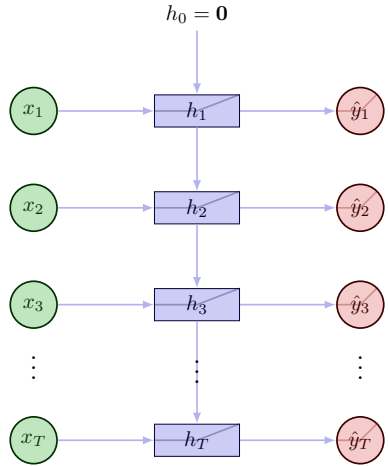
Pontificia Universidad JAVERIANA Cali

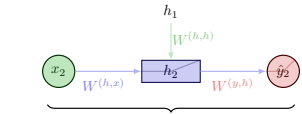
# Encuentro Sincrónico 5

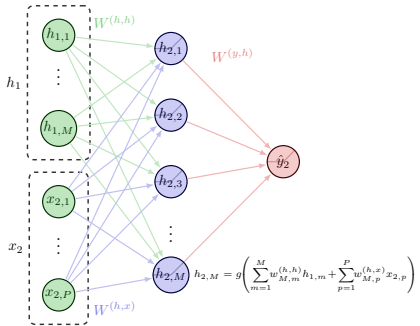
Modelos secuenciales y mecanismos de atención

# Redes recurrentes–(RNN)

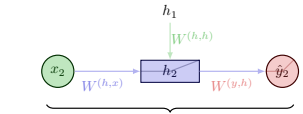


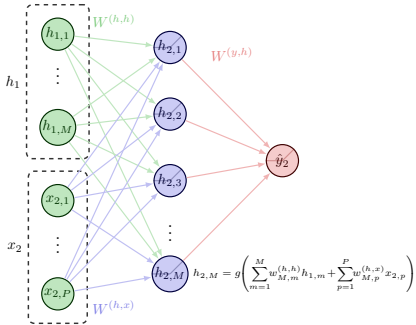
# Redes recurrentes–(RNN)

