

# Procesamiento Digital de Señales (DSP)

Una de las ramas mas importantes de la electronica y sistemas de control es el procesamiento Digital de señales. Son muchísimas las posibilades de aplicacón al rprocesar señales de distintas fuentes: Sonido, transductores, señales eléctricas o electromagnéticas puras entre muchas otras.



Los manatís no se pueden ver en ríos de aguas turbias, por lo que se identifican por sus sonidos. Al igual que las personas, los manatís emiten sus sonidos con ligeras variaciones, aunque cada uno tiene una "voz" única. Por esto, contar cuántos manatís hay en un río se puede realizar usando técnicas de clustering y DSP, procesando los sonidos que se captan, pensando en que cada cluster corresponde a un manatí en particular.

En este caso se cuenta con una base de datos de muestras desonidos de manatíes captados en un río.

```
In [2]: # Imports
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

%matplotlib inline
```

## Datos a procesar

Se lee el archivo csv con el dataset

```
In [3]: manatíes = pd.read_csv('../datasets/dataset_manatíes.csv',engine='python')
manatíes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 4 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Frecuencia Fundamental (Khz)          150 non-null    float64
1   Longitud de sonido (s)                 150 non-null    float64
2   Intensidad promedio (W/m2)            150 non-null    float64
3   Frecuencia promedio por hora(veces)   150 non-null    float64
dtypes: float64(4)
memory usage: 4.8 KB
```

```
In [4]: # Observar datos de forma rápida
manatíes.head()
```

Out[4]:	Frecuencia Fundamental (Khz)	Longitud de sonido (s)	Intensidad promedio (W/m2)	Frecuencia promedio por hora(veces)
0	7.9	3.8	6.4	2.0
1	7.7	3.8	6.7	2.2
2	7.7	2.6	6.9	2.3
3	7.7	2.8	6.7	2.0

Frecuencia Fundamental (Khz)	Longitud de sonido (s)	Intensidad promedio (W/m2)	Frecuencia promedio por hora(veces)
4	7.7	3.0	6.1
			2.3

Preprocesar los datos

Ahora se tiene los datos en formato leide de csv, pero resulta mas facil trabajar con ellos ya normalizados, y en formato de dataset. Como se puede ver, se tienen 4 variables para cada muestra. Se obtienen tambien las metricas estadisticas para cada columna y las mas significativas.

```
In [5]: manaties.describe()
```

	Frecuencia Fundamental (Khz)	Longitud de sonido (s)	Intensidad promedio (W/m2)	Frecuencia promedio por hora(veces)
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
In [6]: # Normalizar los datos
manaties_normalizado = (manaties - manaties.min())/(manaties.max() - manaties.min())
manaties_normalizado.describe()
```

	Frecuencia Fundamental (Khz)	Longitud de sonido (s)	Intensidad promedio (W/m2)	Frecuencia promedio por hora(veces)
count	150.000000	150.000000	150.000000	150.000000
mean	0.428704	0.440556	0.467458	0.458056
std	0.230018	0.181611	0.299203	0.317599
min	0.000000	0.000000	0.000000	0.000000
25%	0.222222	0.333333	0.101695	0.083333
50%	0.416667	0.416667	0.567797	0.500000
75%	0.583333	0.541667	0.694915	0.708333
max	1.000000	1.000000	1.000000	1.000000

Ahora que se tienen los datos normalizados entre 0 y 1, se pueden obtener sus metricas estadisticas en dos o tres variables PCA

```
In [7]: from sklearn.decomposition import PCA

pca = PCA(n_components=3) # tres componentes principales
pca_manaties = pca.fit_transform(manaties_normalizado)
DataFrame_manaties = pd.DataFrame(data= pca_manaties, columns=['Componente_1', 'Componente_2', 'Componente_3'])
```

Actividad propuesta:

- 1. Ahora que se tienen los datos procesados ordenados y normalizados, elija un método de agrupamiento y realice el análisis correspondiente, puede utilizar PCA (2 o 3 elementos) o usar los datos normalizados directamente.
- 2. Obtenga los diferentes clusters y cree un archivo csv con los resultados en la carpeta Results.
- 3. Realice gráficas del procedimiento y ajuste lo mejor posible el método. Justifique su elección.
- 4. Indique sus principales Conclusiones.