

Inteligencia Artificial 2026

Lab 01

02.febrero.2026

En este laboratorio vamos a implementar varios métodos de clustering y aplicarlos a conjuntos de datos.

1. Programar desde cero (sin usar librerías de datos ni funciones especializadas), un método de agrupamiento k -means, para vectores en \mathbb{R}^d .

Como input se deberá dar a la función la matriz de datos, una matriz numérica de $n \times d$, así como el número de grupos o clústers a construir. Como output, su algoritmo debe producir:

- un vector con las clases o *labels* de cada dato (vector de tamaño n).
- un vector con los centroides de cada grupo (matriz de $k \times d$, donde cada fila de esta matriz es un centroide).

2. Evaluar el funcionamiento de su algoritmo con los siguientes conjuntos de datos:

- a) El conjunto de datos **Iris**:

```
from sklearn.datasets import load_iris
import pandas as pd
iris = load_iris()
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['target'] = iris.target
```
- b) El conjunto de datos **penguins** de la librería `seaborn` import `seaborn as sns` `penguins = sns.load_dataset("penguins")`
- c) El conjunto de datos **winequality-red**, dentro del conjunto de datos `wines`:
`https://archive.ics.uci.edu/ml/datasets/Wine+Quality`

En cada caso, elegir un número apropiado de clústers para el agrupamiento. Contrastar los resultados de su algoritmo de k -means contra los de la librería `scikit-learn`, y discutir las semejanzas o diferencias.

3. Hacer un agrupamiento jerárquico con los datos de los países `countries_binary.xlsx`.

Visualizar los resultados o dendrogramas de diferentes métodos de agrupamiento, variando los siguientes:

- el método de agrupamiento (simple, completo, promedio, Ward)
- la métrica utilizada (euclidean, Hamming).

4. Realizar un análisis de agrupamiento k -means, nuevamente para los datos de los países, que están disponibles en el archivo `countries_binary.xlsx`.

Contrastar con el ejercicio anterior. ¿Son iguales las agrupaciones? ¿Por qué? Justificar.

5. Generar un conjunto de datos sintéticos, de tamaño 200, con `make_moons` o círculos concéntricos de Scikit-learn. (Aquí la idea es realizar dos clúster, con 100 observaciones cada una).

Una vez tenga fijado el conjunto de datos, comparar los algoritmos de clustering.

- (a) Explicar lo siguiente: ¿Por qué k -means falla con formas de "luna" o anillos concéntricos?
- (b) Comparar k -means vs. agrupamiento jerárquico. ¿Cuál produce mejores resultados? ¿Por qué? ¿Cuál es el método y la métrica que funcionan mejor en el agrupamiento jerárquico?

6. Realizar un algoritmo de cuantización de colores para imágenes RGB, usando como base un algoritmo de agrupamiento.

Ilustrar los resultados obtenidos con 3 imágenes de su elección. En cada una mostrar:

- El mapa de clases resultado del agrupamiento.
- La imagen cuantizada resultante.