



**TOULOUSE**  
**LAUTREC**

# REGRESION LOGÍSTICA Y ARBOLES DE DECISIÓN



# Objetivo de la clase



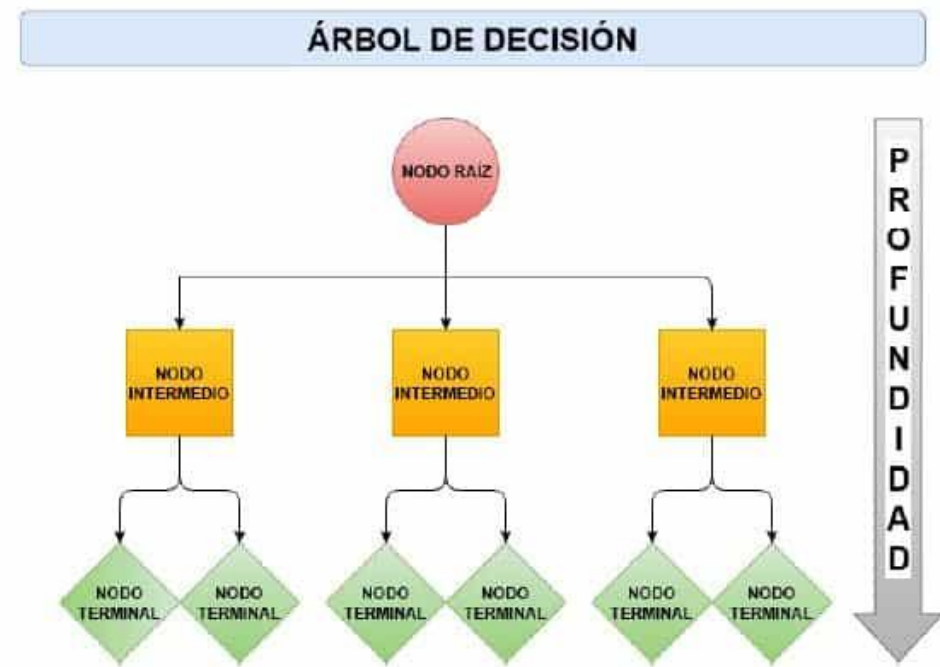
Aplicar algoritmos de clasificación: Entender la parte teórica y de código de los algoritmos de reg. Logística y Árboles de Decisión

Palabras clave: aprendizaje supervisado, regresión lineal, regresión múltiple, regresión no lineal, evaluar modelo

# CLASIFICACION BINARIA Y MULTICLASE

**Clasificación binaria:** Es la tarea de clasificar los elementos de un conjunto en dos grupos sobre la base de una regla de clasificación. Los problemas típicos de clasificación binaria incluyen pruebas médicas para determinar si un paciente tiene cierta enfermedad o no, control de calidad en la industria, y en la recuperación de información, decidir si una página debe estar en el conjunto de resultados de una búsqueda o no

**Clasificación multiclase:** A diferencia de los problemas de clasificación binaria, en la clasificación multiclase se clasifican las muestras en más de dos clases. La respuesta predicha es la clase (por ejemplo, etiqueta) con la puntuación máxima predicha. Las métricas típicas que se utilizan en la multiclase son las mismas que se utilizan en el caso de la clasificación binaria.

