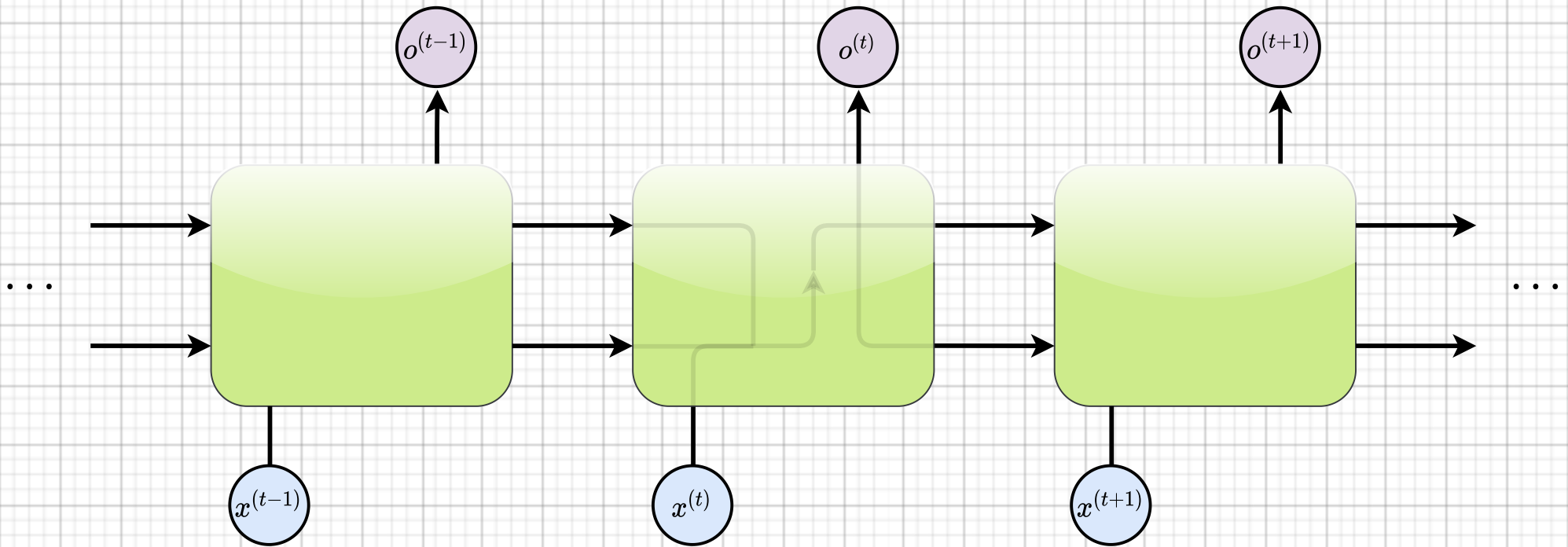


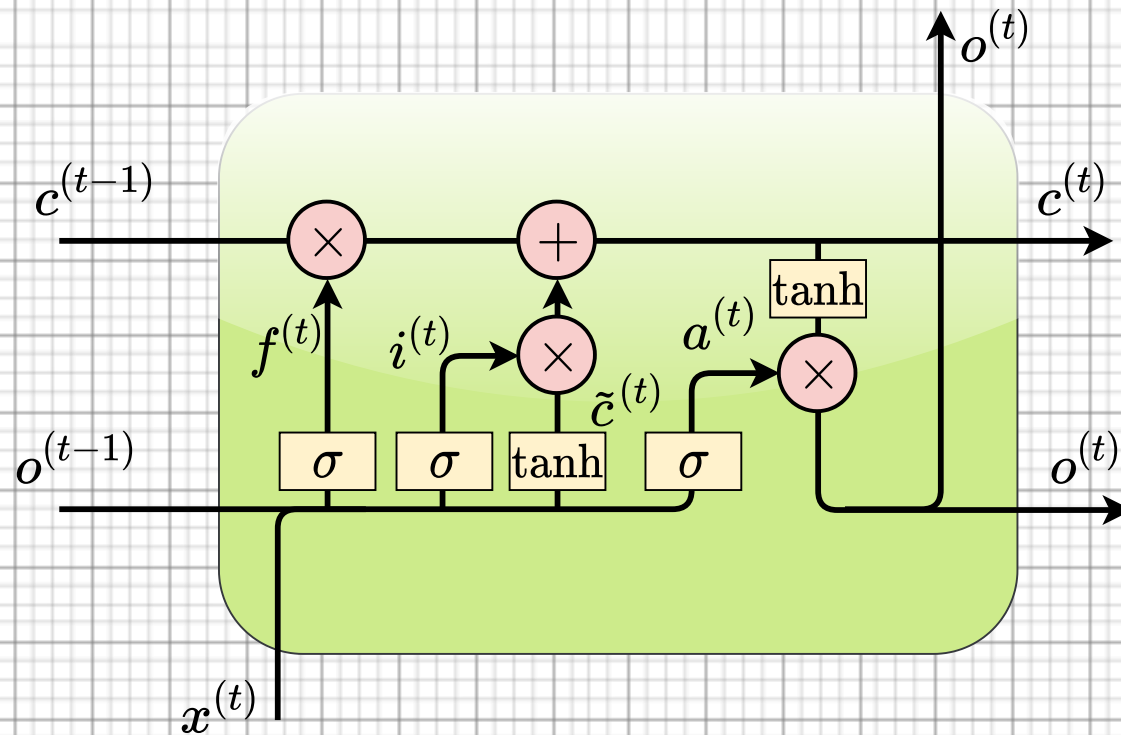
LSTM

Long Short Term Memory networks son un tipo especial de RNNs capaces de aprender a preservar información a largo plazo, precisamente están diseñadas para lidiar con ese problema de retención por un largo período de tiempo.



