

Práctica 4 - TDS

Características (*features*) de la voz

Extrae_Características.m

Práctica 4

OBJETIVOS

1. A partir del espectrograma Mel calcularemos los coeficientes MFCC (*Mel Frequency Cepstral Coefficients*)
2. A partir del espectrograma lineal extraeremos otras características (*features*) espectrales.
3. Construiremos el vector de *features* que va a “describir” al fichero de audio: Cada fichero *.wav se “representará” por un único vector de 85 componentes.
4. Para realizar la práctica **es necesario tener instalada la versión 2019b de Matlab o superior**, o bien acceder desde Matlab online subiendo los ficheros de la práctica al Matlab cloud. Recordad que **tenéis que tener instalada la “Audio Toolbox”**.

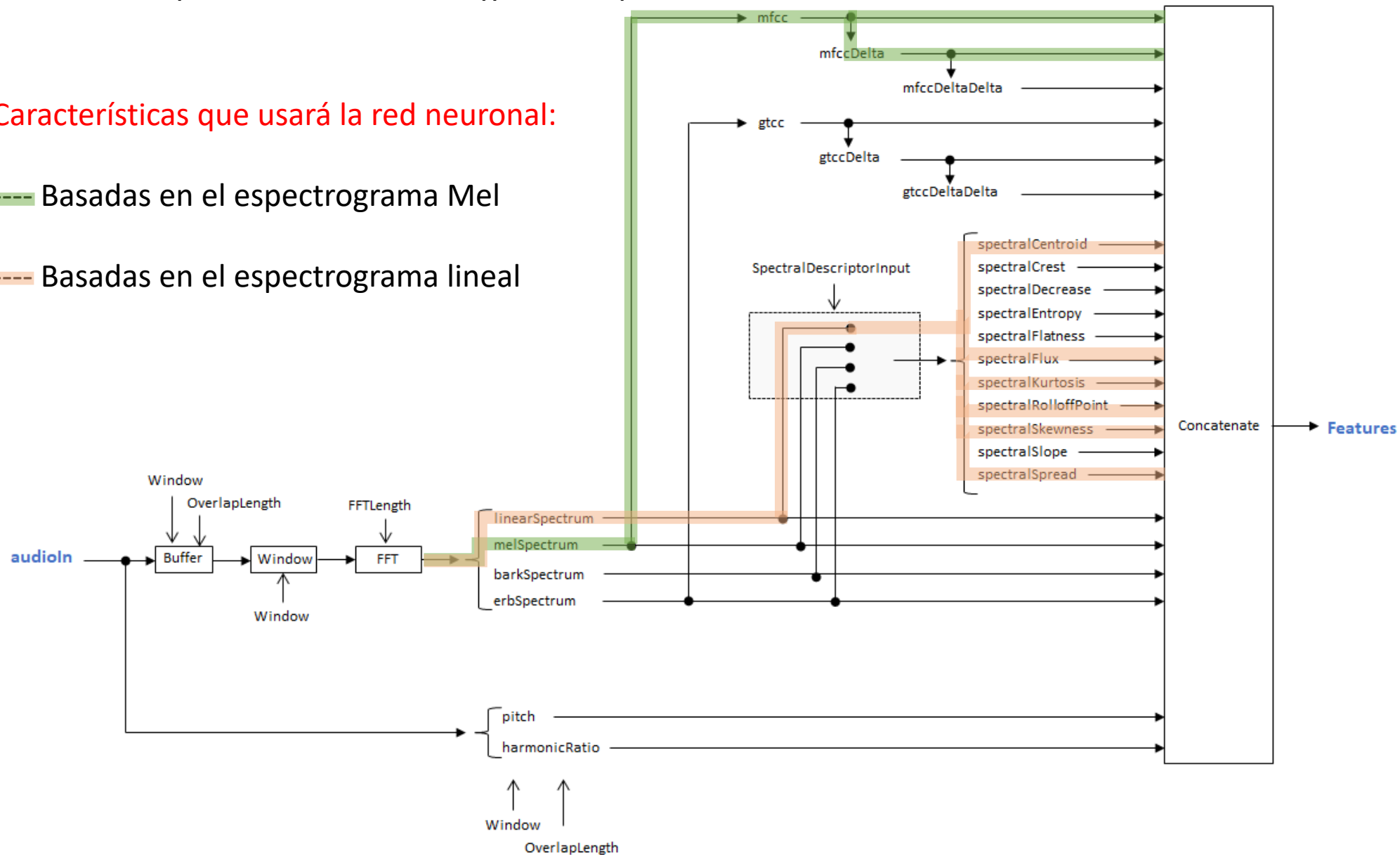
Importante:

- Se proporciona el fichero livescript [mi practica4.mlx](#) para realizar la práctica.
- **Este fichero se debe entregar antes de la siguiente sesión junto con su impresión digital en pdf** (se puede hacer desde Matlab). Se valorará que el fichero tenga comentarios explicativos. También hay que subir el fichero **Extrae_Caracteristicas.m** del ejercicio 4.3.

