

Reconocimiento de patrones y aprendizaje automático

Tarea 3: Clustering

Fecha de entrega: Lunes 31 de enero de 2022

Profesor: Andrés Aldana Gonzáles
Ayudante: Felipe Navarrete Córdova

1. Introducción

El archivo **iris.csv** contiene información sobre las mediciones de 150 lirios: 3 clases diferentes con 50 instancias cada una. Cada columna del archivo representa la siguiente información:

- Longitud del sépalo (cm.)
- Ancho del sépalo (cm.)
- Longitud del pétalo (cm.)
- Ancho del pétalo (cm.)
- Tipo de lirio

2. Ejercicios

1. Utiliza los algoritmos *Hierarchical clustering*, *K-Means* y *DBSCAN* para obtener los grupos (*clusters*) de las diferentes clases de lirios.
2. Utiliza al menos tres formas de evaluar los resultados de los algoritmos utilizados (Disimilitud, pureza, entropía, inercia, Silhouette).
3. Recuerda que K-Means es sensible a las posiciones iniciales de los centroides. Realiza varias ejecuciones con posiciones iniciales diferentes.
4. Recuerda que DBSCAN es sensible al radio (ϵ) y al número de elementos en la vecindad (**MinPts**). Prueba diferentes valores de estos parámetros hasta que estés conforme con el agrupamiento realizado.
5. ¿Cuál es el número óptimo de clusters para este conjunto de datos? Justifica tu respuesta.
6. Compara los resultados entre los diferentes algoritmos de agrupamiento, ¿Cuál es el mejor algoritmo para clasificar este conjunto de datos?

7. Interpreta los resultados:

- ¿Cuáles son las características comunes de cada grupo? ¿Cuáles son sus diferencias?
- ¿Los grupos coinciden con los tipos de lirios? De no ser así, ¿hay más o menos grupos que tipos de lirios? ¿Cuál es el significado de esta diferencia?

3. Software recomendado

Por la simplicidad, facilidad de uso, y compatibilidad se recomienda utilizar *python* como lenguaje de programación y las siguientes bibliotecas:

- numpy
- pandas
- scikit-learn

4. Entregables

La tarea se debe entregar en un notebook de Jupyter con los resultados y las gráficas correctamente discutidos. Los archivos de datos deben estar en el mismo nivel de directorio del notebook para facilitar la ejecución de los programas.

5. Bibliografía recomendada

- Aurelien Geron - Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow-O'reilly (2019). Chapter 9 - Unsupervised Learning Techniques
- Tom Mitchell - Machine Learning. Chapter 8 - Instance Based Learning