

Introducción al Deep Learning

Día 3: Redes neuronales recurrentes

Manuel Germán y David de la Rosa
Universidad de Jaén



Universidad
de Jaén



`(mgerman, drrosa)@ujaen.es`

Expectativas

Tras esta sesión, sabremos:

- Qué es una Red Neuronal Recurrente (RNN).
- Cómo procesa una RNN una secuencia.
- Qué problemas presentan las RNN.
- Cuáles son los diferentes tipos de celdas recurrentes.
- Cómo implementar todo lo anterior usando *Pytorch* y *Pytorch Lightning*.

1

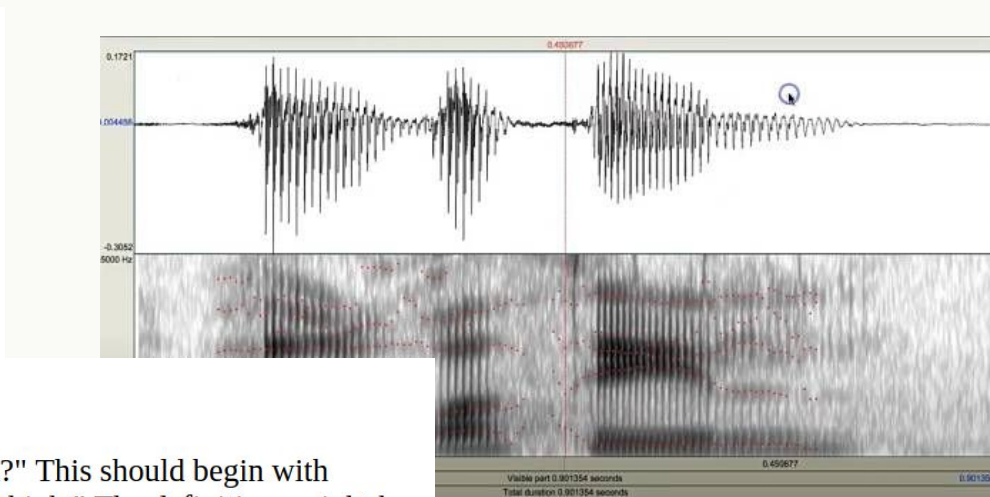
Datos secuenciales

Yellowstone Park

Visitors per month

Temperature (F)

Visitors per month — Temperature (F)



I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the

Secuencias

Una gran cantidad de dominios presentan datos secuenciales



Secuencias

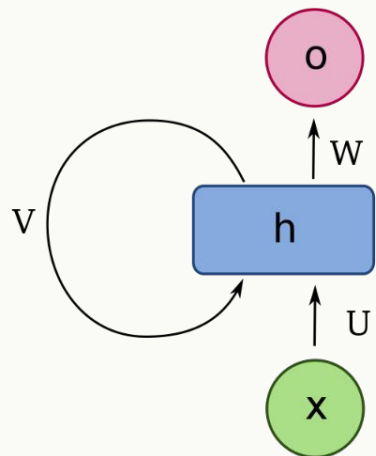
Una gran cantidad de dominios presentan datos secuenciales



