

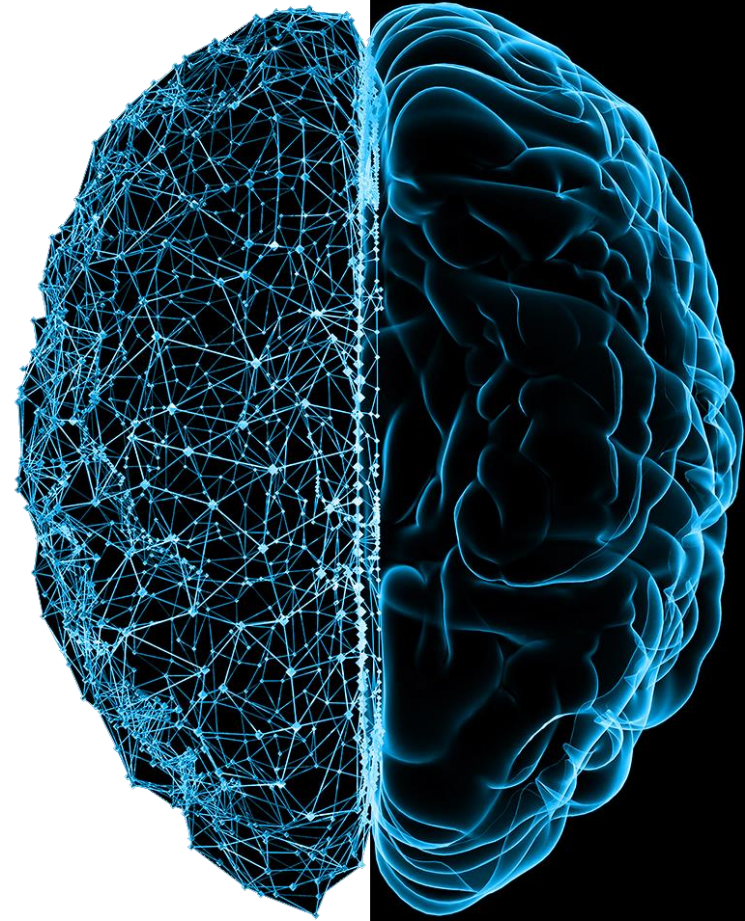
Procesamiento de Lenguaje Natural

Clase 13 – Modelos
Secuenciales

Ph.D. Rubén Manrique

rf.manrique@uniandes.edu.co

Maestría en Ingeniería de Sistemas
y Computación



$c_{t=0}$

4 jh
4c
2l

$h_{t=0}$

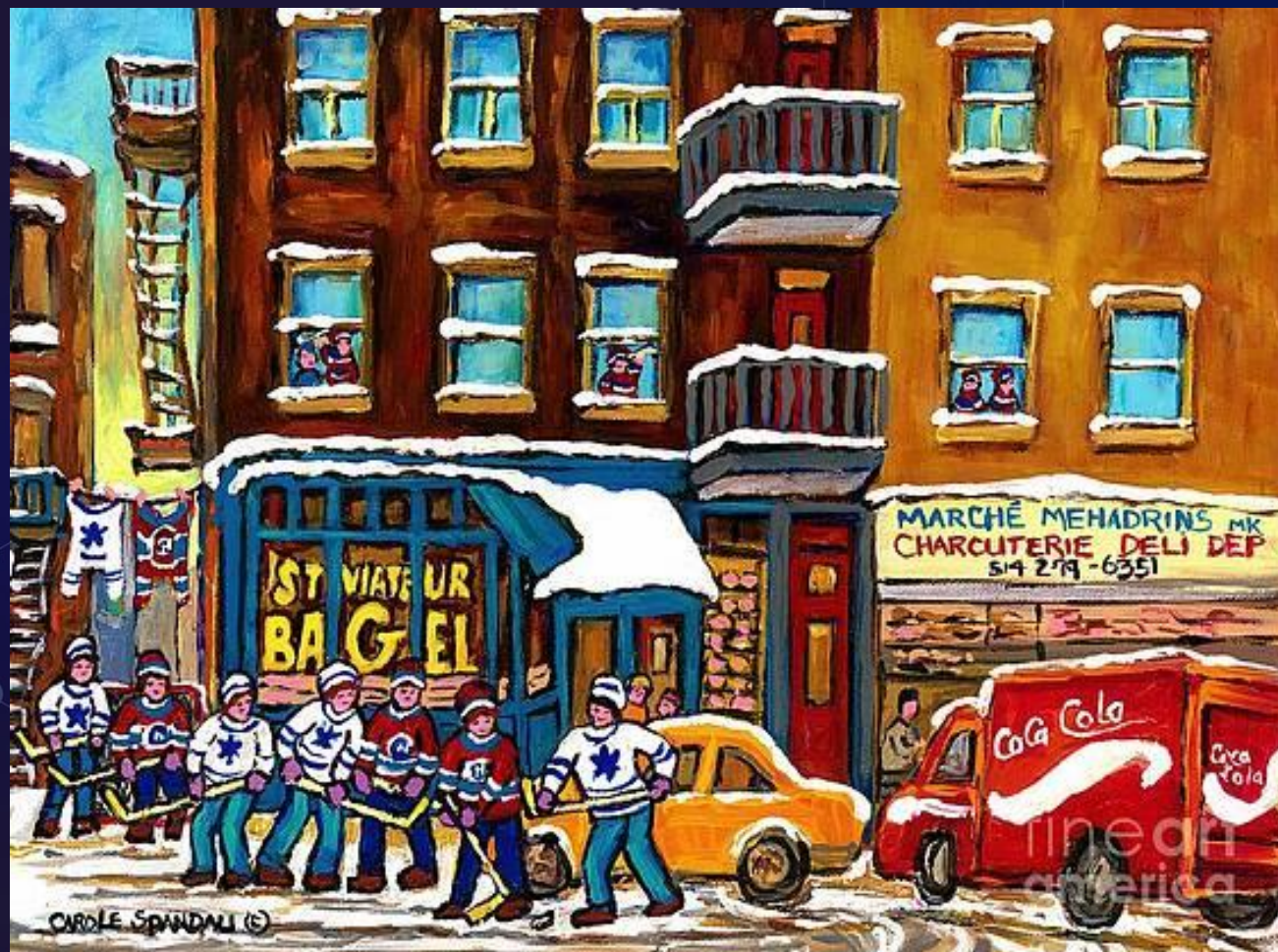
4jh

$c_{t=1}$

4jh,
7jh
2l
1CC

$h_{t=1}$

11jh



$c_{t=0}$

4jh

4c

2l

$h_{t=0}$

4jh

$c_{t=1}$

4jh,

7jh

2l

