Prueba bonus de laboratorio

1. Descarga de datos

Todos los estudiantes trabajarán sonre el mismo fichero de datos llamado `data.npz`, que podrán descargarse de Aula Global.

Podrá cargar el contenido de los ficheros .npz mediante los siguientes comandos:

```
>>> import numpy as np
>>> data = np.load('data.npz')
>>> variable1 = data['variable1']
>>> variable2 = data['variable2']
>>> ...
```

2. Decisión

Se le ha proporcionado un cierto conjunto de datos con K observaciones unidimensionales etiquetadas $\{(x^{(k)}, h^{(k)}), k = 0, \ldots, K-1\}$, siendo $h^{(k)} \in \{0, 1\}$ la hipótesis a la que pertenece la observación $x^{(k)}$. Todas estas observaciones se han almacenado en los array unidimensionales `xD` y `hD`.

(8 %) 1. Estime las probabilidades a priori de ambas hipótesis. Guarde los resultados en las variables `PH0` y `PH1`.

(8 %) 2. Suponga que todos los datos se han generado de acuerdo con verosimilitudes

$$p_{X|H}(x|0) = \frac{1}{B-A}, \qquad A \le x \le B$$

$$p_{X|H}(x|1) = \frac{2(x-A)}{(B-A)^2}, \qquad A \le x \le B$$

Estime los parámetros A y B como los valores mínimo y máximo de los datos disponibles, respectivamente, es decir,

$$\hat{A} = \min\{x^{(0)}, \ldots x^{(K-1)}\},$$
$$\hat{B} = \max\{x^{(0)}, \ldots x^{(K-1)}\}$$

y guarde el resultado en las variables `Ahat` y `Bhat`.

(8 %) 3. Utilizando los valores estimados para los parámetros de las funciones de verosimilitud, así como los valores estimados de las probabilidades a priori, implemente el decisor de máximo a posteriori (MAP), y aplíquelo a los datos proporcionados. Guarde en el array `DMAP` las decisiones del clasificador, de modo que la componente k-ésima contenga la decisión correspondiente a la entrada $x^{(k)}$.

(8 %) 4. Determine el coste promedio del decisor, promediando los costes asociados a cada decisión, de acuerdo con los valores $c_{00} = -1$, $c_{11} = 0$, $c_{01} = 3$, $c_{10} = 2$. Guarde resultado en la variable `CM0`.

3. Estimación

Se le ha proporcionado un cierto conjunto de datos con K observaciones uni-dimensionales $\{(x^{(k)}), k = 0, \ldots, K-1\}$, generadas de acuerdo con la distribución

$$p_{X|S}(x|s) = s \exp(-sx), \qquad x > 0$$

siendo s un parámetro desconocido, generado de acuerdo con una distribución a priori

$$p_S(s) = \exp(-s), \qquad s > 0$$

Las observaciones se encuentran en el array `xR`.

(8 %) 5. Determine el estimador de máximo a posteriori de s a la vista de los datos. Guarde el resultado en la variable `sMAP`

(8 %) 6. Determine la log-verosimilitud $\log(p_{\mathbf{X}|S}(x^{(0)}, \ldots, x^{(K-1)}|s))$ para s igual al estimador obtenido en el apartado anterior (tomando logaritmos neperianos). Guarde el resultado en la variable `Lxs`

(8 %) 7. Suponga que las observaciones $x^{(k)}$ se ha generado con dos valores de s diferentes:

- Un valor s_0 para las observaciones $x^{(k)}$ con k par.
- Un valor s_1 para las observaciones $x^{(k)}$ con k impar.

Determine el estimador ML de s_0 y s_1 dados los datos. Guarde el resultado en las variables `s0ML`, `s1ML`.

4. Filtrado

Cierto registro de señal $x[n]$ ha sido generado filtrando otro registro $u[n]$ con un filtro lineal e invariante de respuesta al impulso finita (FIR), $s[n]$, con `M` coeficientes (que pueden almacenarse en un vector **s** de longitud `M`), añadiendo un ruido `r[n]` aditivo, blanco, Gaussiano e independiente de la señal, con media cero y varianza `varN` a la salida del filtro.

Se sabe que **s** tiene una distribución a priori Gaussiana de media **0** y matriz de varianzas dada por la matriz identidad multiplicada por `varS`.

Los vectores `u` (de longitud N_u) y `x` (de longitud N) almacenan, las primeras muestras de $u[n]$ y $x[n]$, respectivamente, para $n = 0, 1, 2, \ldots$ (hasta $N_u - 1$ o $N - 1$). La longitud de estos registros puede variar, pero en todo caso podrá comprobar que $N_u \geq N$. Puede suponer que $u[n] = 0$ para $n < 0$.

(8 %) 8. Determine el estimador MAP de **s** a la vista de `x` y guárdelo en la variable `sMAP`.

(8 %) 9. Determine la predicción MMSE de $x[N+10], x[N+11], \ldots, x[N+49]$ basada en $x[0], \ldots, x[N-1]$. Guarde el resultado en la variable `xMSE10`.

(8 %) 10. Determine el error cuadrático total dado por

$$SSE = \sum_{n=0}^{N-1} (x[n] - y[n])^2$$

siendo $y[n]$ la salida del último filtro obtenido en los apartados anteriores, para entrada $u[n]$. Guarde el resultado en la variable `SSE`

5. Almacenamiento de resultados

Guarde las variables solicitadas en el fichero `results.npz`, mediante un comando de la forma

```
>>> np.savez('results.npz', var1=var1, var2=var2, ...)
```

Cree el fichero comprimido `labtest.zip` en el que debe incluir su fichero de resultados y el código utilizado para la resolución de la práctica. Suba el fichero .zip a Aula Global antes de que expire el período de tiempo que tiene asignado.