

Tercer examen parcial

Sistemas Inteligentes II

Jorge Alberto Jaramillo Garzón
Universidad de Caldas

El repositorio de aprendizaje de máquina de la Universidad de California en Irvine (*UCI machine learning repository*) almacena más de 500 datasets que pueden ser usados en tareas de aprendizaje de máquina (<https://archive.ics.uci.edu/ml/index.php>).

- 1) **(0.9)** Cada estudiante debe ingresar a la página de la UCI y escoger una base de datos con características categóricas o mixtas (categóricas y numéricas) para clasificación. Entrene un modelo CART o C4.5/ID3 para obtener un árbol de decisión y generar reglas explícitas de clasificación. A continuación, analice las reglas obtenidas, identifique cuáles resultan más útiles (por ejemplo, por pureza, cobertura y simplicidad), y realice una validación cruzada del modelo. Concluya al respecto.
- 2) **(0.9)** Cada estudiante debe ingresar a la página de la UCI y escoger una base de datos para **clustering**. A continuación usará el algoritmo de clustering jerárquico con los diferentes criterios de enlazamiento (*linkage*) para agrupar los datos. Finalmente, usando el criterio de agrupamiento J_4 (el mismo que vimos en la clase de selección de características), analizará los diferentes puntos de corte para el agrupamiento. ¿Qué puede concluir?
- 3) **(0.7)** Finalmente, usando el número óptimo de clusters encontrado en el punto anterior, usará el algoritmo k-medias y evaluará nuevamente el algoritmo usando el criterio J_4 . Concluir al respecto.
- 4) **(2.5)** Sustentación oral en horario de clase.

Entregable: Un notebook de Jupyter con código, salidas y comentarios personales. Puede emplear las librerías que considere pertinentes para la implementación de los modelos de aprendizaje de máquina.