

# Guía para diferenciar entre datasets de regresión y clasificación

## ¿Qué es un problema de regresión?

Un problema de regresión ocurre cuando **queremos predecir un valor numérico continuo**. Ejemplos de valores continuos:

- Precio de una casa
- Temperatura en grados
- Porcentaje de popularidad
- Altura de una persona

## Características de un dataset de regresión:

- La variable objetivo (target) es numérica continua
- No hay un número fijo de clases
- Se puede representar en una recta o plano continuo

## ¿Qué es un problema de clasificación?

Un problema de clasificación ocurre cuando queremos predecir **una categoría o clase**. Ejemplos de categorías:

- "Sí" o "No"
- "Apto", "No apto"
- Tipo de flor: "Iris Setosa", "Iris Versicolor", etc.
- Diagnóstico: "Diabético" o "No diabético"

## Características de un dataset de clasificación:

- La variable objetivo (target) es categórica
- Tiene un número limitado de posibles clases
- Cada ejemplo pertenece a una clase (o varias, si es multiclase/multietiqueta)

## ⚠ Cuidado: Números que en realidad son categorías

A veces, la variable objetivo parece ser un número, pero no representa una cantidad real, sino una categoría codificada con números. En esos casos, el problema es de clasificación, no de regresión.

Ejemplos:

Variable Objetivo	Qué significa?	Tipo de problema
0 = No, 1= si	Compró el producto o no?	Clasificación
1, 2, 3	Tipo de flor (1=Bugambilia, 2 = Rosas, 3= Tulipan)	Clasificación
0-4	Nivel de satisfacción ( 0 = muy baja, 4= muy alta)	Clasificación (ordinal)
0, 1, 2	Número de clases en una escuela	Clasificación (si son categorías distintas, no cantidades)

## Cómo darte cuenta si es un número o una categoría:

- ¿Se puede hacer una media de los valores y que tenga sentido?
  - Sí → Puede ser regresión
  - No → Probablemente es clasificación
- ¿Los valores representan algo cualitativo aunque estén en forma de número?
  - Ej: "1 = perro, 2 = gato, 3 = ave"
  - Entonces es clasificación

### Un tip:

Dentro de el repositorio de UCI Machine Learning Repository (si decidiste utilizar alguna base de datos de ahí) puedes confirmar el tipo de problema en la sección de "Associated Tasks".

Por ejemplo, este dataset contiene datos donde se quiere ayudar a un equipo forense a identificar 6 tipos de vidrio que aparecen en la escena del crimen.

Este sería un problema de clasificación ya que nuestra variable objetivo (tipo de vidrio) es una variable categórica.



# Glass Identification

Donated on 8/31/1987

From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc)

## Dataset Characteristics

Multivariate

## Subject Area

Physics and Chemistry

## Associated Tasks

Classification

## Feature Type

Real

## # Instances

214

## # Features

9