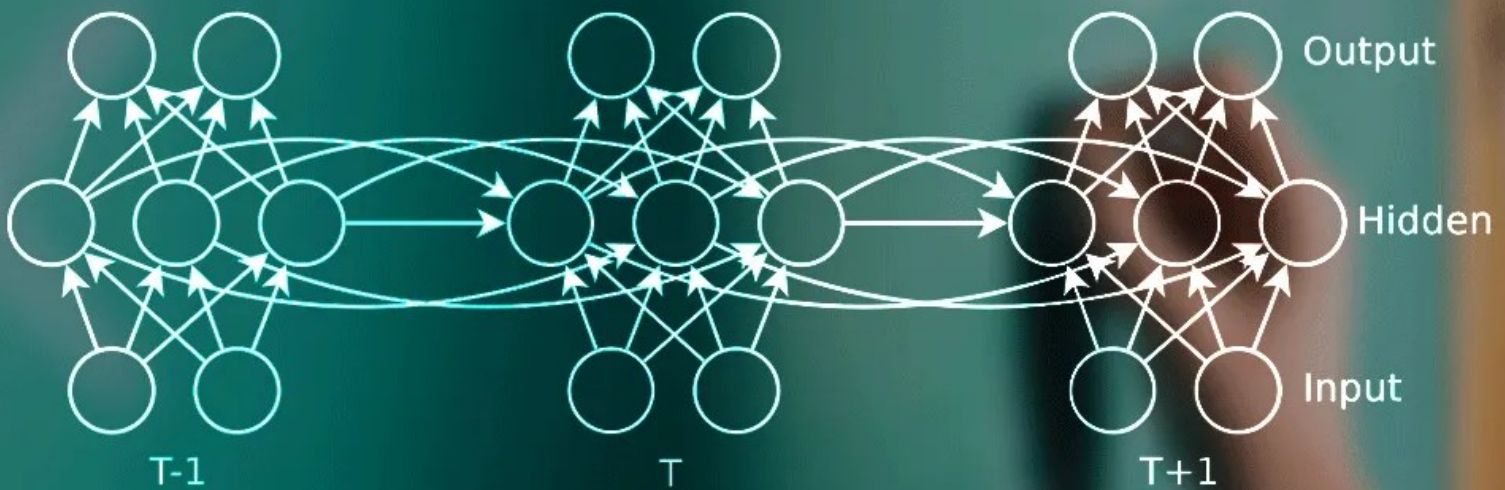


## Práctica 7

# Redes Neuronales Recurrentes



### Objetivos

El objetivo de esta práctica es comprender el funcionamiento de las redes neuronales recurrentes

### Temas

- Redes Neuronales Recurrentes: Redes Recurrentes Simples, Redes LSTM

### Lectura

Material de Lectura: Capítulos 12 a 14 del libro Time Series Forecasting in Python.

## Ejercicio 1

### Preparación de Dataset

El dataset **Beijing Multi-Site Air-Quality Data** registra 6 contaminantes atmosféricos principales y 6 variables meteorológicas relevantes en múltiples emplazamientos de Pekín, China. Incluye datos de contaminantes del aire por hora de 12 sitios de monitoreo de calidad del aire controlados a nivel nacional. Los datos de calidad del aire provienen del Centro de Monitoreo Ambiental Municipal de Beijing. Los datos meteorológicos en cada sitio de calidad del aire se emparejan con la estación meteorológica más cercana de la Administración Meteorológica de China. El período de tiempo va desde el 1 de marzo de 2013 hasta el 28 de febrero de 2017. Los datos faltantes se indican como NA.

El repositorio original del dataset encuentra disponible en el repositorio UCI y se encuentra disponible en la url: <https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data>

Utilice la versión del dataset limitada a la estación meteorológica de **Aotizhongxin** disponible en la carpeta de datos de la cátedra (**Beijing\_air\_quality\_Aotizhongxin.csv**). Prepare los datos para trabajar en modelos con diferentes arquitecturas:

- Cargue el archivo y realice un análisis de los atributos y sus valores faltantes. En particular analice y grafique el atributo CO alrededor de la muestra 16.000 para eliminar todas las muestras anteriores debido a la gran cantidad de valores faltantes.
- Dado que la fecha y hora está separada en otros atributos, cree un nuevo atributo que los unifique. Luego, complete los valores faltantes con el promedio de otras filas con la misma hora. Puede utilizar la función groupby y transform sobre el nuevo atributo.
- Elimine los atributos que no sean numéricos o relevantes para realizar predicciones.
- Implemente una función parametrizada que divida los datos en un grupo de entrenamiento, validación y prueba. Muestre en un gráfico cada grupo de un color diferente.
- Implemente una función parametrizada que permita crear una versión del dataset para trabajar con series. La función recibe el dataset, la cantidad de muestras de entrada a procesar y la cantidad de muestras que deben ser predichas y retorna dos arreglos: uno con las secuencias de datos de entrada para entrenamiento y otro con las secuencias de datos de salida correspondiente.

## Ejercicio 2

Utilizando los datos preprocesados del ejercicio anterior y aplicando normalización lineal, entrene los siguientes modelos de diferente arquitectura para predecir un valor del atributo **NO2** (dióxido de nitrógeno) a partir de los valores correspondientes a las 24 horas previas (univariada-simple):

- Un combinador Lineal.
- Un multiperceptrón.
- Una red recurrente simple.
- Una red LSTM.

Luego haga un gráfico de barra comparando los errores en entrenamiento y validación de los 4 modelos.

### Ejercicio 3

Utilizando los datos preprocesados, genere modelos de redes LSTM para:

- a) Predecir los valores de **NO2** para una secuencia de 4 horas a partir de las 24 horas anteriores (univariada-múltiple)
- b) Predecir un valor de **NO2** a partir de las 24 horas anteriores de los atributos **PM2.5, PM10, SO2, NO2, CO, O3** y **WSPM** (multivariada-simple).
- c) Predecir los valores de **NO2** para una secuencia de 4 horas a partir de las 24 horas anteriores de los atributos **PM2.5, PM10, SO2, NO2, CO, O3** y **WSPM** (multivariada-múltiple).