Gente,

Para aquellos que quieran seguir practicando un poco mas les dejo el siguiente dataset adjunto.

Encontraran la explicación en:

<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

Es un problema de clasificación binaria (las categorías son 0 o 1). La ultima columna del dataset indica si un paciente en particular tiene o no diabetes. El resto de las columnas son datos de la historia clínica del paciente (ver link arriba).

Inicialmente trabajaríamos de manera similar al dataset de Messi visto en clases. Usen como guía lo visto en la AVS 2 y el material subido al foro.

**1.**Elijan 2 o mas features (columnas del CSV) e intenten hacer una predicción sobre el resultado del paciente utilizando regresión logística.

**2.**Graficar los resultados del modelo.

**3.** Utilizar K-folds y sacar un promedio del error.

**4.**Luego, para aquellos que se animen, pueden probar usar KNN con K=2 o Arboles de Decisión. Esto lo veremos mas en detalle recién en la AVS 3.

Estamos en contacto por cualquier duda!

Un abrazo! Enrique