1CC

$h_{t=1}$

11jh

$c_{t=2}$

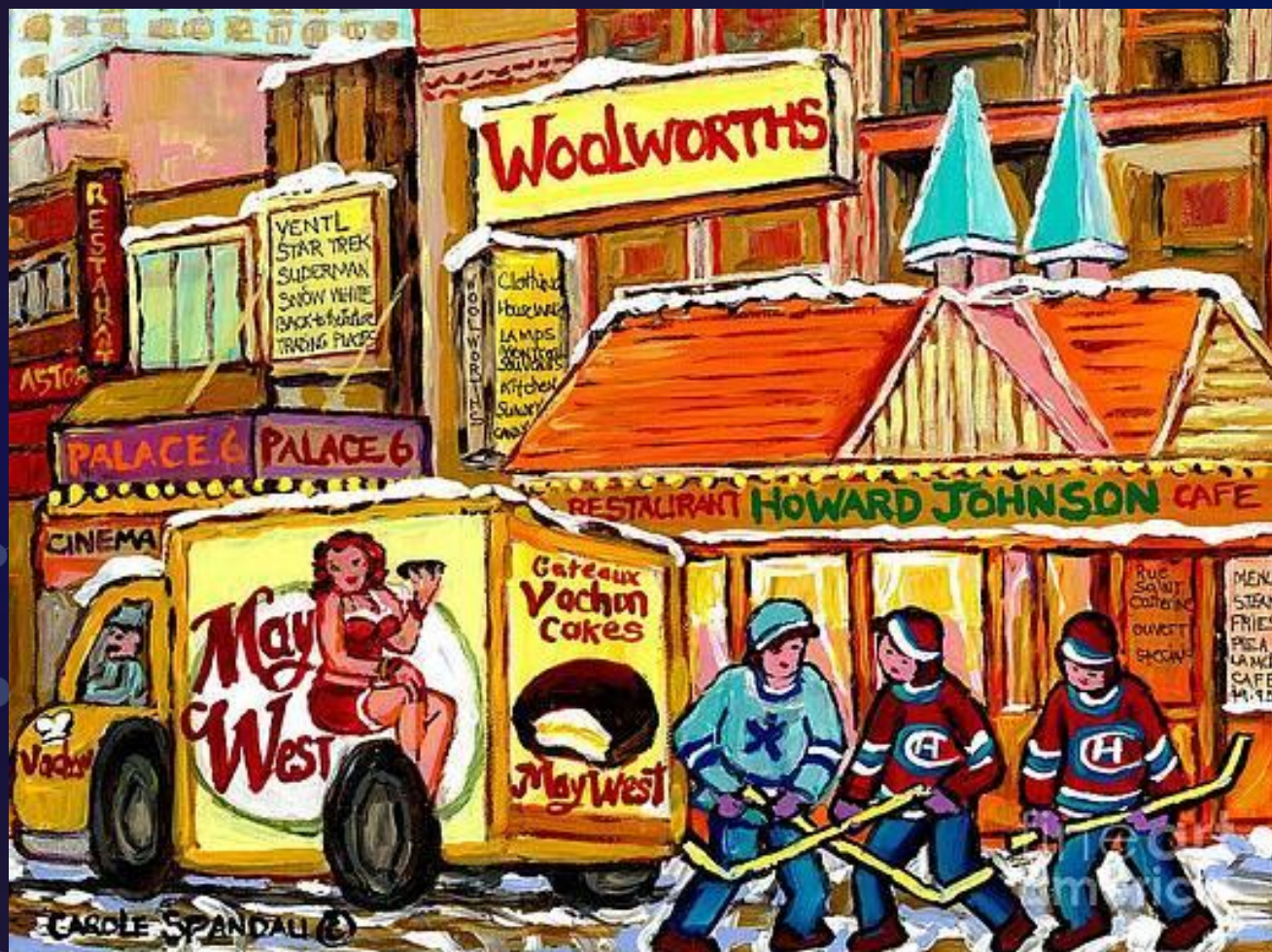
4jh, 7jh,

3jh

2l

$h_{t=2}$

14jh



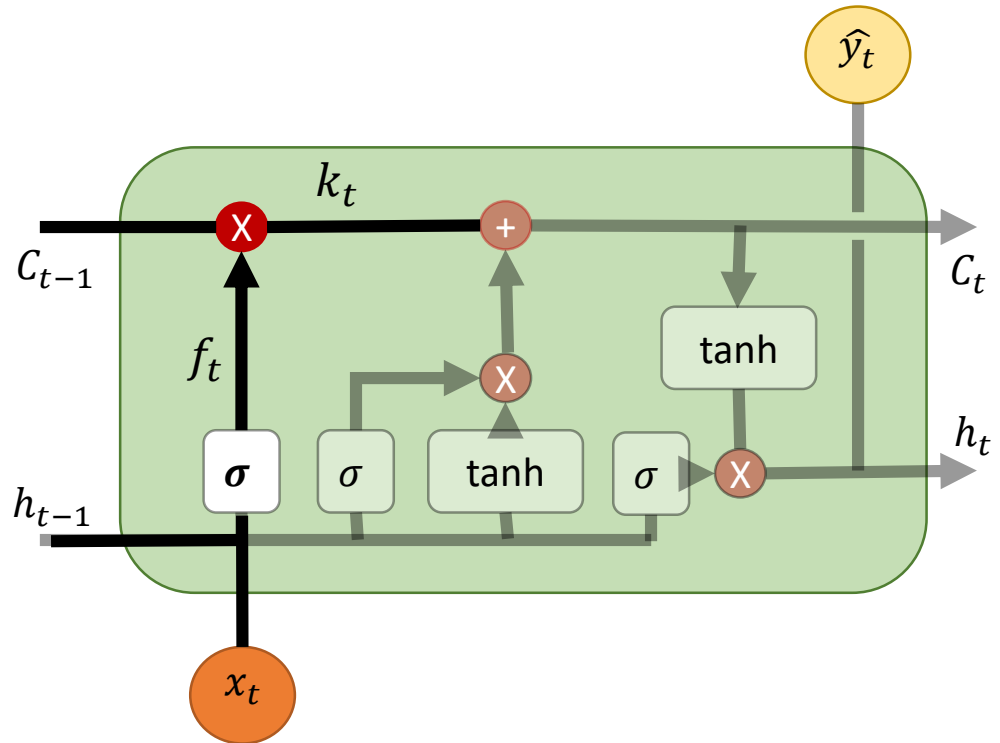
Long Short-Term Memory

- Divide la gestión del contexto en dos problemas:
 - Remover información innecesaria.
 - Adicionar información que es probablemente es necesaria para las futuras decisiones.
- Añade una capa de contexto (adicional a la capa oculta recurrente).
- Usa unidades especializadas de control de flujo denominadas “gates”.
 - Consisten en una capa feedforward seguida de una función de activación sigmoide y una compuerta de multiplicación.
 - La elección del sigmoide como función de activación surge de su tendencia a empujar sus salidas a 0 o 1. Combinar esto con una multiplicación tiene un efecto similar al de una máscara binaria.