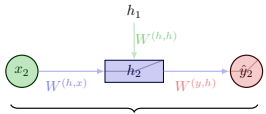


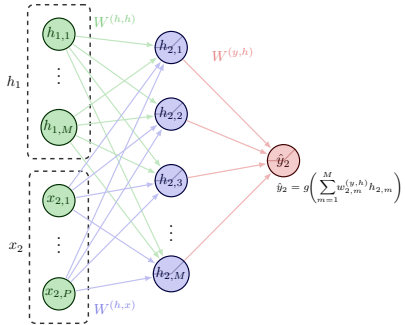



# Redes recurrentes–(RNN)

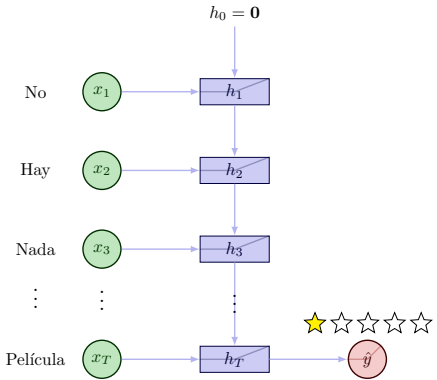


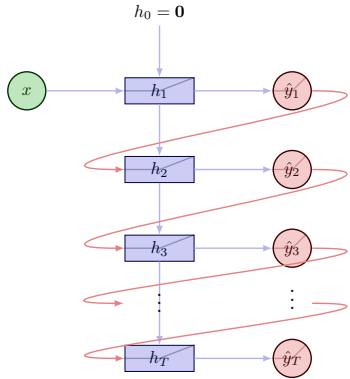



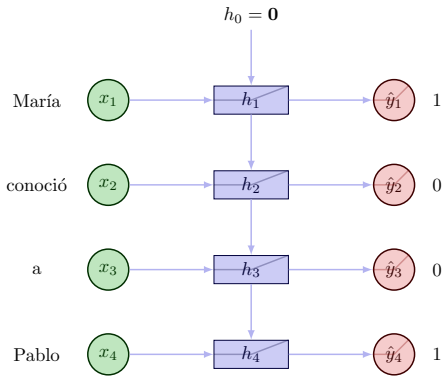



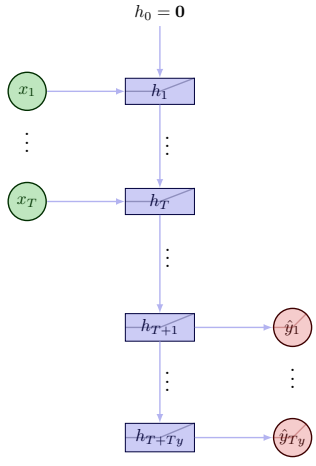
# Variaciones de las redes recurrentes



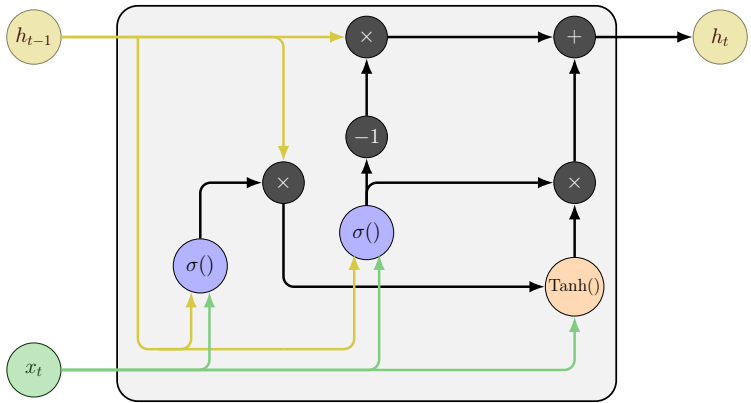



# Variaciones de las redes recurrentes



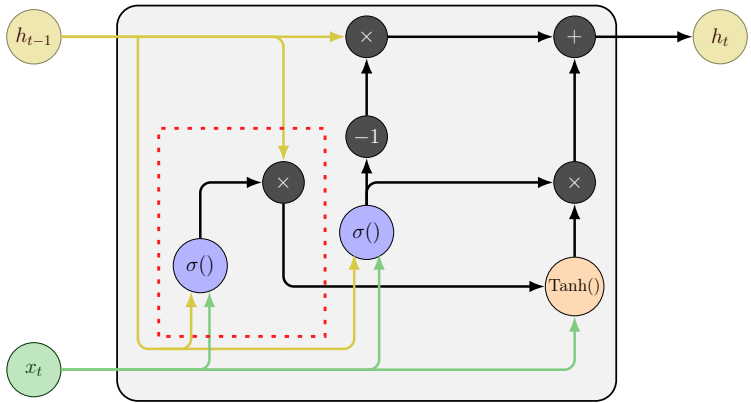



# Unidades recurrentes con compuerta–(GRU)



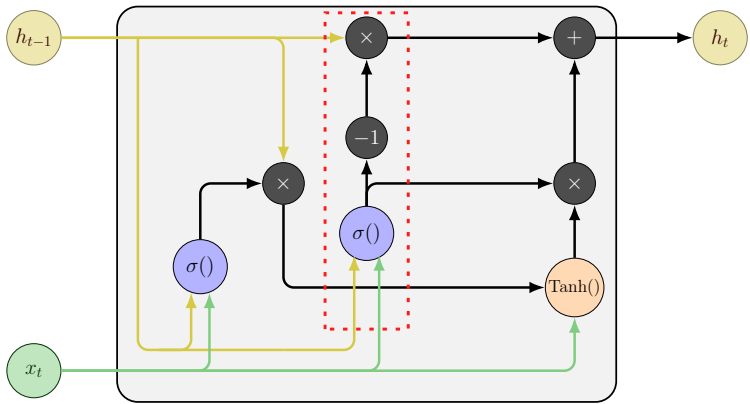
# Unidades recurrentes con compuerta–(GRU)

La compuerta de reinicio $\boldsymbol{r}_t = \sigma\left(\boldsymbol{W}^{(xr)}\boldsymbol{x}_t + \boldsymbol{h}_{t-1}\boldsymbol{W}^{(hr)} + \boldsymbol{b}_r\right), [0, 1]$



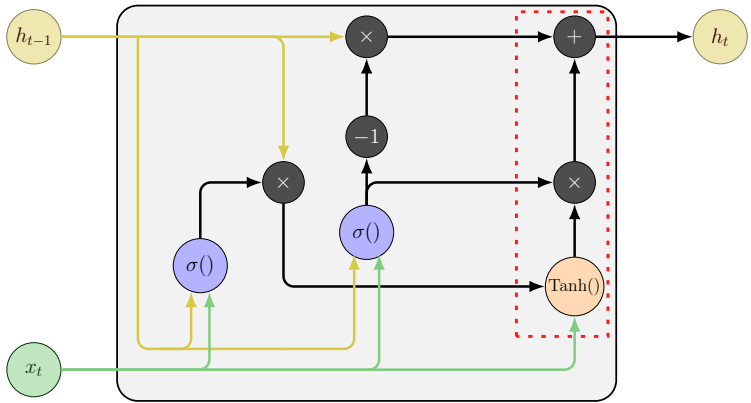
# Unidades recurrentes con compuerta–(GRU)