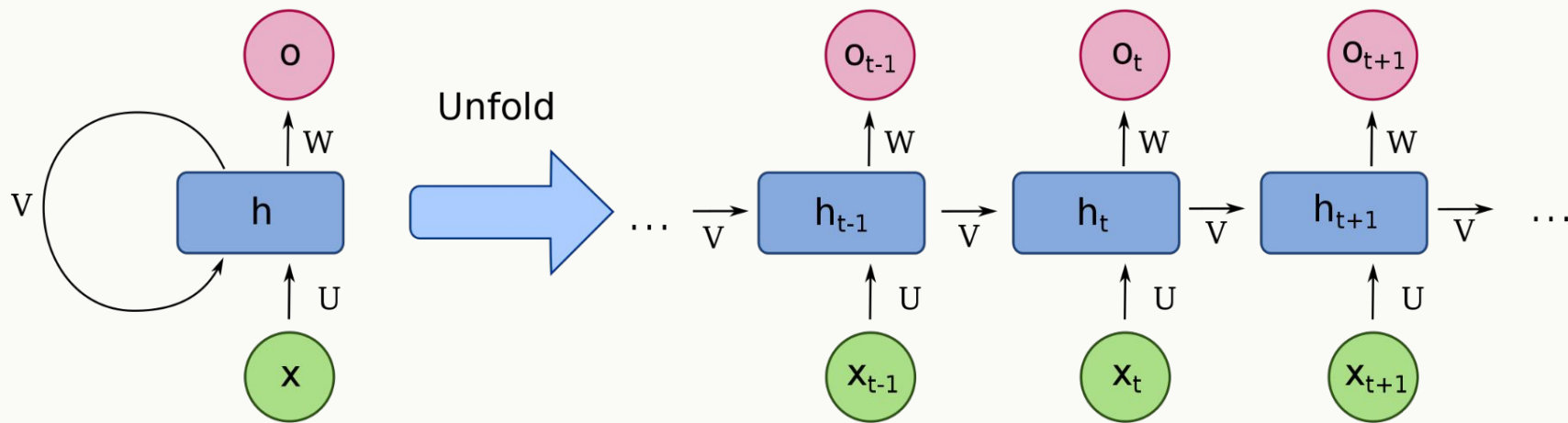
2

Redes neuronales recurrentes

¿Qué es una RNN?



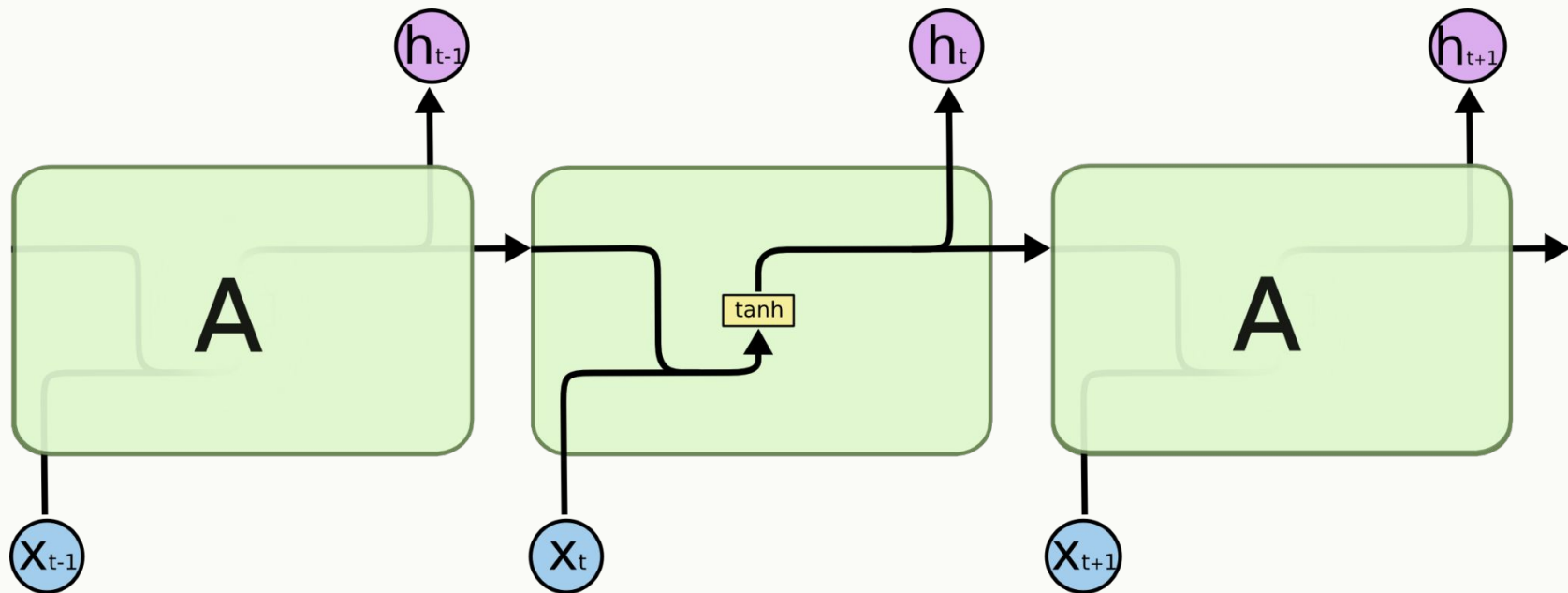
¿Qué es una RNN?



