# Aplicación de regularización

En este laboratorio se hizo el uso de los datasets de regresión y clasificación que se utilizó en los anteriores laboratorios, se repitió los mismo utilizando regularización y sin utilizar regularización, y se presentó los resultados de ambas experiencias, identificando cuales son los mejores hiperparametros en ambos casos, si corresponde.

## Enlace a los datasets:

### Regresión Lineal

Se hizo uso del dataset para predecir el precio de las casas en USA. El enlace al dataset es <https://www.kaggle.com/datasets/harlfoxem/housesalesprediction>.

El conjunto de datos, denominado **kc\_house\_data.csv**, contiene un conjunto de entrenamiento de precios de casas en el condado de King. Este conjunto de datos probablemente incluye varias características relevantes para predecir el precio de las casas, como el tamaño de la vivienda, el número de habitaciones, la ubicación, las características específicas de la casa, etc.

### Regresión Logística

Se hizo uso del para entrenar el modelo de **Regresión Logística** aplicado regularización y predecir si un paciente tiene alguna enfermedad cardiovascular.

Estos problemas a menudo se deben a la aterosclerosis. Esta afección ocurre cuando la grasa y el colesterol se acumulan en las paredes del vaso sanguíneo (arteria). El enlace al dataset es <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>).

El archivo **cardiovascular\_diseases\_dv3.csv** contiene un conjunto de datos de entrenamiento de datos si un paciente tiene una enfermedad cardiovascular.

### Regresión Logística One-vs-All

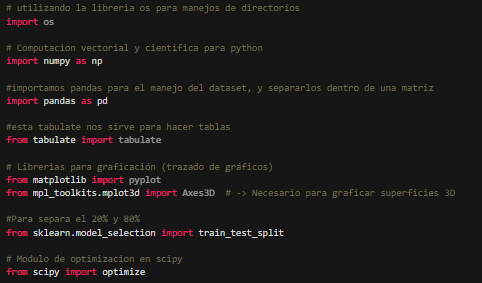
Se hizo uso del para entrenar el modelo de **Regresión Logística one vs all** aplicando regularización y dada la información relacionada con el crédito de una persona, se creó un modelo de aprendizaje automático que pueda clasificar el puntaje crediticio de una persona

Haciendo uso del siguiente dataset: <https://www.kaggle.com/datasets/deepaksaipendyala/credit-score-classification-processed>.

El archivo **credit\_score.csv** contiene un conjunto de datos de entrenamiento para clasificar el puntaje crediticio de una persona.

## Importación de las Librerías necesarias

Primero se importó todas las librerías necesarias, pero la más importante aquí es la de **Pandas** que es una herramienta poderosa y versátil utilizada para manipulación y análisis de datos. Ofrece estructuras de datos flexibles y eficientes para trabajar con datos tabulares, como hojas de cálculo en Excel o tablas SQL:



Ahora se pasará a la aplicación de regularización para cada tipo de regresión:

## Regresión Lineal

### Carga de datos del dataset

cargamos los datos haciendo el uso de la librería **Pandas,** donde se separa todos los datos delimitados por un “,” a una variable **df**:

#cargamos el dataset a un dataframe

df = pd.read\_csv('kc\_house\_data.csv', *delimiter*=',')