

## PRÁCTICA 03

### ALGORITMOS DE AGRUPAMIENTO

Docente: Dr. Edward Hinojosa Cárdenas

11 de Noviembre del 2023

#### 1 CONCEPTOS BÁSICOS

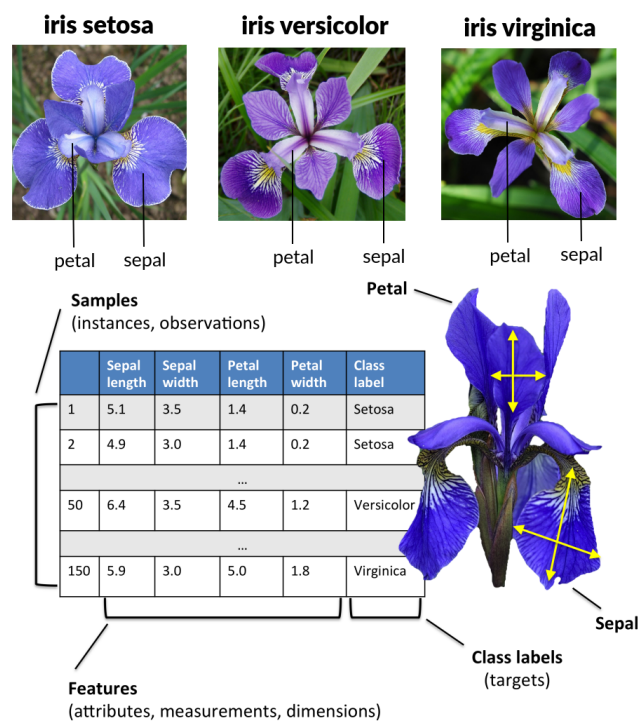
- Agrupamiento no Supervisado
- K-means

#### 2 EQUIPOS Y MATERIALES

- Un computador.
- Material del curso.

#### 3 EJERCICIOS

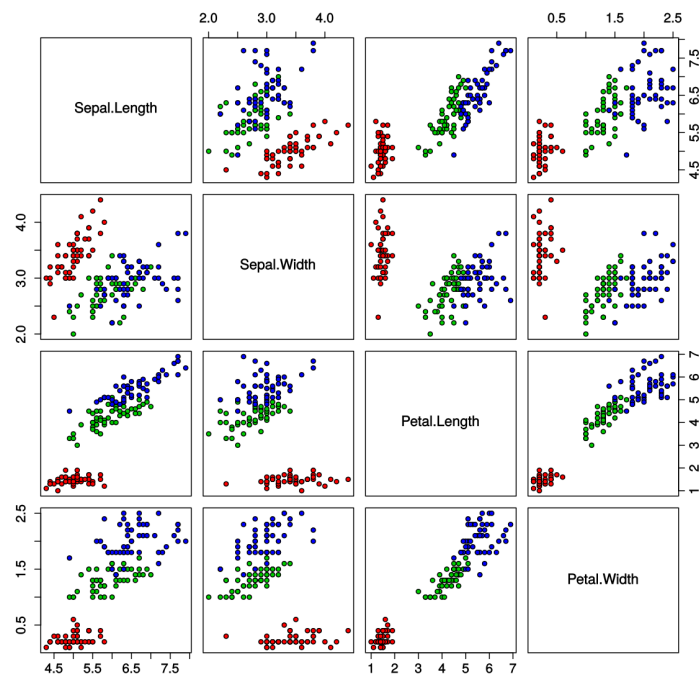
1. Considerar la base de datos iris (como problema de agrupamiento o clustering):





