

---

## Aprendizaje Profundo

### Práctica 3: Recurrent Neural Networks

4 de diciembre de 2024

---

En esta práctica deberás desarrollar un modelo que sea capaz de predecir el precio del dólar en euros con 6h de adelanto. Para ello, contarás con el fichero de datos UDC\_EUR.json, disponible en el canal de Teams de la asignatura.

La idea es hacer un modelo que te permita decidir cuándo comprar y cuándo vender dólares, para que puedas operar de la siguiente manera:



- Si el modelo predice que el dólar bajará: vendo un dólar ahora y lo compro más barato dentro de 6h. La diferencia entre el precio de venta y el de compra es mi ganancia.
- Si el modelo predice que el dólar subirá: compro un dólar ahora y lo vendo más caro dentro de 6h. La diferencia entre el precio de compra y el de venta es mi ganancia.
- Puedes suministrarle al modelo los datos de entrada que estimes para cada predicción, siempre respetando que ninguno esté a menos de 6h del que intentas predecir.

Debes respetar las siguientes restricciones:



- No se deben utilizar datos posteriores a 31/08/2022 23:59:59 para entrenamiento. Los datos desde 01/09/2022 00:00:00 en adelante se podrán utilizar como conjunto de test.
- La predicción se deberá hacer a 6h vista, es decir, se deberá predecir el valor de salida 6 horas después del dato más reciente que se le proporcione al modelo. Se deberán predecir dos valores distintos:
  - Valor de la variable "precio fin".
  - Valor binario que indicará si "precio fin" 6h después será mayor o menor que en el dato más reciente proporcionado a la entrada.

## Trabajo a realizar

1. (1.5 puntos) Carga y examen de los datos:
  - (1 punto) Carga los datos con Pandas y haz un examen preliminar de los mismos. Descarta aquellos que sean erróneos.
  - (0.5 puntos) Haz la partición entre conjunto de entrenamiento y conjunto de test.
2. (2 puntos) Preprocesado:
  - (0.5 puntos) Representa las variables en un formato que facilite el aprendizaje. Para ello estandariza aquellas que lo necesiten y transforma las fechas a un formato adecuado.
  - (0.5 puntos) Transforma los conjuntos de entrenamiento y test a NumPy.
  - (1 punto) Utiliza la función `sliding_window` del Lab11 para ventanear la serie temporal para obtener un lote de muchos pares (vector de entrada, etiqueta) para entrenamiento y otro lote para test:
    - Haz que el tamaño de ventana sea configurable para poder probar varias alternativas.
    - Revisa la continuidad temporal de la serie para que las ventanas representen siempre el mismo periodo de tiempo. Descarta las ventanas incorrectas.
3. (3.5 puntos) Elección del/los modelo/s. Prueba con ventanas de distinto tamaño, y ajusta a los datos de entrenamiento a modelos con arquitecturas de tipo:

- (0.5 puntos) Totalmente conectado.
- (1 punto) LSTM y GRU.
- (1 punto) LSTM y GRU anidados. Se requieren al menos dos capas.
- (1 punto) Transformer (No es necesario entrenarlo, sólo prepararlo para su uso).

4. (4 puntos) Evaluación del rendimiento del/los modelo/s frente a baselines.

- (2 puntos) Compara el rendimiento de tus modelos frente entre sí y con dos baselines sencillas (diseña dos baselines similares para el problema binario):
  - (1 punto) Un modelo que prediga siempre el último valor de entrada recibido (6h antes del dato a predecir).
  - (1 punto) Un modelo que prediga la media de los valores de entrada recibidos.
- (0.5 puntos) Identifica sesgos en tu modelo. Para ello, analiza las distribuciones de entrenamiento y test y las predicciones/errores de entrenamiento/test. Si identificas algún problema, adapta tu pipeline para intentar solventarlo.
- (1.5 puntos) Busca dar respuesta a las preguntas, ¿por cuántos céntimos de dólar falla cada modelo? ¿Cuánto puedo esperar ganar con mi modelo durante el periodo de test, haciendo las operaciones de 1\$?.

- Se entregarán en el mismo *assignment* de Classroom que la práctica 1 (<https://classroom.github.com/a/WdUBDmBc>) como uno o varios cuadernos nuevos. Puedes organizarlos en directorios si lo consideras necesario.
- Se deberá crear un tag llamado **P3**. El contenido de este tag será lo que se evalúe.
- La fecha de entrega límite es el día **20 de diciembre de 2024** a las 23:50 horas.
- Se valorarán las explicaciones de los resultados:
  - El proceso está suficientemente detallado y las decisiones adoptadas están justificadas.
  - Los resultados se comentan e interpretan correctamente.