$$h_t = f(x_t \cdot U + h_{t-1} \cdot V)$$

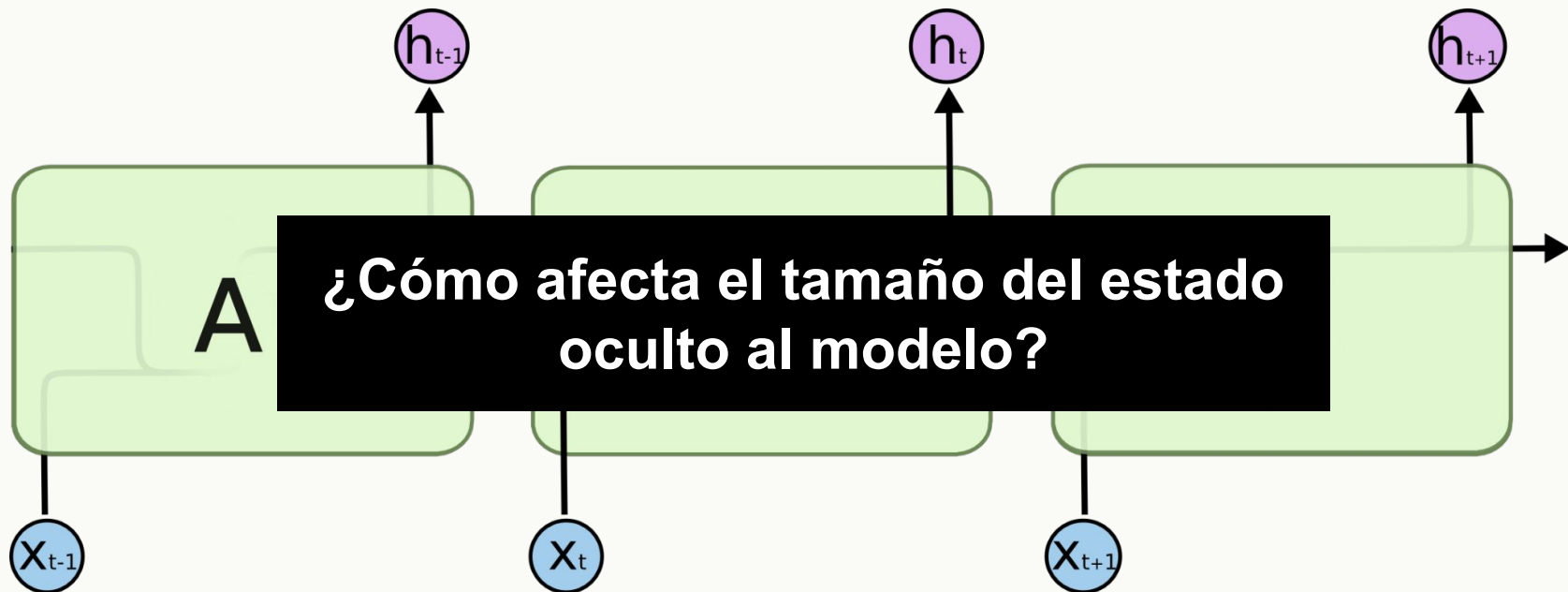
$$o_t = f(h_t \cdot W)$$

¿Qué es una RNN?



$$h_t = \tanh(x_t \cdot U + h_{t-1} \cdot V)$$

¿Qué es una RNN?



