

# Anatomía Matemática de la Celda LSTM: Las 4 Compuertas

**Author:** D.Sc. Aboud Barsekh Onji

**Institution:** Universidad Anáhuac México - Facultad de Ingeniería

**Contact:** aboud.barsekh@anahuac.mx

**ORCID:** 0009-0004-5440-8092

---

La red **LSTM (Long Short-Term Memory)** soluciona el problema del *gradiente desvanecido* presente en las Redes Recurrentes (RNN) simples. A diferencia de una neurona estándar que simplemente aplica una función  $\tanh(Wx + b)$ , una celda LSTM posee una estructura compleja de memoria interna regulada por **4 compuertas (gates)** operando en paralelo.

Esta estructura explica por qué en MATLAB vemos una cantidad de parámetros aprendibles muy superior al número de neuronas declaradas.

**Regla de Parámetros:** Si definimos  $N$  unidades ocultas, los pesos internos se calculan sobre  $4 \times N$ .

(Ej. 128 unidades  $\rightarrow$  512 juegos de pesos en las matrices).

A continuación, se detallan las 4 sub-redes internas:

## 1. Forget Gate (Puerta de Olvido) - $f_t$

Determina qué información del pasado es irrelevante y debe ser eliminada de la memoria a largo plazo (Cell State,  $C_{t-1}$ ).

- **Función de Activación:** Sigmoide ( $\sigma$ ). Salida en el rango  $[0, 1]$ .
- **Operación:** Observa la entrada actual  $x_t$  y el estado oculto anterior  $h_{t-1}$ .
- **Significado:**
  - 0: “Olvidar esto completamente”.
  - 1: “Mantener esto intacto”.
- *Tiene su propio conjunto de Pesos ( $W_f$ ) y Bias ( $b_f$ ).*

## 2. Input Gate (Puerta de Entrada) - $i_t$

Decide qué valores nuevos vamos a permitir entrar en la memoria. Funciona como un filtro de atención para los nuevos datos.

- **Función de Activación:** Sigmoide ( $\sigma$ ). Salida en el rango  $[0, 1]$ .
- **Operación:** Determina la **importancia** del nuevo dato entrante.
- *Tiene su propio conjunto de Pesos ( $W_i$ ) y Bias ( $b_i$ ).*

## 3. Cell/Candidate Gate (Candidato a Memoria) - $\tilde{C}_t$

Genera el vector de **nueva información** potencial a ser almacenada.

- **Función de Activación:** Tangente Hiperbólica ( $\tanh$ ). Salida en el rango  $[-1, 1]$ .
- **Operación:** Crea una representación vectorial de los nuevos datos  $x_t$  y el contexto inmediato  $h_{t-1}$ .
- **Interacción:** Este valor candidato se multiplica por la salida de la *Input Gate* ( $i_t \times \tilde{C}_t$ ) para agregarse solo si la puerta de entrada lo permite.
- *Tiene su propio conjunto de Pesos ( $W_c$ ) y Bías ( $b_c$ ).*

#### 4. Output Gate (Puerta de Salida) - $o_t$

Controla el flujo de información hacia el exterior. Decide qué parte de la memoria interna ( $C_t$ ) se expondrá como “Estado Oculto” ( $h_t$ ) para la siguiente capa o el siguiente paso de tiempo.

- **Función de Activación:** Sigmoide ( $\sigma$ ).
- **Lógica:** Separa la memoria interna (que puede contener mucha historia acumulada) de lo que es relevante *en este instante específico* para realizar la predicción.
- *Tiene su propio conjunto de Pesos ( $W_o$ ) y Bias ( $b_o$ ).*

---

### Conclusión sobre Complejidad Computacional

Debido a estas 4 operaciones matriciales por cada paso de tiempo ( $t$ ), las LSTM son computacionalmente más costosas que las redes simples, pero ofrecen una capacidad superior para aprender dependencias a largo plazo (e.g., el final de una oración dependiendo de una palabra al inicio).

Ecuación simplificada del flujo:

$$\text{LSTM}(x_t, h_{t-1}, C_{t-1}) \rightarrow [h_t, C_t]$$