Forget Gate

- El propósito de esta puerta es eliminar información del contexto que ya no es necesaria.

$$f_t = \sigma(U_f h_{t-1} + W_f x_t)$$
$$k_t = c_{t-1} \odot f_t$$



Input/add gate

Información actual, nótese la función de activación – información candidata a ser incorporada.

$$g_t = \tanh(U_g h_{t-1} + W_g x_t)$$

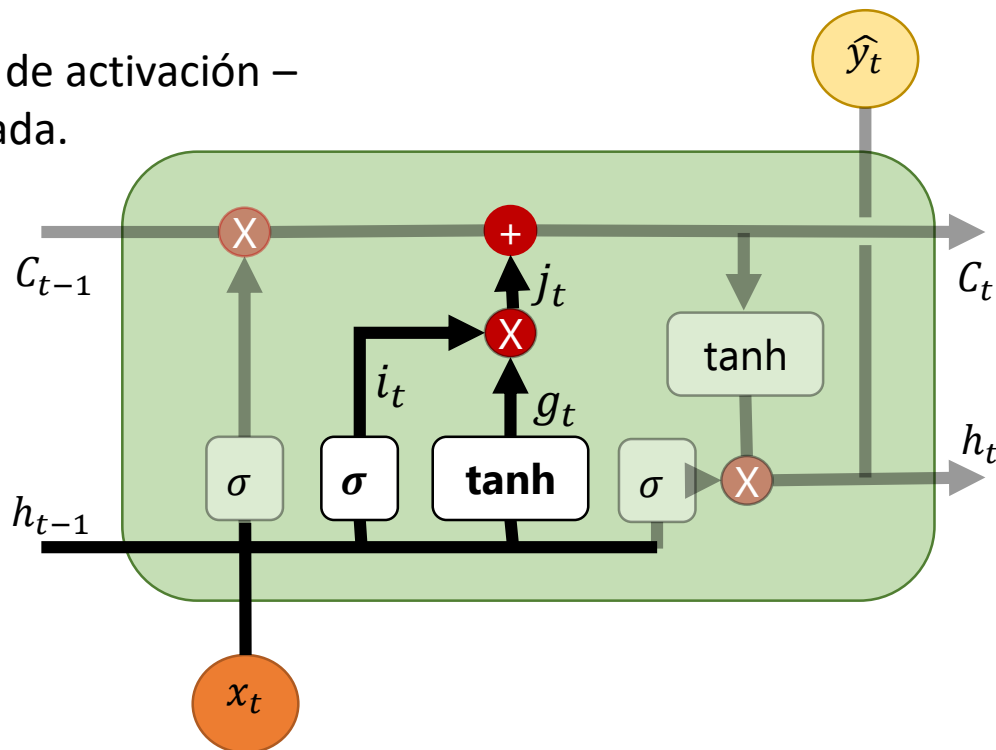
Seleccionar que información vamos a añadir

$$i_t = \sigma(U_i h_{t-1} + W_i x_t)$$

$$j_t = g_t \odot i_t$$

Lo añadimos al contexto:

$$c_t = j_t + c_{t-1}$$

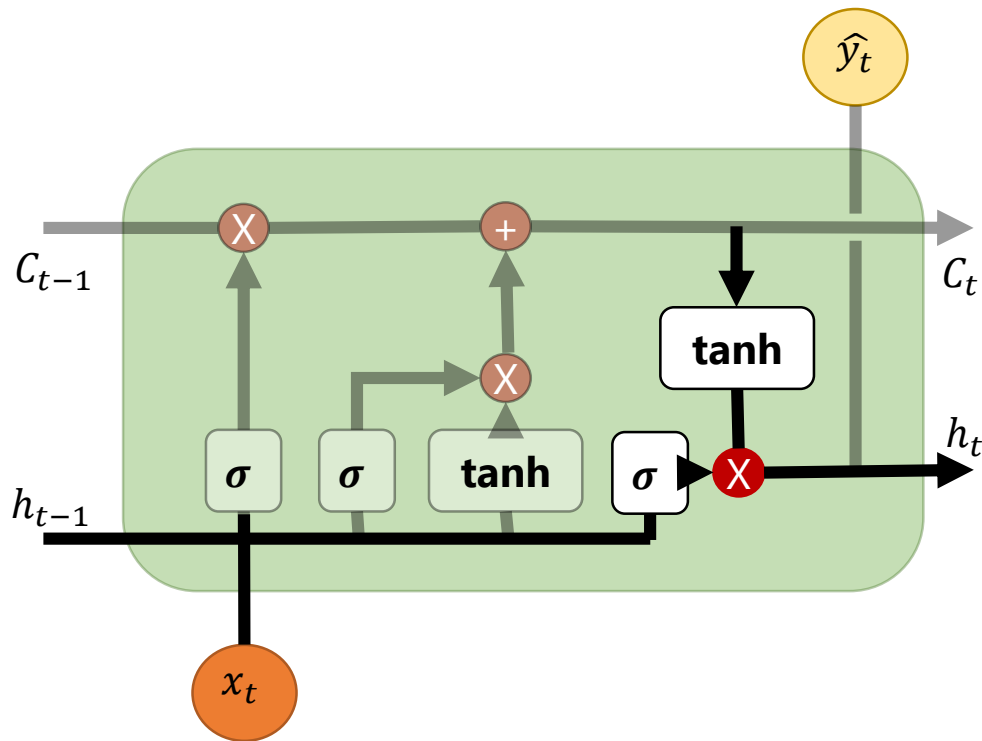


Output Gate

Calculamos el estado oculto para t

$$o_t = \sigma(U_o h_{t-1} + W_o x_t)$$

$$h_t = o_t \odot \tanh(c_t)$$



Resumen

$$f_t = \sigma(U_f h_{t-1} + W_f x_t)$$

$$k_t = c_{t-1} \odot f_t$$

$$g_t = \tanh(U_g h_{t-1} + W_g x_t)$$

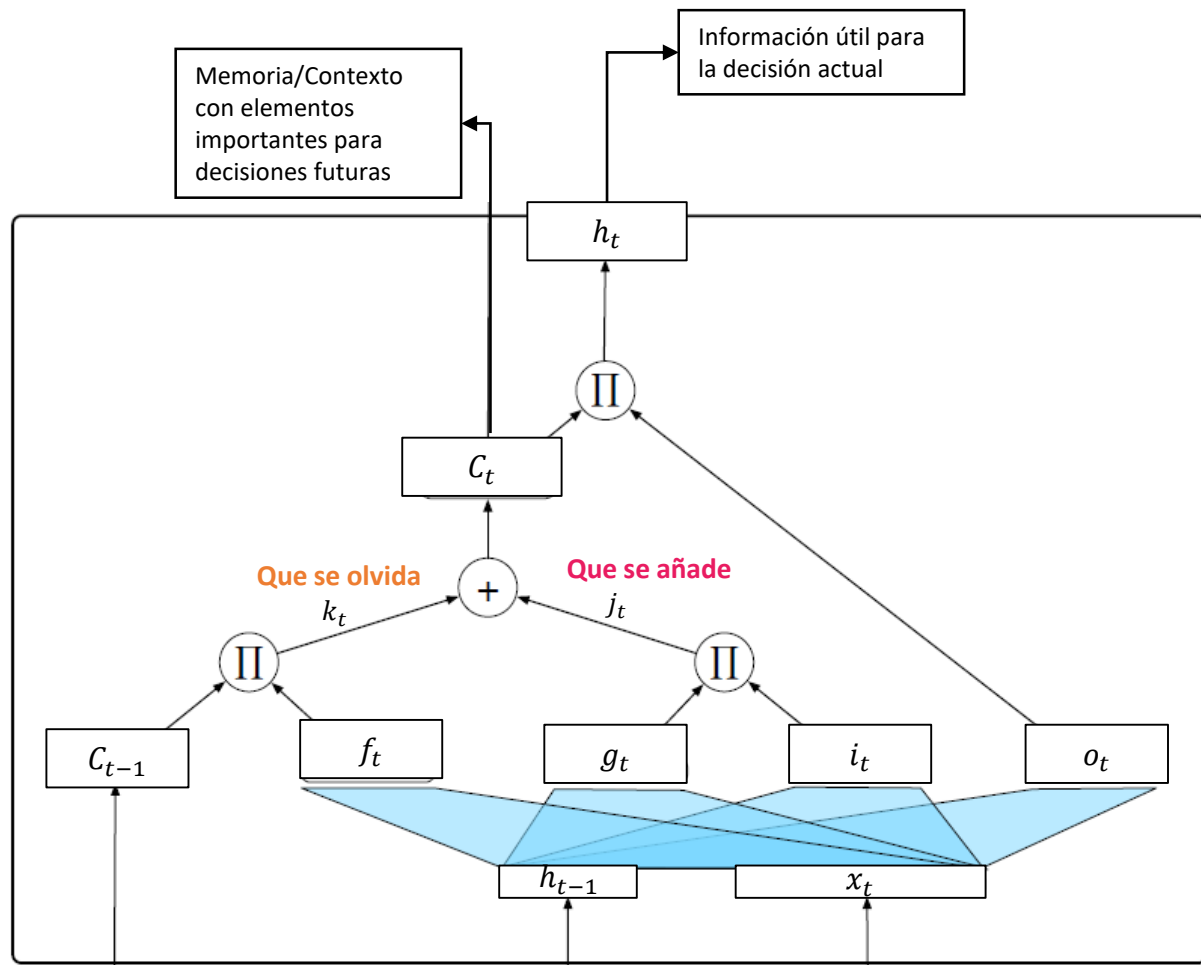
$$i_t = \sigma(U_i h_{t-1} + W_i x_t)$$

$$j_t = g_t \odot i_t$$

$$c_t = j_t + k_t$$

$$o_t = \sigma(U_o h_{t-1} + W_o x_t)$$

$$h_t = o_t \odot \tanh(c_t)$$



8 matrices de parámetros en LSTM vs 2 en RNN

Gated Recurrent Units (I)

- Reducción del número de matrices de parámetros.
 - Se reducen el número de compuertas a 2.