$$h_t = \tanh(x_t \cdot U + h_{t-1} \cdot V)$$

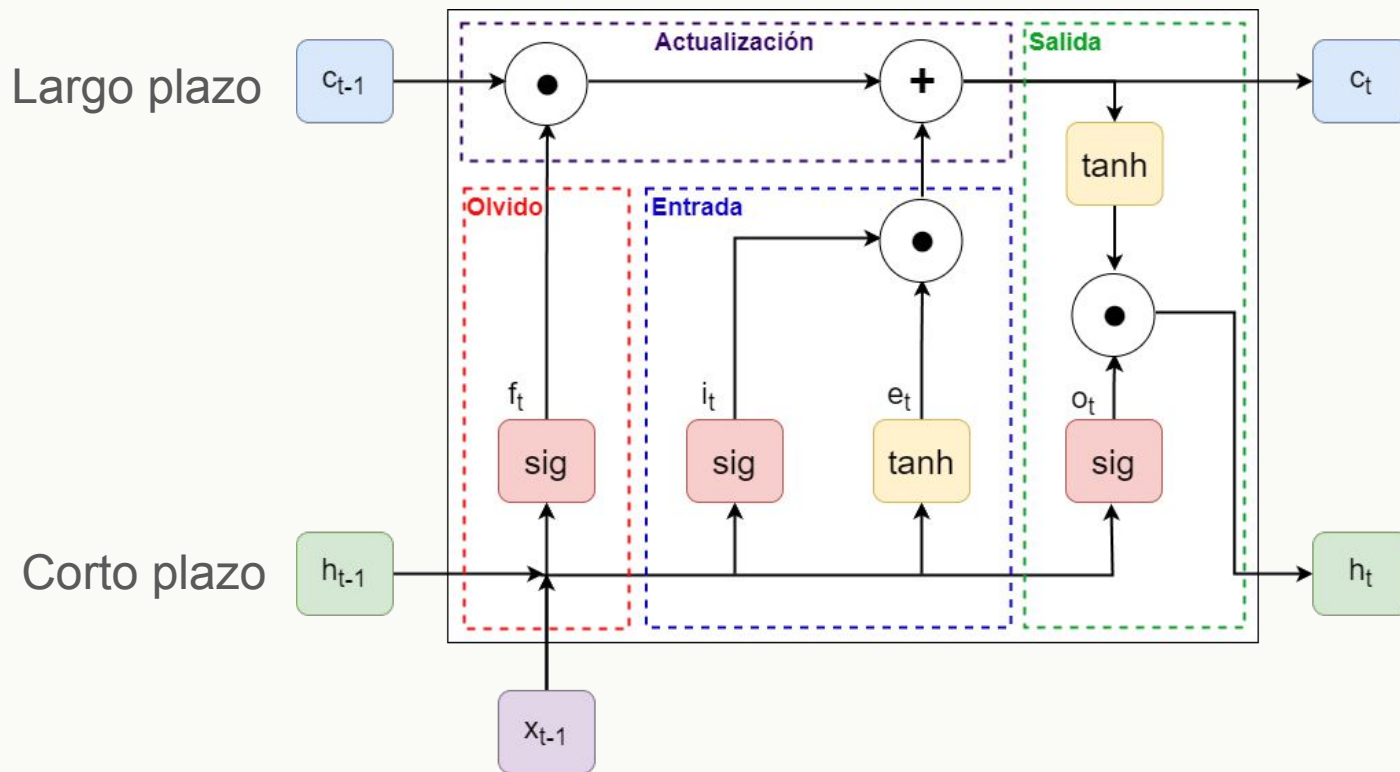
Problemas

- Pérdida de dependencias temporales en ventanas grandes.
- Explosión/desvanecimiento del gradiente.
- Imposibilidad de paralelizar el entrenamiento.

Modificaciones propuestas

Nuevos modelos

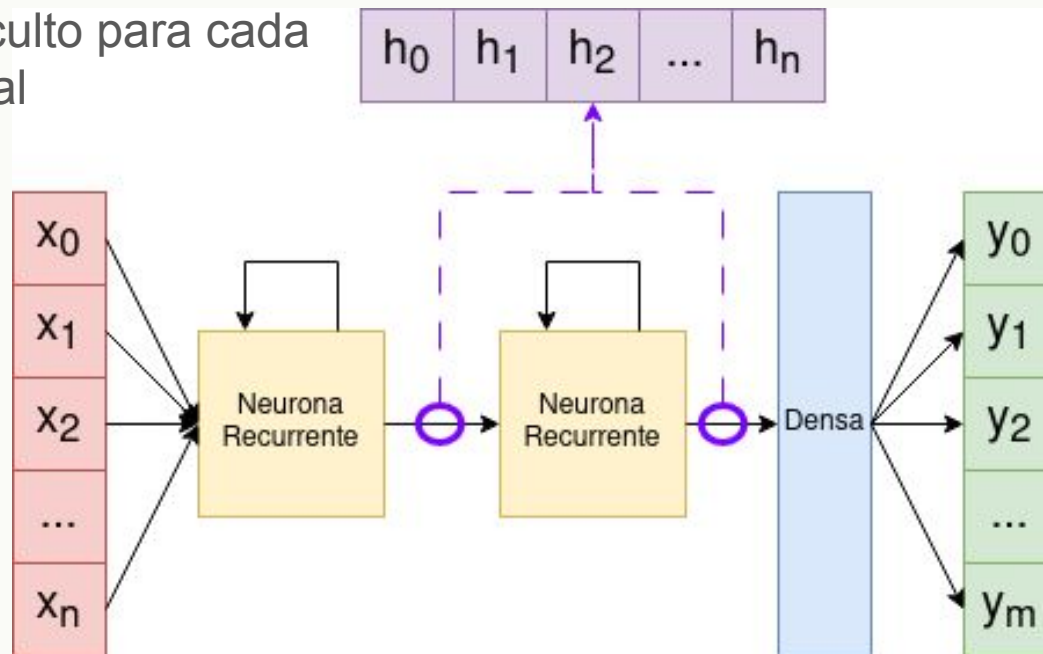
Long-Short Term Memory (LSTM)