La compuerta de actualización $\mathbf{\Gamma}_t = \sigma\left(\boldsymbol{x}_t \boldsymbol{W}^{(xz)} + \boldsymbol{h}_{t-1} \boldsymbol{W}^{(hz)} + \boldsymbol{b}_z\right), [0, 1]$



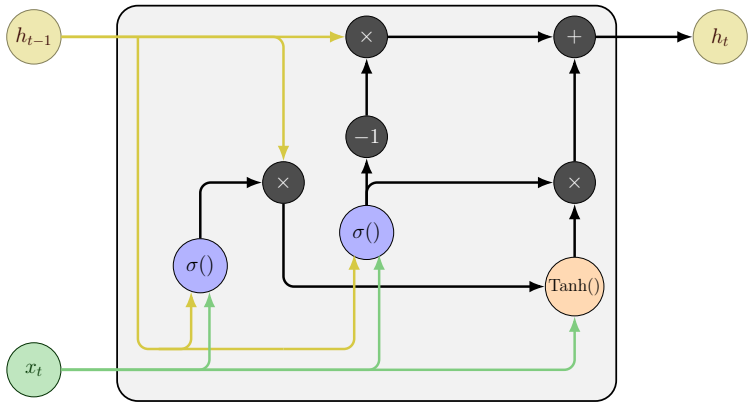
# Unidades recurrentes con compuerta–(GRU)

El candidato $\hat{\boldsymbol{h}}_t = \tanh\left(\boldsymbol{x}_t \boldsymbol{W}^{(xh)} + (\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}) \boldsymbol{W}^{(hh)} + \boldsymbol{b}_h\right)$



# Unidades recurrentes con compuerta–(GRU)

$$\boldsymbol{h}_t = \boldsymbol{\Gamma}_t \odot \boldsymbol{h}_{t-1} + (1 - \boldsymbol{\Gamma}_t)\hat{\boldsymbol{h}}_t$$



# Unidades recurrentes con compuerta–(GRU)

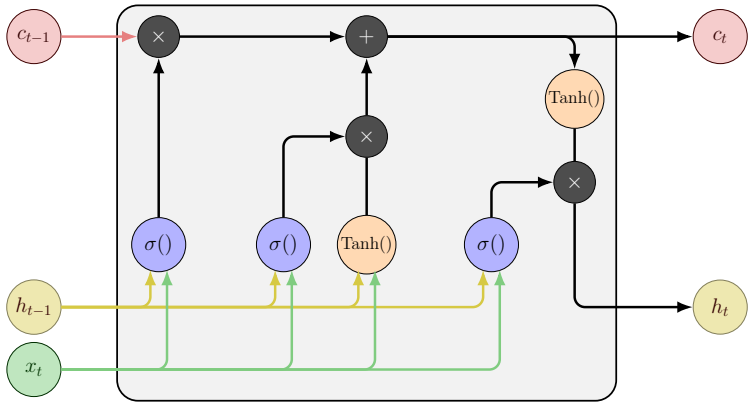
- Esto es útil en problemáticas como las que planteamos en la frase: *El perro que ya había comido ..., estaba lleno*.
- La red GRU podría entrenarse de tal forma que reconozca las palabras que son claves para resolver la tarea y descarte las menos importantes.

| $\Gamma_1 = 1$ | $\Gamma_2 = 0$ | $\Gamma_3 = 1$ | $\Gamma_4 = 1$ | $\Gamma_5 = 1$ | $\Gamma_6 = 1$ | | $\Gamma_{T-1} = 0$ | $\Gamma_T = 1$ |
|---|---|---|---|---|---|---|---|---|
| El | perro | que | ya | había | comido | $\cdots$ | estaba | lleno |

- Así, se eliminan las distancias entre las partes más importantes de la frase y reduce la posibilidad de que los gradientes tiendan a cero.