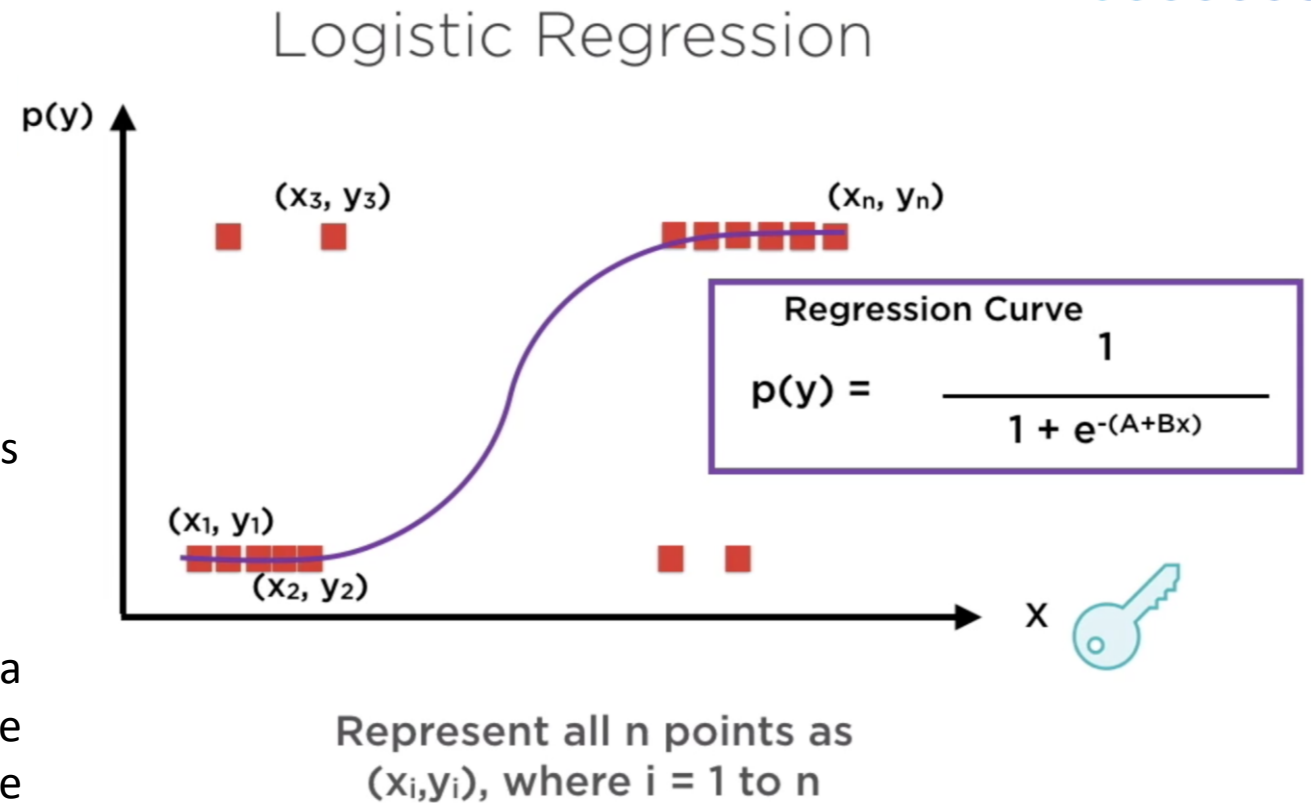


# REGRESIÓN LOGÍSTICA

**Variable dependiente binaria:** En su forma básica, la regresión logística se utiliza para predecir variables dicotómicas (0 ó 1).

- La variable de respuesta es binaria.
- Las observaciones son independientes.
- No existe una multicolinealidad severa entre las variables predictoras.

**Diferencia con la regresión lineal:** A diferencia de la regresión lineal, donde se estima el valor específico de la variable dependiente, en la regresión logística se estima la probabilidad de que se produzca la expresión 1<sup>2</sup>.



# ÁRBOLES DE DECISIÓN

- **Fácil de interpretar:** La lógica booleana y las representaciones visuales de los árboles de decisión los hacen más fáciles de entender y consumir. La naturaleza jerárquica de un árbol de decisión también facilita ver qué atributos son los más importantes, lo que no siempre es claro con otros algoritmos, como las [Redes neuronales](#).

- **Se requiere poca o ninguna preparación de datos:** Los árboles de decisión tienen una serie de características que los hacen más flexibles que otros clasificadores. Puede manejar varios tipos de datos, es decir, valores discretos o continuos, y los valores continuos se pueden convertir en valores categóricos mediante el uso de umbrales.

**Estructura jerárquica:** Tienen una estructura de árbol jerárquica, que consta de un nodo raíz, ramas, nodos internos y nodos hoja<sup>1</sup>.