Apilando capas recurrentes

Un estado oculto para cada paso temporal

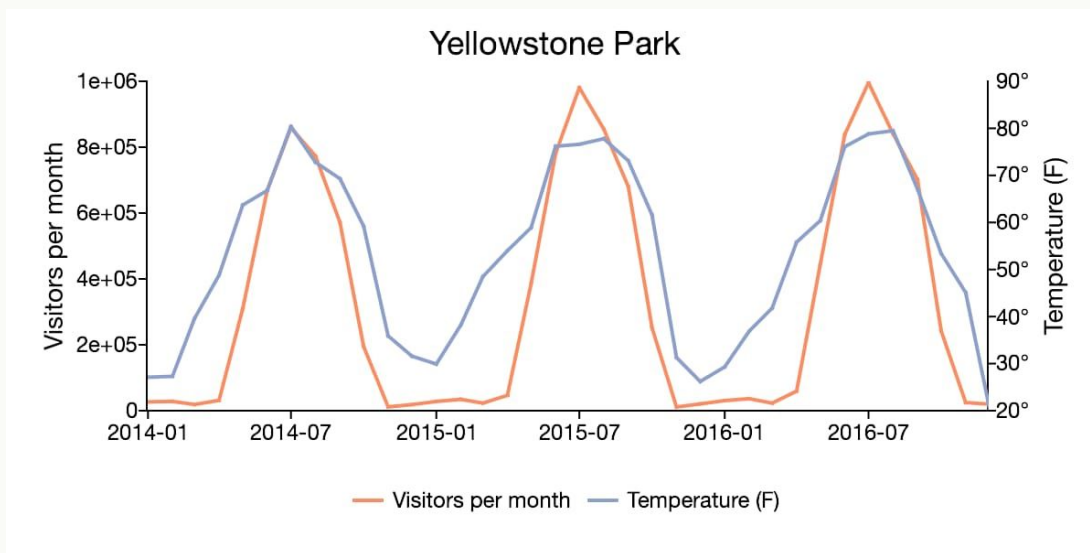


3

Predicción de series temporales

Definición de serie temporal

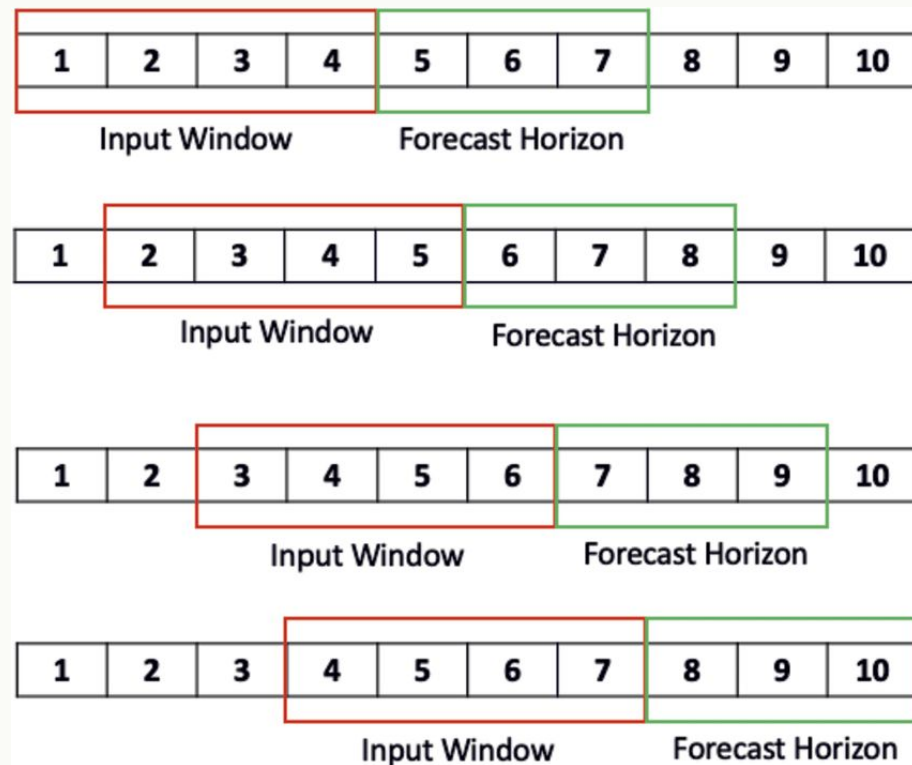
Conjunto de observaciones ordenadas en el tiempo.



