

Día 06 · Tema 2 — Ejercicios: RNN, LSTM y GRU

Cómo se usa esta hoja. Hay dos tipos de ejercicio, y sólo dos.

El ejercicio del día. Uno por hoja. Es el obligatorio y es el que entra en el examen, con el cálculo incluido. Basta con hacer ese para seguir el curso con solvencia.

Los ejercicios de ampliación. El resto. Están para enseñar de qué es capaz el método, no para exigir: **no se piden calculados en el examen.** De ellos sólo puede preguntarse la idea, a nivel conceptual.

Tabla de referencia para los ejercicios de este cuaderno. Se proporcionan los valores de la sigmoide σ y la tangente hiperbólica necesarios para resolver los cálculos a mano.

Sigmoide $\sigma(x) = \frac{1}{1+e^{-x}}$	Tangente hiperbólica $\tanh(x)$
$\sigma(-2) = 0,119$	$\tanh(-1) = -0,762$
$\sigma(-1) = 0,269$	$\tanh(-0,850) = -0,691$
$\sigma(0) = 0,500$	$\tanh(-0,593) = -0,532$
$\sigma(1) = 0,731$	$\tanh(0) = 0$
$\sigma(2) = 0,881$	$\tanh(0,102) = 0,102$
	$\tanh(0,557) = 0,506$
	$\tanh(0,700) = 0,604$
	$\tanh(1) = 0,762$
	$\tanh(1,051) = 0,782$
	$\tanh(1,256) = 0,850$
	$\tanh(2) = 0,964$

EL EJERCICIO DEL DÍA *obligatorio · entra en el examen*

1. Pasada hacia delante en una RNN escalar

Consideremos una RNN con estado oculto escalar (una sola neurona), activación tanh y los siguientes pesos:

$$w_h = 0,5, \quad w_x = 0,3, \quad b = 0,1.$$

La ecuación de recurrencia es $h_t = \tanh(w_h h_{t-1} + w_x x_t + b)$. Partiendo de $h_0 = 0$, calcula h_1 , h_2 y h_3 para la secuencia de entrada $x_1 = 2$, $x_2 = -1$, $x_3 = 3$.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

2. Conteo de parámetros en una RNN

Una RNN simple tiene $d = 5$ rasgos de entrada y $d_h = 10$ unidades ocultas. La capa de salida produce un único escalar $\hat{y}_t = W_o h_t + b_o$. Calcula el número de parámetros de cada matriz y vector, y el total.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

3. Gradiente evanescente en la RNN escalar

Usando la misma RNN escalar del ejercicio 1, calcula la magnitud de $\partial h_3 / \partial h_1$ mediante la regla de la cadena. Interpreta el resultado en relación con el problema del gradiente evanescente.

Pista: para la activación tanh, la derivada de $h_t = \tanh(z_t)$ respecto a h_{t-1} es $\frac{\partial h_t}{\partial h_{t-1}} = w_h (1 - h_t^2)$.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

4. Pasada hacia delante en una LSTM escalar

Consideremos una LSTM con estado oculto y estado de celda escalares (una sola unidad). Las ecuaciones son:

$$\begin{aligned} f_t &= \sigma(w_f^x x_t + b_f), & i_t &= \sigma(w_i^x x_t + b_i), \\ \tilde{C}_t &= \tanh(w_c^x x_t + b_c), & o_t &= \sigma(w_o^x x_t + b_o), \\ C_t &= f_t C_{t-1} + i_t \tilde{C}_t, & h_t &= o_t \tanh(C_t). \end{aligned}$$

Los parámetros son:

Puerta de olvido:	$w_f^x = 0, b_f = 1,$
Puerta de entrada:	$w_i^x = 1, b_i = 0,$
Candidato:	$w_c^x = 1, b_c = 0,$
Puerta de salida:	$w_o^x = 0, b_o = 0.$

Con $h_0 = 0, C_0 = 0, x_1 = 1$ y $x_2 = 2$, calcula C_1, h_1, C_2 y h_2 paso a paso. Comenta qué papel juega cada puerta.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

5. Pasada hacia delante en una GRU escalar

Una GRU escalar sigue las ecuaciones:

$$\begin{aligned}z_t &= \sigma(w_z^x x_t + b_z), \\r_t &= \sigma(w_r^x x_t + b_r), \\\tilde{h}_t &= \tanh(w_h^h (r_t \cdot h_{t-1}) + w_h^x x_t + b_h), \\h_t &= (1 - z_t) h_{t-1} + z_t \tilde{h}_t.\end{aligned}$$

Los parámetros son:

Puerta de actualización:	$w_z^x = 1, b_z = 0,$
Puerta de reset:	$w_r^x = 1, b_r = 0,$
Candidato:	$w_h^h = 1, w_h^x = 1, b_h = 0.$

Con $h_0 = 0, x_1 = 1$ y $x_2 = -1$, calcula h_1 y h_2 paso a paso. Explica qué ocurre con las puertas cuando el input cambia de signo.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

6. Conteo de parámetros: LSTM vs. GRU

Supongamos una capa recurrente con tamaño de entrada $d = 8$ y tamaño oculto $d_h = 16$. Calcula el número total de parámetros de la parte recurrente (sin contar la capa de salida) para una LSTM y para una GRU.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

7. Cálculo del MAPE

Un modelo predice ventas semanales para cuatro productos. Los valores reales y las predicciones son:

Producto	A	B	C	D
Real y_i	100	150	200	80
Predicción $\hat{y}_i$	110	140	220	90

Calcula el MAPE (*Mean Absolute Percentage Error*). Comenta por qué el MAPE puede ser problemático si algún valor real fuera cercano a cero.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

8. Diseño de una RNN para predicción de ventas

Un minorista dispone de datos de ventas semanales durante 52 semanas para 100 SKUs (*Stock Keeping Units*). Cada semana se observan $d = 3$ rasgos por SKU (unidades vendidas, precio medio, indicador de promoción). Se desea entrenar una RNN simple con $d_h = 32$ unidades ocultas para predecir las ventas de la semana siguiente.

- ¿Cuáles son las dimensiones de W_x y W_h ?
- ¿Cuál es la forma del tensor de entrada para un batch completo?
- Si se sustituye la RNN simple por una LSTM, ¿cuántos parámetros tiene la parte recurrente?