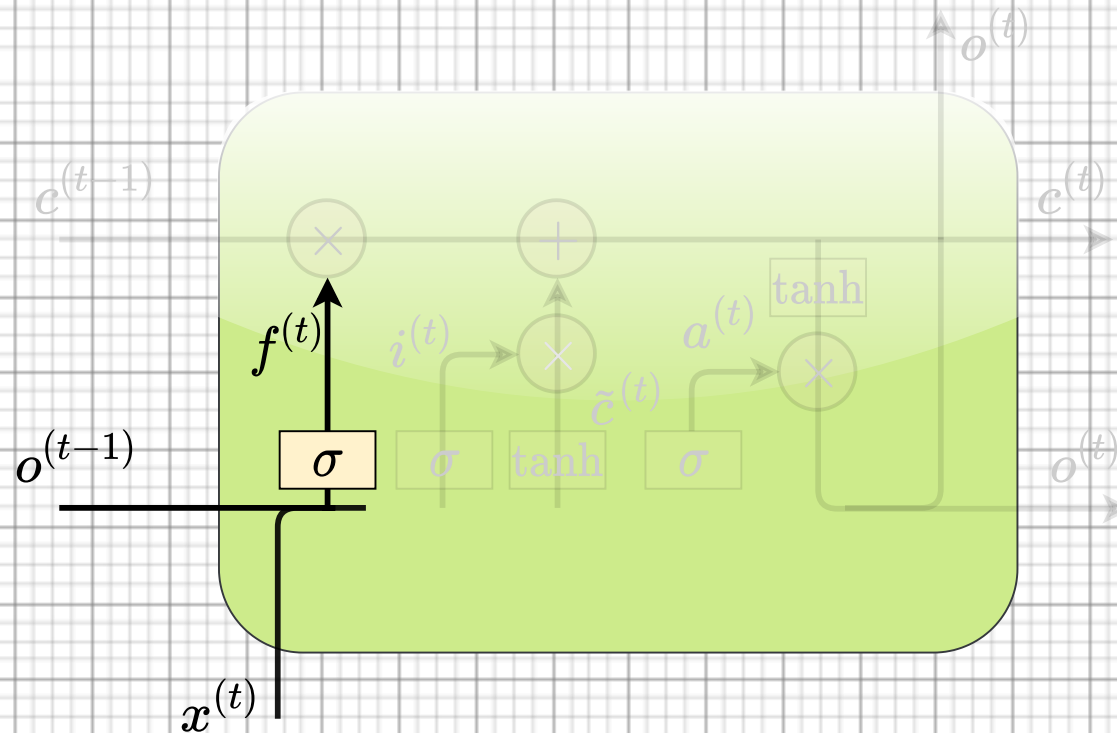
Célula LSTM

Por medio de compuertas, *gates* (σ) esta célula es capaz de regular el flujo y retención de la información.



Paso 1

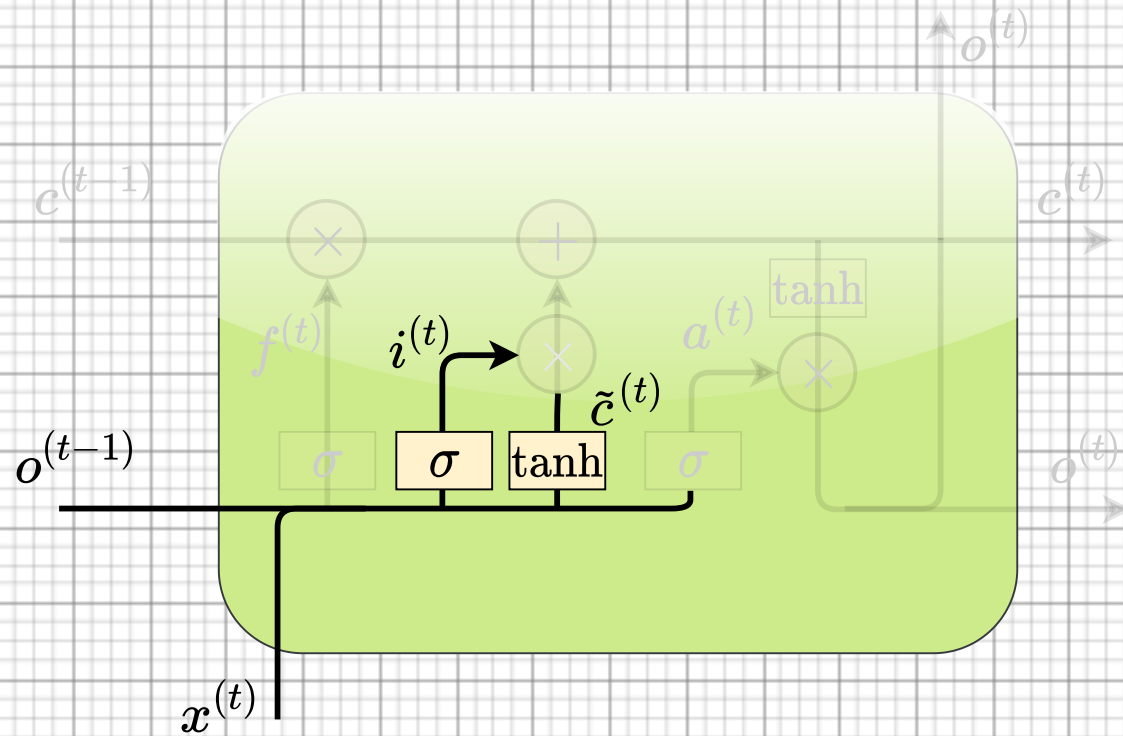
Aquí decide que tanta información se conserva del estado de la unidad anterior ($t - 1$), así como la información que viene del *input* al tiempo t .



$$f^{(t)} = \sigma \left(W_f \cdot [o^{(t-1)}, x^{(t)}] + b_f \right)$$

Paso 2

Aquí la primera compuerta, *input gate layer*, decide que información se va actualizar. La segunda compuerta selecciona valores que son candidatos a estar en el estado de la célula.

