Aprendizaje Automático I

1er cuatrimestre 2024

Prof. Lic. Javier Di Salvo

Programa de la materia:

- Introducción al aprendizaje automático.
- Tipos de aprendizaje automático.

- **Técnicas de aprendizaje automático supervisado.**
    - Arboles de clasificación, árboles de regresión, bosques aleatorios de clasificación y de regresión. Clasificador de Bayesiano. Clasificador de regresión: lineal, logística y polinómica. (6-7-8-9)

- **Técnicas de aprendizaje automático no supervisado.**
    - Modelo de Clusteing jerárquico: dendogramas. Modelo de Clusteing no jerárquico: k-means, k-medoid. (10-11-12)

Clase N°1. Introducción al aprendizaje automático

El **aprendizaje automático** es una rama de la inteligencia artificial, la cual basada en algoritmos estadísticos permite que los sistemas informáticos puedan aprender a tomar decisiones basadas en algoritmos de aprendizajes. Es así que el algoritmo por ejemplo aprende a clasificar, agrupar, dividir objetos para poder realizar predicciones. El objetivo es imitar el comportamiento humano optimizando el tiempo de respuesta, la precisión y el costo de un proceso. El sistema para poder mejorar sus predicciones necesita nutrirse de nueva información y de esta forma logra mejorar sus resultados.

Por ejemplo, mientras que un conductor decide de manera intuitiva cual es la velocidad óptima de su auto en la ruta un día de lluvia, el algoritmo de un auto automático en base a información observable como: nivel de precipitaciones, hora y cantidad de tráfico, permite calcular la velocidad del auto apropiada para evitar un accidente. El aprendizaje automático se puede clasificar en dos grandes grupos: el aprendizaje supervisado y el aprendizaje no supervisado. **El aprendizaje supervisado** se caracteriza porque el algoritmo es ingestado por datos etiquetados y aprende de ellos para predecir el comportamiento de la variable respuesta en base a un conjunto de variables observables o factores.

> *Datos etiquetados: identificar los datos sin procesar (es decir, imágenes, archivos de texto, vídeos) y luego añadirles una o más etiquetas para especificar su contexto para los modelos, lo que permite que el modelo de machine learning realice predicciones precisas. Conoce los inputs y los outputs o variables objetivo a predecir.*

En el caso de los sistemas **no supervisados**, el objetivo es identificar patrones de comportamiento en datos ***no etiquetados****,* pudiendo encontrar tendencias no esperadas en el análisis.

> *Datos no etiquetados: No cuenta con variables objetivos a predecir o clasificar. Conoce los inputs pero no los outputs.*

Dentro de estos dos grandes grupos existen diferentes **técnicas de clasificación y regresión, como por ejemplo:** la regresión lineal múltiple, la regresión logística, máquinas de vectores de soporte, *random forest*, árboles de decisión, redes neuronales, entre otros.

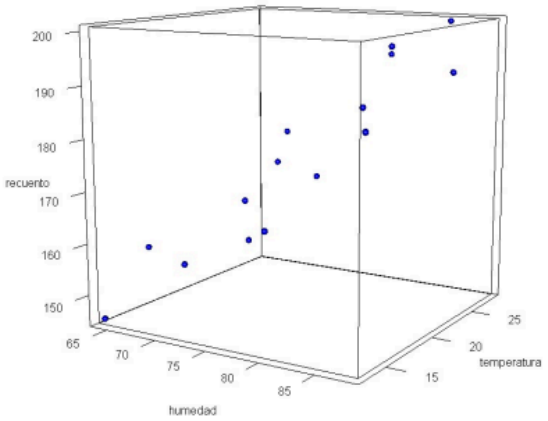
EJEMPLO DE REGRESIÓN LINEAL MÚLTIPLE.

Ejemplo: En un estudio sobre la población de un parásito se hizo un recuento de parásitos en 15 localizaciones con diversas condiciones ambientales.
Los datos obtenidos son los siguientes:

Temperatura	15	16	24	13	21	16	22	18	20	16	28	27	13	22	23
Humedad	70	65	71	64	84	86	72	84	71	75	84	79	80	76	88
Recuento	156	157	177	145	197	184	172	187	157	169	200	193	167	170	192

Parece que la humedad y la temperatura son dos factores que afectan a la riqueza de la especie. ¿por qué no utilizamos toda la información que tenemos e intentamos explicar el comportamiento de la riqueza de parásitos a partir de ambas variables?

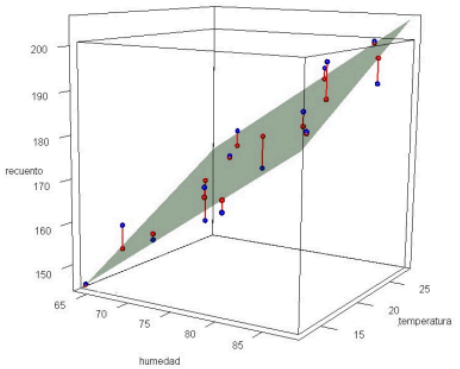
$$\text{Recuento} = \beta_0 + \beta_1 \text{Temperatura} + \beta_2 \text{Humedad} + \epsilon$$

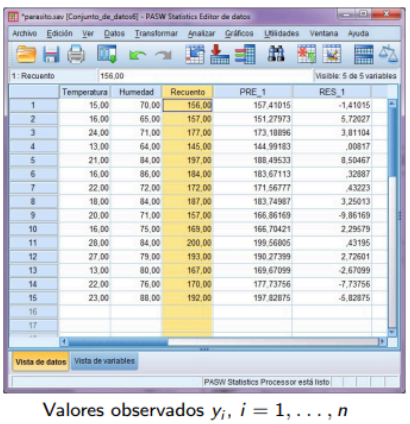


- **Inputs:** Temperatura, Humedad (variable observable).
- **Outputs:** Recuento de parásitos (variable respuesta).

