Aprendizaje de Máquina

Parcial 1(Máximo 2 Estudiantes).

**Parte 1(25 puntos)**

Utilizando el siguiente Dataset:

https://raw.githubusercontent.com/prashant-kikani/breast-cancer-detection/master/breast-cancer-data.csv

Tomar las 30 primeras muestras del dataset de diabetes y utilizar como variables de entrada

| **radius\_mean** | **texture\_mean** | **perimeter\_mean** |
| --- | --- | --- |

Variable a predecir:

**diagnosis**

Realizar el cálculo manual para todo el proceso(sin usar sklearn, aunque puede escribir su propio código de apoyo o una hoja de cálculo)

a) Realizar validación cruzada de cv=3 para un árbol de decisión clasificador de max\_depth=3:

- En cada CV encontrar un árbol de decisión de max\_depth=3 utilizando entropía. Mostrar cada árbol resultante, evaluar Accuracy en train y Accuracy en validación.

**En cada proceso de entrenamiento se debe mostrar la entropía para todos los posibles condicionales y cuál se escogió.(12 puntos)**

b) Corroborar los resultados usando Sklearn (2 puntos)

c) Dibujar la frontera de decisión a mano o en el notebook para el árbol de la primera cv.(No se pueden usar bibliotecas que realizan automáticamente este proceso).(8 puntos)

d) Corroborar los resultados con Sklearn.**(2 puntos)**

**Parte 2(25 puntos)**

Utilizando el dataset de Grades, construir un modelo de árbol de decisión con sklearn que prediga la nota final(G3) de un estudiante. Utilizar sólo las variables numéricas(seleccionar las que considere más importantes). No utilizar las columnas G1 y G2. Utilizar validación cruzada de 5. Probar con al menos 3 alturas diferentes.(Expresar resultados en términos de promedios y desviación estándar).

https://raw.githubusercontent.com/camilousa/datasets/master/school\_grades\_dataset.csv