

## EJERCICIO:

Sobre el conjunto de datos X-Y de la práctica (sin considerar el número de Euler), determina las 3 características que proporcionan de forma conjunta la mayor separabilidad en los datos y representa las muestras en ese espacio de características. Hacer para los siguientes casos:

- Círculos vs cuadrados vs triángulos
- Círculos vs cuadrados
- Círculos vs triángulos
- Cuadrados vs triángulos
- Círculos-triángulos vs cuadrados

Para cada caso, aplica el siguiente procedimiento implementado a través de la siguiente función:

```
[espacioCcas, JespacioCcas]=funcion_selecciona_vector_ccas_3_dim(XoI,YoI)
```

donde: XoI y YoI, son los datos del problema específico a tratar

### 1. Ranking de características:

- Cuantificación individual de características: para cada característica, determina el grado de separabilidad J que proporciona de forma individual.
- Utiliza los valores de J para ordenar el conjunto de características de mayor a menor importancia de acuerdo a este criterio de evaluación individual de características.

### 2. Pre-selección de características: de las 22 características, selecciona las 9 más relevantes de acuerdo al ranking establecido en el apartado anterior. Este número también se puede pasar como parámetro de entrada a la función a través de la variable numDescriptoresOI.

### 3. Selección final de características: considerando únicamente las 9 características preseleccionadas, encontrar las 3 características que proporcionan la mayor separabilidad de forma conjunta.

**Observación:** Dado un vector de N elementos, `vector`, la función de Matlab `combnk(vector,k)` devuelve todas las combinaciones de los N valores del `vector` tomados en grupos de k.

Para cada apartado, se debe guardar el espacio de característica resultante, así como el conjunto de datos específicos del problema (`XoI-YoI`, `nombresProblemaOI`).

## ANEXO II

# MEDIDA DEL GRADO DE SEPARABILIDAD ENTRE CLASES A PARTIR DE LA ESTIMACIÓN DE MATRICES DE DISPERSIÓN ENTRE Y DENTRO DE LAS CLASES

The within-class scatter matrix ( $S_w$ ) indicates the distribution of sample points around their respective mean vectors and is defined as:

$$S_w = \sum_{i=1}^C S_i \quad (4.2)$$

$$S_i = \sum_{n \in C_i} (X_n - M_i)(X_n - M_i)^T \quad \text{and} \quad M_i = \frac{1}{N_i} \sum_{n \in C_i} X_n \quad (4.3)$$

where  $C$  is the number of classes,  $N_i$  represents the number of examples in class  $C_i$ ,  $X_n$  refers to the sample  $n$ , and  $M_i$  is the mean of class  $C_i$ . The between-class scatter matrix ( $S_b$ ) represents the scatter of samples around the mean vector of the class mixture and is defined as:

$$S_b = \sum_{i=1}^C N_i (M_i - M)(M_i - M)^T \quad (4.4)$$

$$M = \frac{1}{N} \sum_{n=1}^N X_n = \frac{1}{N} \sum_{i=1}^C N_i M_i \quad (4.5)$$

while  $N = \sum_i N_i$  shows the total number of sample points in the dataset. After within-class and between-class matrices are measured the following metric  $J$  can be obtained:

$$J = \text{trace} \left( \frac{S_b}{S_w} \right) \quad (4.6)$$

## Implementación con Matlab:

Grado de separabilidad de un conjunto de muestras de distintas clases descritas por un vector de atributos (datos dados por X-Y) mediante CSM - ("Class Scatter Matrix"):

```
J = indiceJ (inputs, outputs)
```

donde:

- inputs = X' y outputs = Y'
- J: grado de separabilidad.