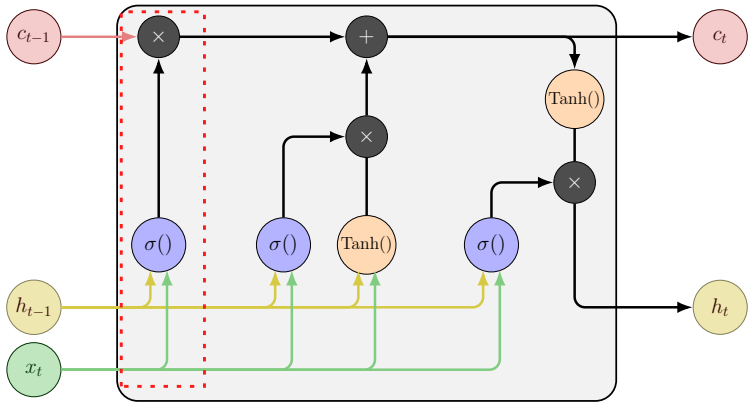
# Long Short-Term Memory–(LSTM)

$\boldsymbol{c_t}$ :Memoria a largo plazo, $\boldsymbol{h_t}$ :Memoria a corto plazo.



# Long Short-Term Memory–(LSTM)

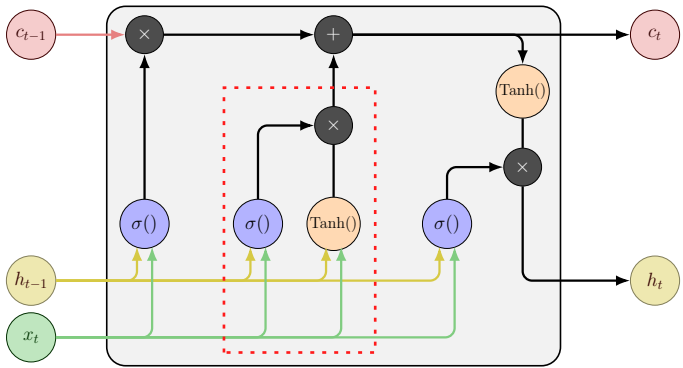
$\boldsymbol{f_t}$ :Compuerta de olvido, $\boldsymbol{f_t} : \sigma\left(\boldsymbol{x_t}\boldsymbol{W}^{(xf)} + \boldsymbol{h_{t-1}}\boldsymbol{W}^{(hf)} + \boldsymbol{b_f}\right),\ [0, 1]$



# Long Short-Term Memory–(LSTM)

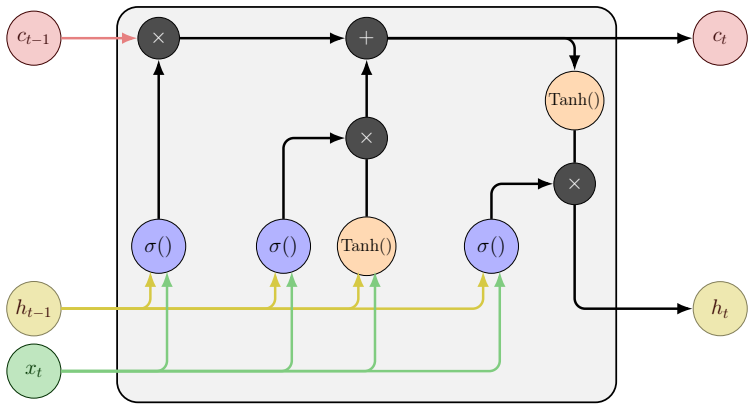
$\boldsymbol{i}_t$ :Compuerta de entrada,
$\boldsymbol{i}_t : \sigma\left(\boldsymbol{x}_t\boldsymbol{W}^{(xi)} + \boldsymbol{h}_{t-1}\boldsymbol{W}^{(hi)} + \boldsymbol{b}_i\right), [0, 1]$

$\hat{\boldsymbol{c}}_t$ :Candidato
$\hat{\boldsymbol{c}}_t = \tanh\left(\boldsymbol{x}_t\boldsymbol{W}^{(xc)} + \boldsymbol{h}_{t-1}\boldsymbol{W}^{(hc)} + \boldsymbol{b}_c\right)$



# Long Short-Term Memory–(LSTM)

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \hat{\boldsymbol{c}}_t$$



# Long Short-Term Memory–(LSTM)

$\boldsymbol{o_t}$ :Compuerta de salida,

$\boldsymbol{o_t} : \sigma\left(\boldsymbol{x_t}\boldsymbol{W}^{(xo)} + \boldsymbol{h_{t-1}}\boldsymbol{W}^{(ho)} + \boldsymbol{b_o}\right), [0, 1]$

$\boldsymbol{h_t}$ :Memoria a corto plazo

$\boldsymbol{h_t} = \boldsymbol{o_t} \odot \tanh\left(\boldsymbol{c_t}\right)$