 - Una compuerta de reset “r”. El propósito de la puerta de reinicio es decidir qué aspectos del estado oculto anterior es relevante para el contexto actual y lo que se puede ignorar.
 - Una compuerta de actualización “z”.

$$r_t = \sigma(U_r h_{t-1} + W_r x_t)$$

$$z_t = \sigma(U_z h_{t-1} + W_z x_t)$$

Gated Recurrent Units (II)

Estado intermedio del estado oculto en tiempo t

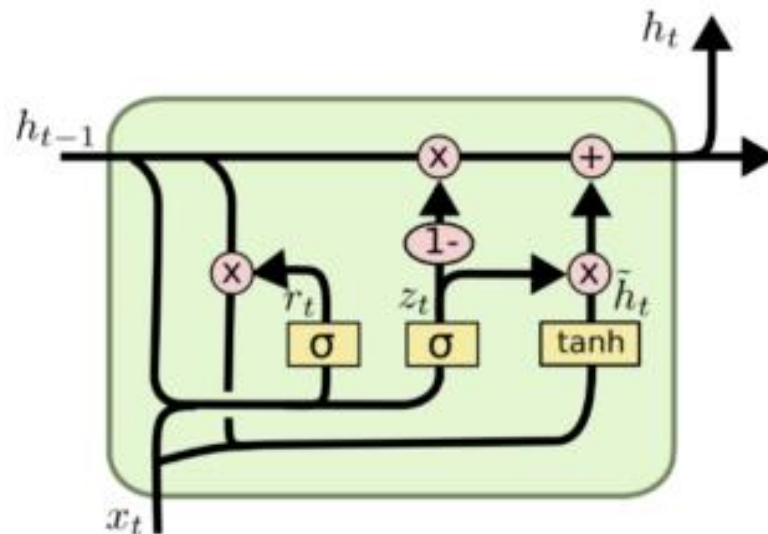
$$\tilde{h}_t = \tanh(U(r_t \odot h_{t-1}) + Wx_t)$$

Con ayuda de la compuerta de Actualización se determina que aspectos del estado intermedio se preservan para su futuro uso.

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t$$

$$r_t = \sigma(U_r h_{t-1} + W_r x_t)$$

$$z_t = \sigma(U_z h_{t-1} + W_z x_t)$$



Gracias por la atención

¿Tiene alguna pregunta?

