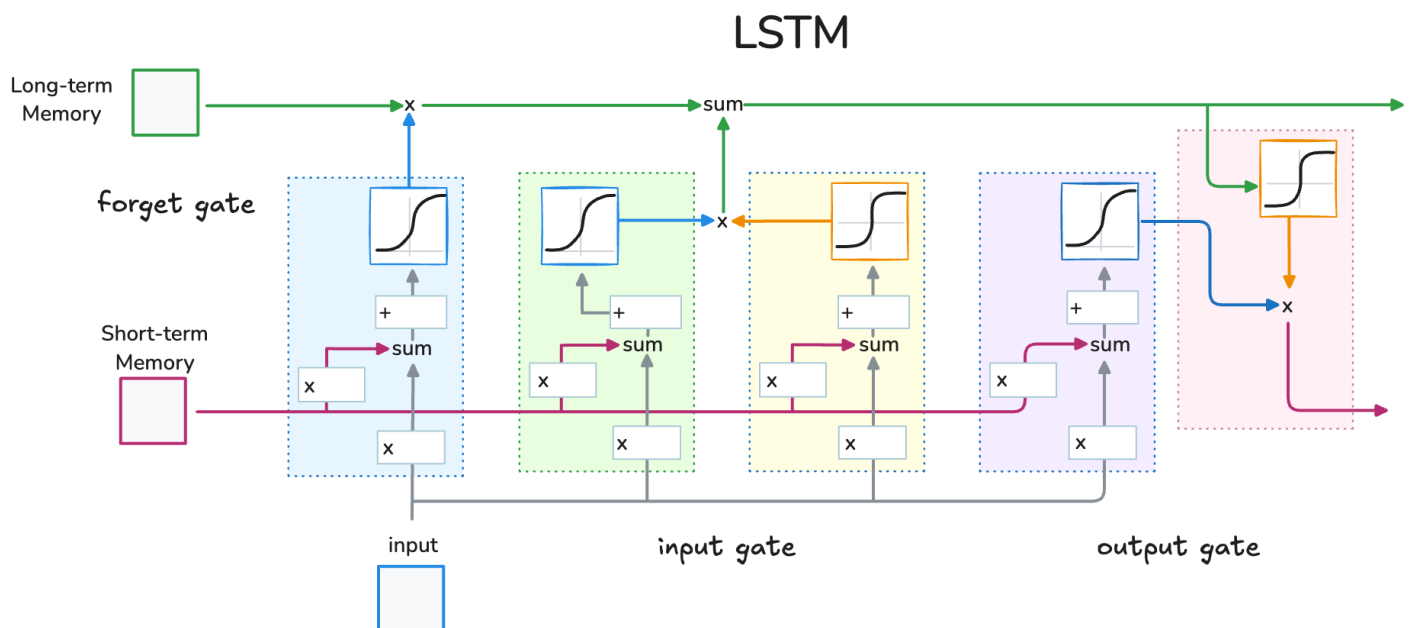


## Estrutura e Funcionamento da LSTM

As **Long Short-Term Memory Networks (LSTMs)** representam uma evolução das redes neurais recorrentes (RNNs), concebidas para resolver o problema do esquecimento de informações ao longo do tempo — o chamado *vanishing gradient*. Enquanto uma RNN simples processa as entradas em sequência e tende a perder gradualmente as dependências mais distantes, a LSTM foi idealizada para preservar e controlar o fluxo de memória. Sua estrutura permite armazenar informações relevantes por longos intervalos, descartando o que não é mais útil de forma autônoma.

A LSTM é composta por **células de memória** que se encadeiam ao longo da sequência temporal. Cada célula mantém dois estados: o **estado de célula** ( $c_t$ ), que funciona como memória de longo prazo, e o **estado oculto** ( $h_t$ ), que representa o contexto imediato ou memória de curto prazo. O diferencial da LSTM está nas **portas de controle** que regulam a passagem dessas informações. Essas portas — de esquecimento, de entrada e de saída — atuam como filtros que decidem quanto do passado deve ser mantido, o que deve ser atualizado e o que será exposto como saída.

Figure 1 Representação esquemática de uma célula LSTM. As portas de esquecimento, entrada e saída regulam o fluxo de informações entre a memória de longo e curto prazo.



Fonte: elaboração própria com base em Hochreiter e Schmidhuber (1997).

A **porta de esquecimento** define quais informações antigas deixarão de ser relevantes. Ela avalia o estado anterior e a nova entrada, produzindo valores entre 0 e 1. Quando o resultado se aproxima de zero, a informação é descartada; quando se aproxima de um, ela é mantida. A **porta de entrada**, por sua vez, decide quais novas informações serão incorporadas à memória de longo prazo. Ela gera uma informação candidata e, por meio de um fator de ponderação, define o quanto dessa nova informação será realmente armazenado. Já a **porta de saída** regula a quantidade de informação que será utilizada naquele instante, atualizando o estado oculto e definindo o que será transmitido para a próxima célula.

Essas decisões são sintetizadas na atualização do estado de célula. O novo estado combina parte da memória antiga com a nova informação selecionada, formando um equilíbrio dinâmico entre persistência e adaptação. Esse mecanismo cria um canal de gradiente mais estável durante o treinamento, evitando que as informações se percam ao longo de muitas iterações. Em termos práticos, isso significa que a LSTM pode aprender dependências temporais longas — por exemplo, reconhecer padrões de comportamento de clientes ao longo de vários meses ou capturar tendências de séries históricas.

A operação das portas é mediada por funções de ativação específicas. A **sigmoide** é usada para limitar as decisões das portas entre 0 e 1, enquanto a **tangente hiperbólica** mantém os valores de atualização do estado entre -1 e 1. Essa combinação de funções garante estabilidade numérica e suavidade no fluxo de informações. Além disso, o gradiente flui de forma quase direta pelo estado de célula, o que preserva a capacidade de aprendizado mesmo em sequências longas.

Do ponto de vista conceitual, a LSTM pode ser comparada a uma “planilha de memória inteligente”. Cada linha dessa planilha contém um registro de conhecimento acumulado, e as portas funcionam como mecanismos que decidem se esse registro deve ser mantido, alterado ou apagado. Esse processo é contínuo e adaptativo, permitindo que a rede aprenda o que deve ser lembrado com base na tarefa e nos dados apresentados.

Durante o treinamento, os pesos associados às portas são ajustados de modo que o modelo aprenda padrões temporais recorrentes. Por exemplo, se uma mudança de comportamento no início da sequência tende a influenciar fortemente o resultado final, a porta de esquecimento aprende a preservar essa informação. Caso contrário, ela a elimina. Assim, a LSTM não apenas memoriza o passado, mas aprende **quando e por que lembrar**, desenvolvendo uma representação temporal eficiente do fenômeno analisado.

A LSTM é, portanto, uma arquitetura que combina memória seletiva e estabilidade de aprendizado. Sua capacidade de controlar o fluxo de informação em múltiplas escalas temporais a torna particularmente útil para problemas em que o tempo e a sequência importam. Em contextos práticos, como previsão de demanda, análise de engajamento ou modelagem de séries financeiras, ela se destaca por capturar dependências que outros modelos ignorariam. É essa habilidade de “lembrar com propósito” que consolidou a LSTM como um marco na evolução das redes neurais.

## Referência

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural Computation**, [S. l.], v. 9, n. 8, p. 1735–1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).