	A	B	C	D	E
1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Class
2	5.1	3.5	1.4	0.2	Iris-setosa
3	4.9	3	1.4	0.2	Iris-setosa
4	4.7	3.2	1.3	0.2	Iris-setosa
5	4.6	3.1	1.5	0.2	Iris-setosa
6	5	3.6	1.4	0.2	Iris-setosa
7	5.4	3.9	1.7	0.4	Iris-setosa
8	4.6	3.4	1.4	0.3	Iris-setosa
9	5	3.4	1.5	0.2	Iris-setosa
10	4.4	2.9	1.4	0.2	Iris-setosa
11	4.9	3.1	1.5	0.1	Iris-setosa
12	5.4	3.7	1.5	0.2	Iris-setosa
13	4.8	3.4	1.6	0.2	Iris-setosa
14	4.8	3	1.4	0.1	Iris-setosa
15	4.3	3	1.1	0.1	Iris-setosa
16	5.8	4	1.2	0.2	Iris-setosa
17	5.7	4.4	1.5	0.4	Iris-setosa
18	5.4	3.9	1.3	0.4	Iris-setosa
19	5.1	3.5	1.4	0.3	Iris-setosa
20	5.7	3.8	1.7	0.3	Iris-setosa
21	5.1	3.8	1.5	0.3	Iris-setosa
22	5.4	3.4	1.7	0.2	Iris-setosa
23	5.1	3.7	1.5	0.4	Iris-setosa
24	4.6	3.6	1	0.2	Iris-setosa
25	5.1	3.3	1.7	0.5	Iris-setosa

**Iris Data (red=setosa, green=versicolor, blue=virginica)**



- Considere el los archivos iris-5-fold.zip y iris-10-fold.zip ([Link](#)) que contienen los conjuntos de datos de entrenamiento y los datos de test para realizar una validación cruzada de 5 y 10 folds respectivamente.
- Implemente y ejecute los algoritmos K-means, DBSCAN y Fuzzy c-means usando el archivo iris-5-fold.zip y defina los valores de precisión para cada fold. Considere los valores obtenidos para los datos de entrenamiento y test, calcule la media y la desviación estándar. Muestre los resultados obtenidos en tabla. Defina los parámetros para cada algoritmo con el objetivo de mejorar los resultados obtenidos.
- Implemente y ejecute los algoritmos K-means, DBSCAN y Fuzzy c-means usando el archivo iris-5-fold.zip y defina los valores de precisión para cada fold. Considere los valores obtenidos para los datos de entrenamiento y test, calcule la media y la desviación estándar. Muestre los resultados obtenidos en tabla. Defina los parámetros para cada algoritmo con el objetivo de mejorar los resultados obtenidos.
- Recuerde que para la generación del modelo no se considera la clase de cada ejemplo por ser un algoritmo no supervisado.
- Genere un artículo con la comparación de los algoritmos implementados que por lo menos contenga: Título, Abstract, Introducción, Algoritmos implementados, Base de Datos utilizada, comparación (discusión de resultados), conclusiones, trabajos futuros y referencias. El artículo considerar dos columnas. El artículo debe tener por lo menos 3 páginas completas y como máximo 10 páginas completas. Considere por lo menos 6 referencias. Se utilizará software de similitud el cual no debe superar el 30% de similitud, caso contrario la nota será 0 puntos.
- **IMPORTANTE:** No utilice ningún framework o librería del algoritmo K-means, DBSCAN o Fuzzy c-means. Debe utilizar sus implementaciones. Pueden considerar implementar adicionalmente otros algoritmos y utilizar otras bases de datos.

#### 4 ENTREGABLES

Al finalizar el estudiante deberá:

1. Generar el artículo en fomato .pdf (Ejemplo: [Link](#), Formato IEEE: [Link](#)).
  2. Comprimir en un archivo .zip todos los archivos anteriores (además de todo código fuente, sin el cual no se revisará el laboratorio, separado por algoritmo) y subirlo el archivo al aula virtual según las indicaciones del profesor:  
Practica\_XX\_ApellidoPaterno\_ApellidoMaterno\_PrimerNombre\_IA\_MCC\_2023\_UNSA.zip
- **IMPORTANTE** En caso de copia o plagio o similares todos los alumnos implicados tendrán sanción en toda la evaluación del curso.