(en el menú izquierdo haz click en "Algorithms")

----- Basadas en el espectrograma Mel

Basadas en el espectrograma lineal



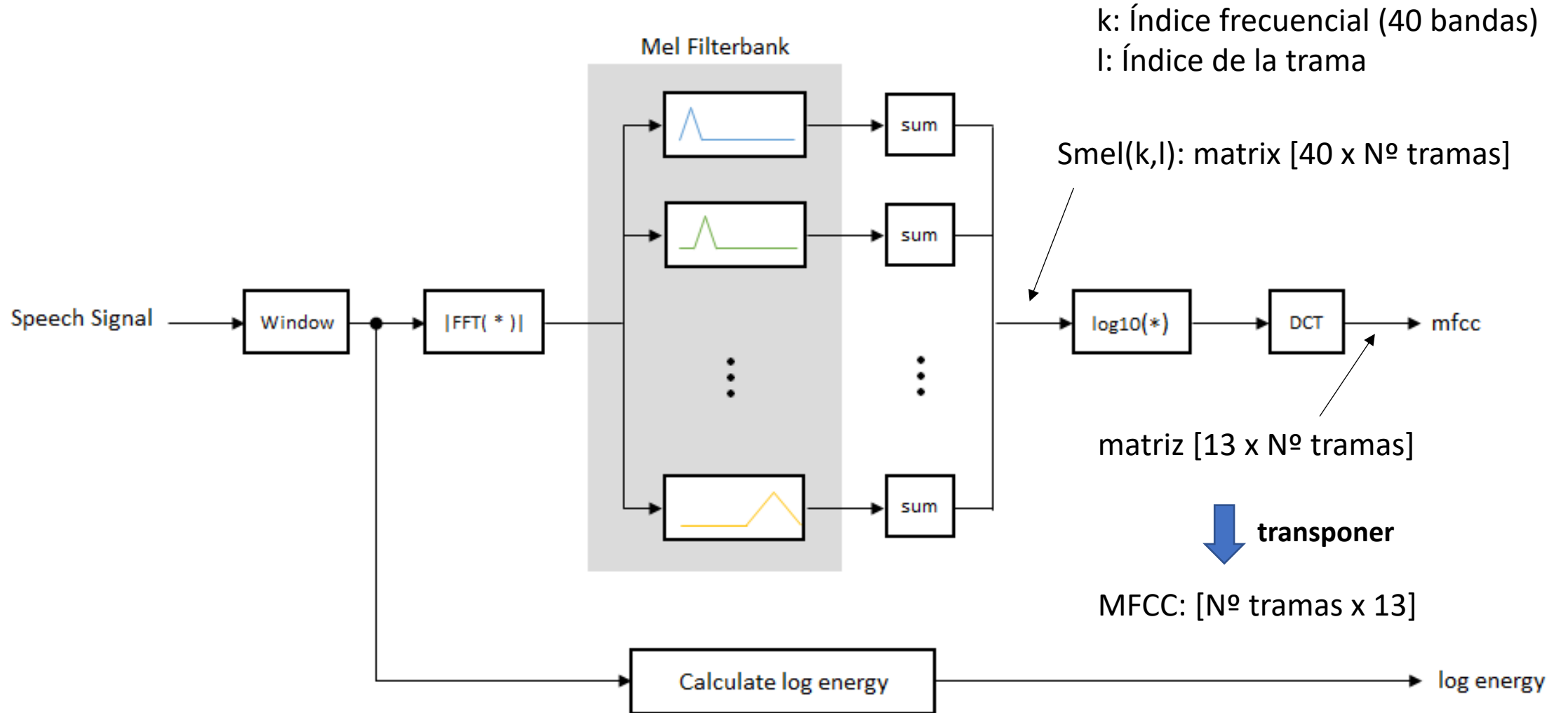
Ejercicio 4.2: Calcular los MFCC

- **MFCC: *Mel Frequency Cepstral Coefficients***
- Procedimiento:

A partir de la matriz “Smel” que contiene el espectrograma de Mel (en la siguiente figura), es necesario realizar para cada trama (columna) de “Smel”:

1. Calcular el logaritmo en base 10
2. Calcular la transformada de coseno discreto (*Discrete Cosine Transform*, DCT) mediante la función `dct` de Matlab. La función `dct` aplicada a una matriz realiza la DCT por columnas, al igual que sucedía con la función `fft`.
3. Los coeficientes MFCC son los coeficientes 1 al 13 de la matriz obtenida a la salida de la transformada DCT.