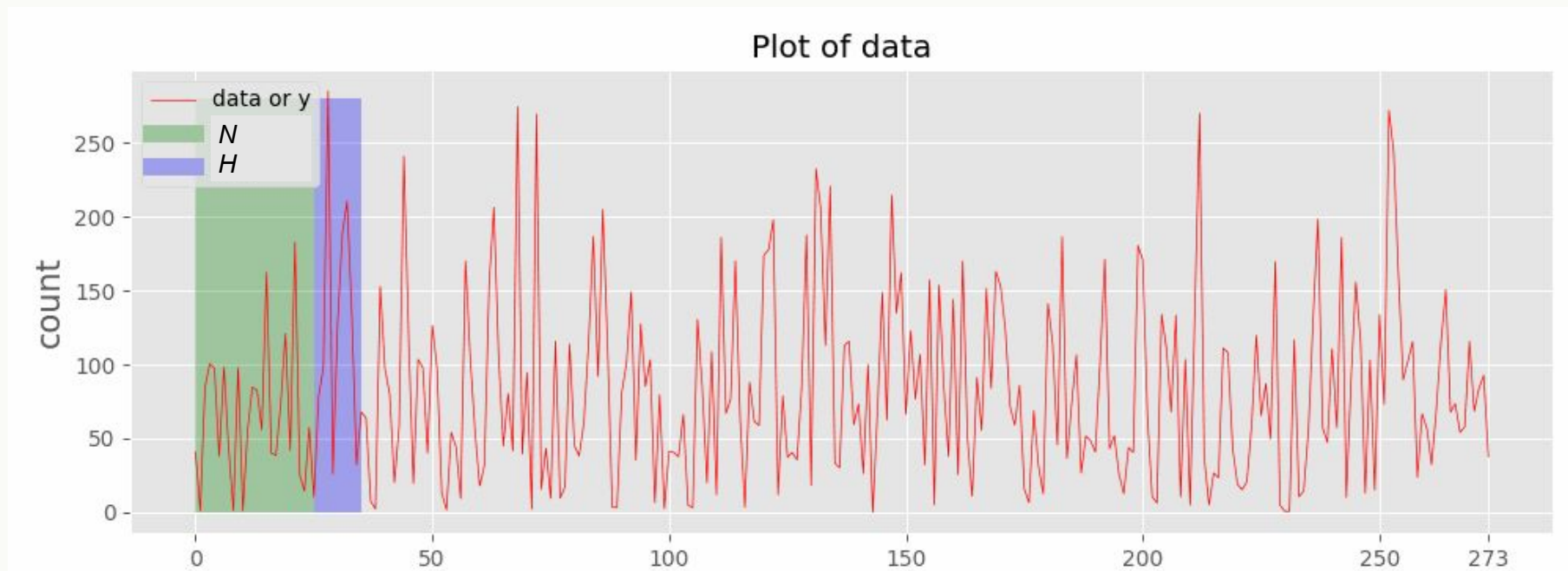
Conversión a problema supervisado: Ventanas deslizantes

Una ventana está compuesta por N instantes (*lags*) de **entrada** y H instantes de **salida** (*horizonte de predicción*).

N instantes actúan como **entradas** para el modelo. H instantes serán la “etiqueta” que queremos que el modelo tiene que predecir.



