2:

#### Máster en Big Data

## Asignatura: Casos de analítica

## Fecha: 3/05/2021

Apellidos: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Nombre: \_\_\_\_\_\_\_\_\_\_ Login : \_\_\_\_\_\_\_\_\_\_ Curso: **2021/22**

**Caso 2: UrbanSound 8K *dataset***

En este caso vamos a clasificar eventos acústicos de la ciudad de Nueva York. La contaminación acústica en entornos urbanos es un problema en auge que las administraciones públicas quieren solventar. Por esto, desde la NYU se ha confeccionado un *dataset* con más de 8000 ficheros de audio de 10 fuentes de ruido distintas para el desarrollo de modelos de clasificación automática (<https://urbansounddataset.weebly.com/urbansound8k.html>). Con estos datos, se os propone que extraigáis características acústicas de la señal (el espectrograma) i que utilicéis técnicas de *deep learning* para la clasificación de eventos acústicos. Explicitad y detallad todos los pasos hechos para responder a cada pregunta.

Primera Parte: Análisis Cuantitativo.

* 1. Primer examen preliminar del *dataset*. ¿Qué clases componen *dataset*?
  2. Mostrad en una tabla u histograma el número de ficheros totales y tiempo por clase de cada *fold* del dataset.
  3. Ahora sumad el número de ficheros y de tiempo por clase de todo el *dataset* y razonad si éste está balanceado.

Segunda Parte: Extracción de Características y Análisis Cualitativo.

Al trabajar con señales de audio, es muy frecuente que un primer paso sea extraer características frecuenciales (frecuencia máxima, frecuencia mínima, MFCC, …), temporales (ZCR, envolvente acústica, …) o espectro-temporales (espectrograma, espectrograma mel, …) de las señales. Para solventar este caso, vamos a utilizar los espectrogramas de la señal como características para poder clasificarlas. Podéis utilizar o bien los espectrogramas o bien los espectrogramas mel de la señal.

* 1. Mostrad un espectrograma o espectrograma mel de cada clase del dataset.
  2. ¿Qué rango frecuencial tiene el espectrograma? ¿Y temporal? ¿Cómo vais a abordar el hecho de que hay ficheros de audio en el *dataset* con distintas duraciones?
  3. Generad los espectrogramas para todo el dataset y guardároslo en disco. Se recomienda el uso de ficheros HDF5. Tened en cuenta que para la clasificación vamos a utilizar redes neuronales que esperan entradas de 224x224, no es necesario que guardéis más píxeles.

Tercera Parte: Análisis Predictivo.

Ahora vais a clasificar los espectrogramas. Para hacerlo, debéis utilizar el método de *fine-tuning* utilizando una red neuronal profunda. Debéis tener en cuenta que típicamente estas redes han sido entrenadas con imágenes con canales R, G y B, pero que vuestros espectrogramas únicamente nos proporcionan información en cuanto a variaciones temporales (eje horizontal), variaciones frecuenciales (eje vertical) e intensidad. Por lo tanto, no tiene sentido utilizar el color del *plot* del espectrograma como característica. Siendo así, una posible solución es utilizar el espectrograma replicado tres veces para tener las tres capas (R, G y B) que espera como entrada la red neuronal. Os recomendamos que utilicéis la librería *pytorch* para solventar el problema. Podéis utilizar el siguiente enlace como referencia: <https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html>

* 1. Seleccionad como mínimo una red neuronal y explicad su arquitectura (número de capas, estructura de cada capa, etc.).
  2. Entrenad el modelo para clasificar el *dataset* UrbanSound 8K[[1]](#footnote-1).
  3. Explicad qué valores de hiperparámetros principales habéis utilizado (*batch size, learning rate, optimizer, número de epochs*) para obtener los resultados.
  4. Mostrad las gráficas de *accuracy* de entrenamiento y validación del modelo. Opcionalmente, también podéis mostrar las gráficas de *loss*.
  5. Mostrad la matriz de confusión obtenida tras la clasificación. Razonad qué clases clasifica mejor o peor.

Consideraciones:

* Este ejercicio puede realizarse o individualmente o en parejas.
* **Puntos extra**: seleccionad más de una red neuronal pre-entrenada o cread vuestra propia red neuronal profunda desde cero para hacer una comparativa con los resultados obtenidos con la red que habéis utilizado en el ejercicio.
* La fecha límite para entregar el ejercicio es el 15 de junio de 2022.
* Podéis entregar o bien esta memoria rellenada con las respuestas a cada apartado y un fichero con el código aparte, o bien un *notebook* bien estructurado en el que se puedan ver claramente las respuestas a las preguntas.

1. La idea es que realicéis el ejercicio con el *dataset* entero, pero si veis que os tarda mucho tiempo en entrenar podéis optar por reducir el número de *folds* que utilicéis. Es importante que no mezcléis los ficheros de audio entre *folds* de entrenamiento y clasificación y que indiquéis en la memoria del caso qué *folds* concretos habéis utilizado. [↑](#footnote-ref-1)