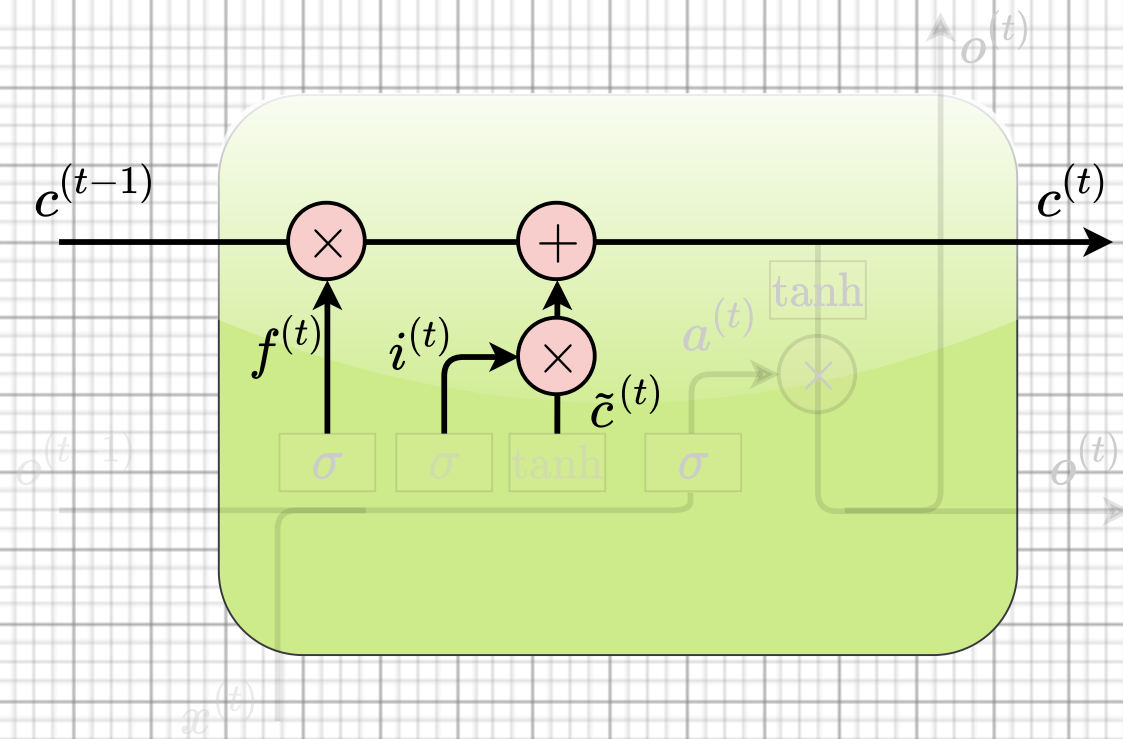


$$i^{(t)} = \sigma \left(W_i \cdot [o^{(t-1)}, x_t] + b_i \right)$$

$$\tilde{c}^{(t)} = \tanh \left(W_c \cdot [o^{(t-1)}, x_t] + b_c \right)$$

Paso 3

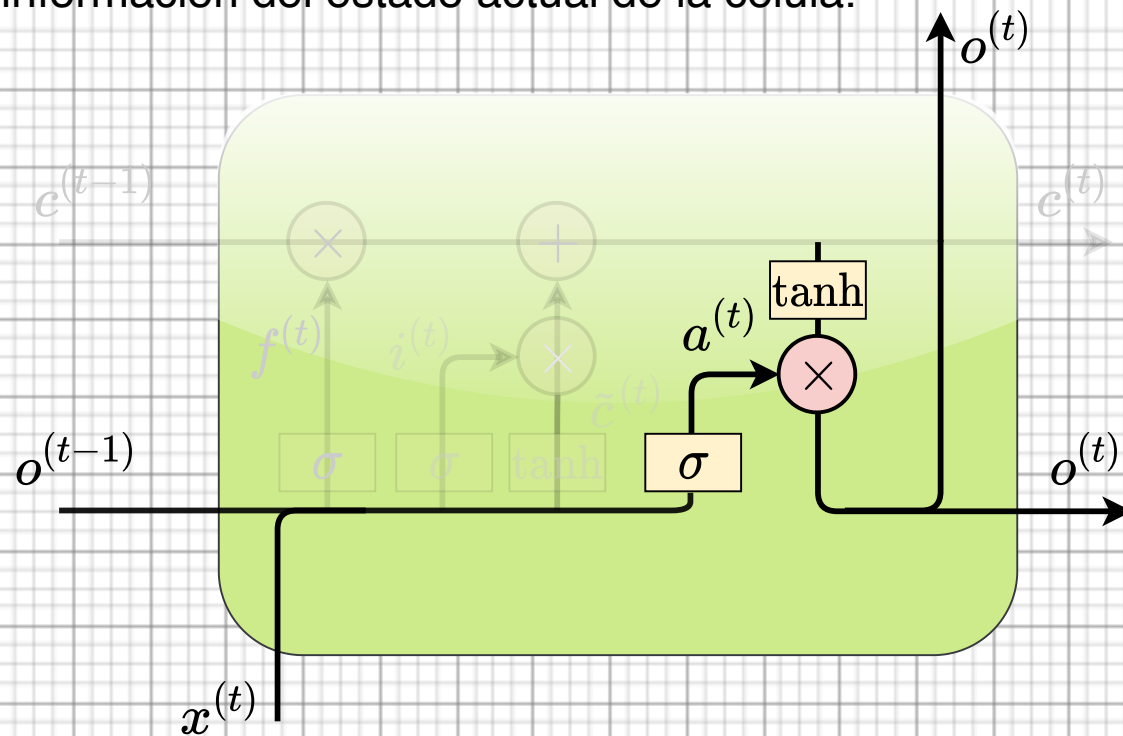
Aquí se actualiza el estado de la célula para el tiempo t . Se realiza una combinación entre el estado anterior ($t - 1$) y el estado auxiliar de la célula al tiempo t .



$$c^{(t)} = f^{(t)} \cdot c^{(t-1)} + i^{(t)} \cdot \tilde{c}^{(t)}$$

Paso 4

Finalmente se decide el *output* de la célula, que será una versión filtrada del estado de la célula al tiempo t . La primera compuerta decide que valores de la información actual y el estado anterior pasan a la siguiente unidad. La siguiente parte funciona como discriminador de la información del estado actual de la célula.



$$a^{(t)} = \sigma \left(W_a [o^{(t-1)}, x^{(t)} + b_a] \right)$$

$$o^{(t)} = a^{(t)} \tanh \left(c^{(t)} \right)$$

En resumen ...

LSTM representa una mejora con respecto a las arquitecturas *vanilla RNN*, pues lidia con el problema de retención sin que explote o desvanezca el gradiente.

$$f^{(t)} = \sigma \left(W_f \cdot [o^{(t-1)}, x^{(t)}] + b_f \right)$$

$$i^{(t)} = \sigma \left(W_i \cdot [o^{(t-1)}, x_t] + b_i \right)$$

$$c^{(t)} = f^{(t)} \cdot c^{(t-1)} + i^{(t)} \cdot \tilde{c}^{(t)}$$

$$\tilde{c}^{(t)} = \tanh \left(W_c \cdot [o^{(t-1)}, x_t] + b_c \right)$$

$$a^{(t)} = \sigma \left(W_a [o^{(t-1)}, x^{(t)}] + b_a \right)$$

$$o^{(t)} = a^{(t)} \tanh \left(c^{(t)} \right)$$

