

Aprendizaje no supervisado y detección de anomalías

Práctica: Técnicas de agrupación (*clustering*)

Curso 2022-2023

Descripción

En este trabajo se propone la resolución de un problema de agrupación sobre un conjunto de datos seleccionado por el estudiante. Se puede utilizar cualquier de los conjuntos de datos del UCI Machine Learning Repository no empleado en clase. En caso de duda de si el conjunto de datos es adecuado, se puede contactar con el profesor para confirmarlo (triguero@decsai.ugr.es).

El objetivo del trabajo es poner en práctica los conocimientos adquiridos en el bloque de *clustering* de la asignatura, demostrando que se conoce la metodología y las herramientas para la creación de modelos de agrupación y que se es capaz de describir e interpretar los resultados obtenidos.

El trabajo se desarrollará individualmente. Los estudiantes deberán seleccionar previamente que se desea realizar este trabajo en PRADO.

Criterios de evaluación

Para la evaluación, se entregará un cuaderno interactivo de Python que combina tanto explicaciones como código necesario para reproducir los experimentos.

Se evaluará la calidad del cuaderno presentado y la elaboración de las soluciones propuestas. Se valorará especialmente la claridad en la redacción y el apoyo con materiales visuales: diagramas, gráficas, etc. Más concretamente, se tendrán en cuenta los siguientes aspectos:

- Definición del problema: datos, objetivos, hipótesis, resultados esperados
- Descripción de la metodología empleada para abordar el problema
- Variedad y comparativa entre los algoritmos utilizados
- Utilización de medidas de calidad del *clustering*
- Aplicación de metodologías para ajuste de hiperparámetros
- Explicación de los resultados obtenidos

Como recomendación general, es más relevante argumentar la utilización y los resultados de los algoritmos que aplicar exhaustivamente todos los presentados en clase.

La entrega se realizará a través de PRADO, en el enlace habilitado al efecto en la página de la asignatura.

Entrega

Límite: sábado 14 de enero de 2023, 23:59

Contenidos: Un cuaderno Jupyter (.ipynb), incluyendo:

- Introducción general y objetivos

- Descripción del conjunto de datos (incluyendo estudio exploratorio de datos)
- Metodología
- Experimentación
- Discusión de resultados
 - * (puede hacerse una sola sección Experimentación + Discusión)
- Conclusiones
- Bibliografía