Conversión a problema supervisado: Ventanas deslizantes



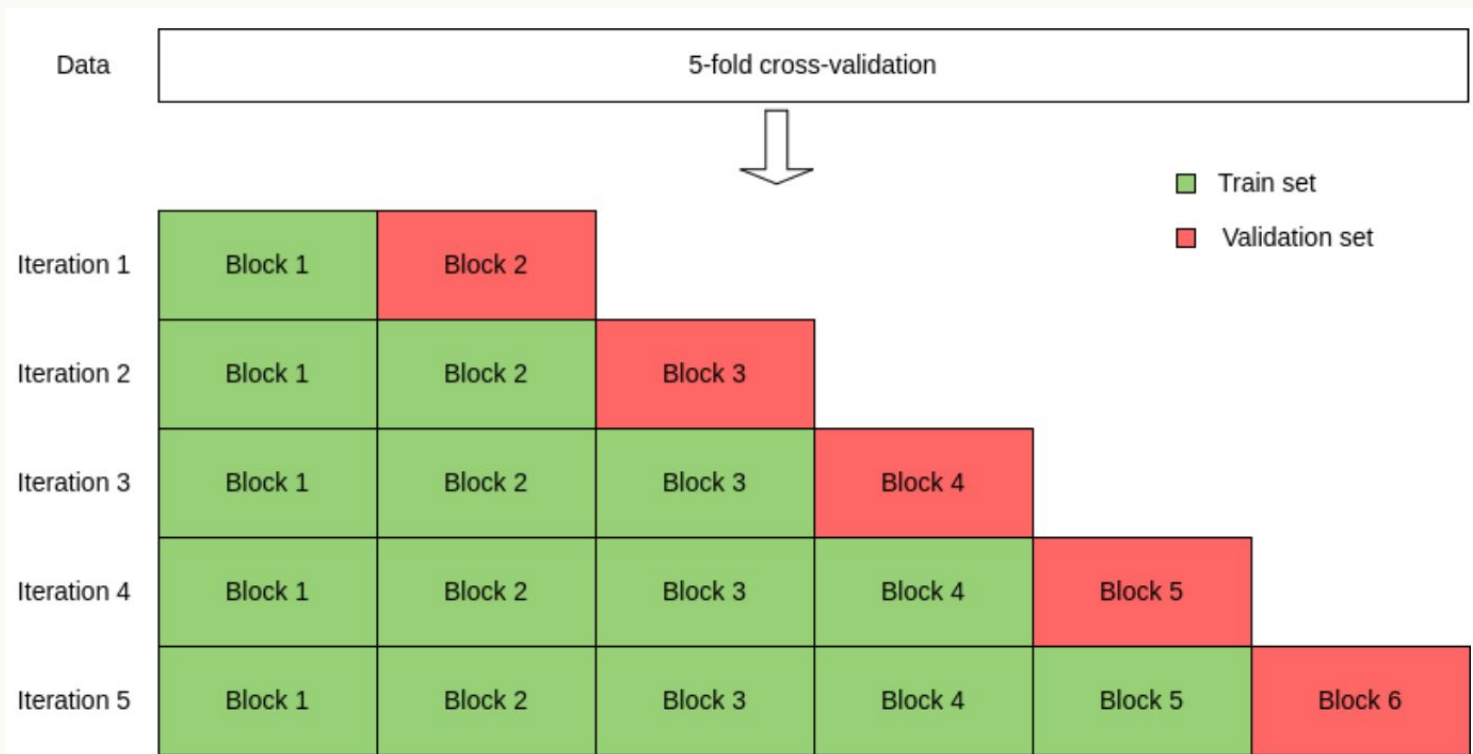
Métricas de evaluación

Métrica	Fórmula
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{n} \sum_{t=1}^n y_t - \hat{y}_t $
Root Mean Squared Error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$
Mean Squared Error (MSE)	$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$
Mean Absolute Scaled Error (MASE)	$\text{MASE} = \frac{\text{MAE}}{\text{MAE}_{\text{naive}}}$
Mean Absolute Percentage Error (MAPE)	$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left \frac{y_t - \hat{y}_t}{y_t} \right \times 100$
Symmetric Mean Absolute Percentage Error (sMAPE)	$\text{sMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{ y_t - \hat{y}_t }{(y_t + \hat{y}_t)} \times 200$