Cálculo de los MFCC



Ejercicio 4.3: *Spectral Flux* y *Spectral Centroid*

- En este ejercicio calcularemos dos características (*features*) espectrales
- *Spectral Flux*: es una medida de la rapidez con la que cambia el espectro en potencia de una trama a otra. Se calcula mediante la siguiente fórmula donde $S[k, l]$ es el valor **del espectro en potencia** para el índice k de la FFT y la trama l -ésima:

$$\text{flux}(l) = \sqrt{\sum_{k=0}^{N_{fft}/2} |S[k, l] - S[k, l-1]|^2} \quad \text{con} \quad \text{flux}(0) = 0$$



- *Spectral Centroid*: se calcula como el centro de masas del espectro para cada trama, donde f_k es la frecuencia en Hz correspondiente al índice k de la FFT. Perceptualmente, este valor se relaciona con el brillo del sonido:

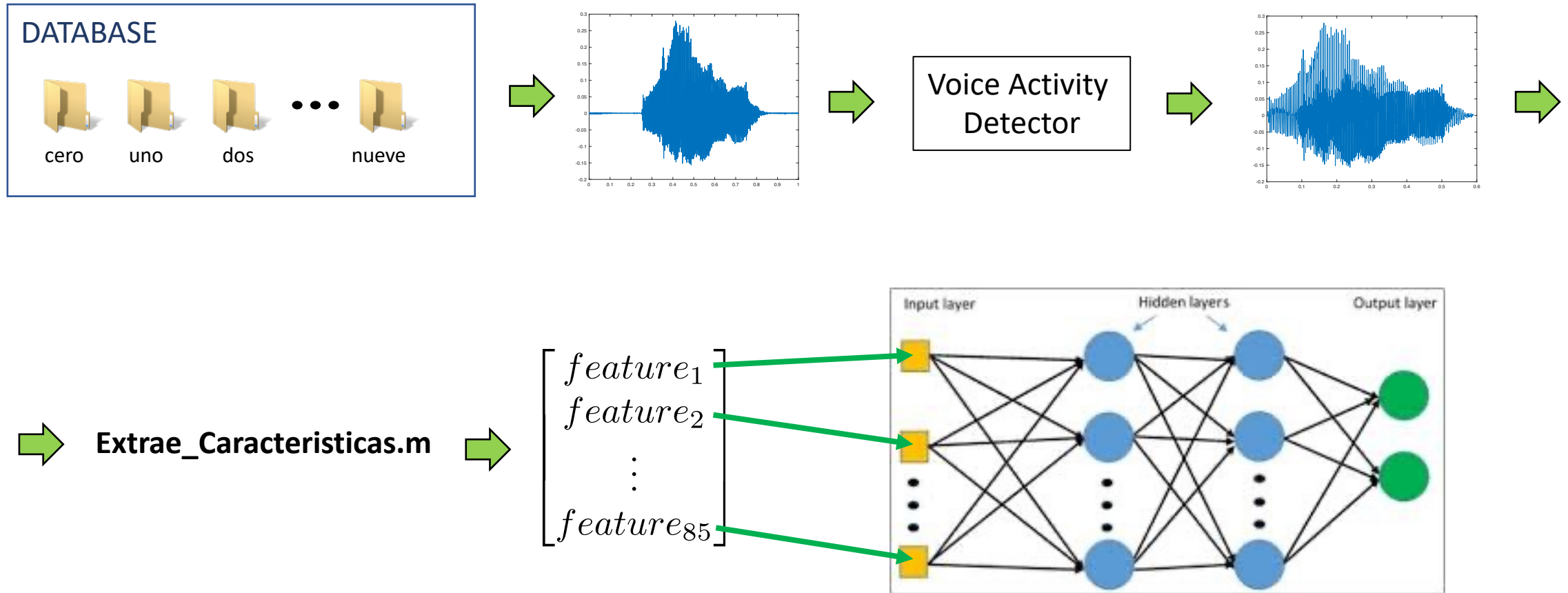
$$\text{centroid}(l) = \frac{\sum_{k=0}^{N_{fft}/2} f_k S[k, l]}{\sum_{k=0}^{N_{fft}/2} S[k, l]}$$



Ejercicio 4.4: Función **Extrae_Caracteristicas.m**

- En este ejercicio se extraerán todas las características necesarias para alimentar la red neuronal.
- Se utilizará el objeto de Matlab **audioFeatureExtractor**
- Se calcularán los promedios temporales de todas las características, y para algunas de ellas además calcularemos la varianza en el tiempo y sus valores máximo y mínimo.
- Con todo ello se construirá la función **Extrae_Caracteristicas.m** que obtendrá al final un único vector de 85 components que describirá cada fichero de audio de la base de datos. Los pasos a seguir están explicados en el livescript.
- Una vez llegados a este punto...

¡Lo tenemos todo para entrenar la red!



Práctica 5 (2 sesiones)

IMPORTANTE

- **Recuerda que antes del 14 de noviembre**, debías haber realizado **20 grabaciones (a 4-5 personas distintas)** de cada dígito y haberlos subido a la carpeta “Carpeta compartida” de la asignatura en PoliformaT.
- Esta tarea vale 1 punto, al igual que cada una de las memorias de prácticas.