

ELE050-B: Aprendizaje de Máquina

Tarea 2: Aprendizaje No supervisado

Daniel San Martín

1. Objetivo

El objetivo de esta tarea es aplicar los contenidos revisados en la asignatura *Aprendizaje de Máquina* para problemas de aprendizaje **no supervisado**. Específicamente, abordaremos las tareas de **reducción de dimensionalidad** y **agrupamiento** (*clustering*).

2. Problemas

La tarea es descubrir estructuras y patrones ocultos en los problemas indicados a continuación, mediante la visualización en una dimensión menor y aplicando métodos de agrupamiento. Para ambos conjuntos de datos, considere la etiqueta como una columna adicional en el procesamiento.

2.1. Semillas

Este conjunto fue creado como parte de un estudio agronómico para analizar las características físicas de semillas de trigo de tres variedades distintas: *Kama*, *Rosa* y *Canadiense*.

Cada muestra en el conjunto de datos representa una semilla y está descrita mediante 7 atributos numéricos calculados a partir de imágenes obtenidas mediante análisis morfológico. Estos atributos incluyen, por ejemplo, el área, perímetro, compacidad, longitud y ancho del núcleo, entre otros [1]. Detalles y fuente de datos: <https://archive.ics.uci.edu/dataset/236/seeds>.

2.2. Identificación de billetes

Este conjunto de datos fue diseñado para el estudio de técnicas de aprendizaje automático aplicadas a la detección de billetes falsificados. Los datos fueron obtenidos a partir de imágenes escaneadas de billetes genuinos y falsos, utilizando transformadas *wavelet* para extraer características estadísticas.

Cada muestra representa un billete y está descrita por 4 atributos numéricos derivados de las imágenes, como: varianza, asimetría, curtosis y entropía [2]. Detalles y fuente de datos: <https://archive.ics.uci.edu/dataset/267/banknote+authentication>.

3. Instrucciones

- El trabajo debe ser desarrollado y presentado por los equipos definidos en Aula.
- La fecha de presentación es el día **9 de junio del 2025** en el horario de clases.
- La tarea debe ser desarrollada usando *Jupyter Notebook*. Los archivos deben estar debidamente organizados y ser **autocontenidos**, considerando las explicaciones, código, etc. Se sugiere crear un *notebook* por cada problema.
- Todo el código debe ser almacenado y gestionado en un repositorio como *GitHub*, *GitLab* o alguna alternativa similar. El repositorio debe ser privado hasta el día de la presentación y liberado antes de presentar. Se revisará el aporte de cada integrante al repositorio.

- Puede usar contenidos del curso *Ciencia de Datos y Aprendizaje de Máquina*. Si utiliza cualquier otro modelo o técnica, debe documentarla y explicarla en el cuadernillo.
- Se habilitará un buzón en Aula para que ingresen el enlace al repositorio.

4. Rúbrica

Los criterios a evaluar son los siguientes:

- Presentación de resultados en el repositorio: **30 puntos**.
- Claridad en la presentación y defensa de los resultados: **50 puntos**.
- Evidencia de trabajo en equipo tanto en el repositorio como en la presentación: **20 puntos**.

Más detalles en la rúbrica del buzón de Aula.

Referencias

- [1] M. Charytanowicz, J. Niewczas, P. Kulczycki, P. Kowalski, and S. Lukasik, “Seeds.” UCI Machine Learning Repository, 2010. DOI: <https://doi.org/10.24432/C5H30K>.
- [2] V. Lohweg, “Banknote Authentication.” UCI Machine Learning Repository, 2012. DOI: <https://doi.org/10.24432/C55P57>.