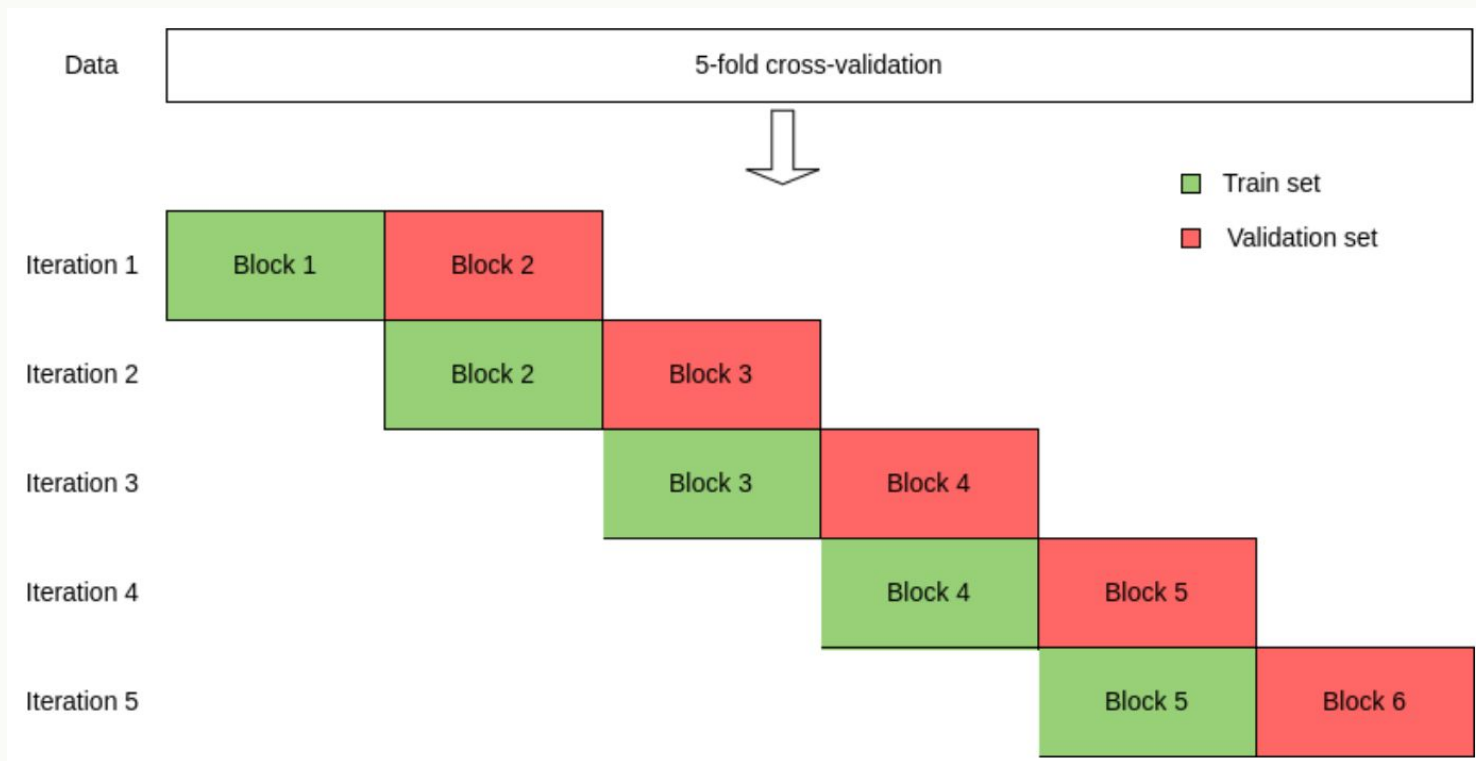
MAE, RMSE, MSE son absolutas. MASE, MAPE y sMAPE relativas.

También es interesante realizar un **análisis de los residuos** para comprobar si el modelo ha sido capaz de aproximar los comportamientos de la serie.

Validación cruzada en TSF



Validación cruzada en TSF



Esquema general

- **Particionamos** conjunto de datos en Train-Test.
- Asignamos rango de hiper-parámetros a optimizar.
- **Aplicamos CV** con el conjunto de entrenamiento, agregamos las métricas por *fold*.
- Creamos modelo con los **mejores parámetros**. Lo ajustamos con el **100%** del conjunto de train.
- **Evaluamos** el modelo mediante el conjunto de test.

Introducción al Deep Learning

Día 3: Redes neuronales recurrentes

Manuel Germán y David de la Rosa
Universidad de Jaén



Universidad
de Jaén



`(mgerman, drrosa)@ujaen.es`

Material complementario

- En <https://mgermanm0.github.io/intro-ts/> se detallan algunas técnicas para pre-procesar y analizar con más detalle series temporales.
- En GitHub se subirá un cuaderno sobre el manejo de tensores en Pytorch.
- Existe una versión mucho más extendida de esta sesión. Si te interesa, dímelo y te la comparto.