Según el ajuste anterior:

$$\text{Recuento} = 25.7115 + 1.5818 \text{Temperatura} + 1.5424 \text{Humedad}$$

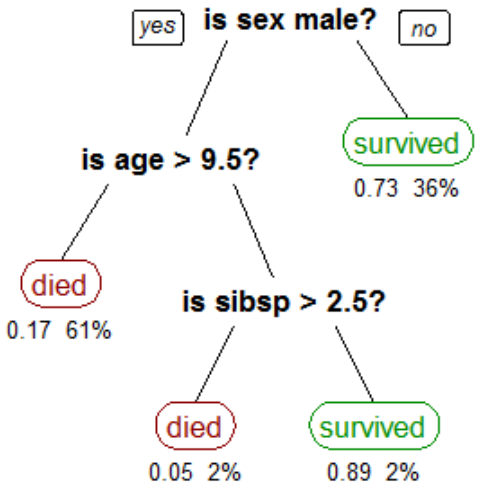




Valores observados y_i, $i = 1, \ldots, n$

EJEMPLO DE ÁRBOL DE DECISIÓN.

Queremos estimar la probabilidad de que los sobrevivientes a un naufragio mueran o sobrevivían en base a su edad y sexo.

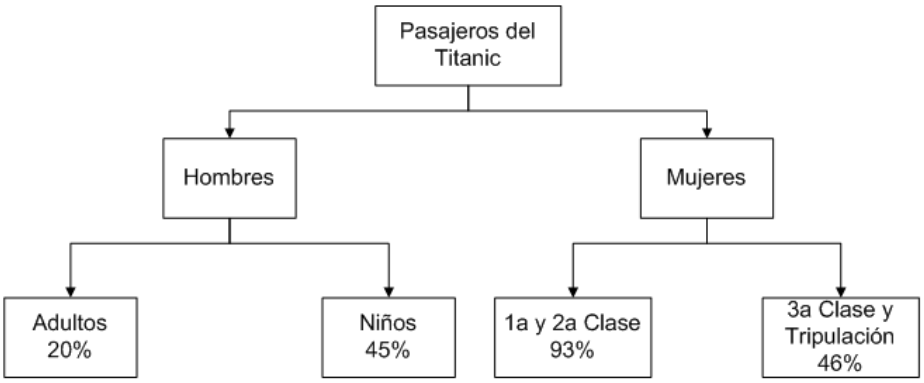


- **Inputs:** sexo, edad , cónyuges o hermanos a bordo (nodos de decisión, variables observables)
- **Outputs:** muere o sobrevive (nodos hojas, variables respuesta)

Por ejemplo si un pasajero no es hombre tiene una probabilidad de sobrevivir del 36%.

EJEMPLO DE ÁRBOL DE REGRESIÓN.

Queremos estimar cual es la probabilidad de que los sobrevivientes al naufragio del Titanic sobrevivían en base al sexo, edad y tipo de clase.



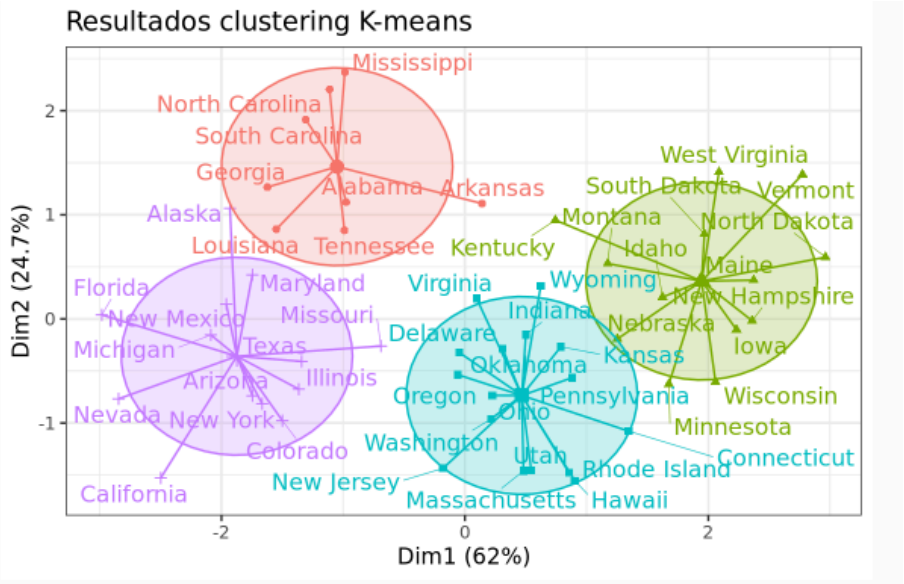
- **Inputs:** sexo, edad, tipo de clase (nodos de decisión, variables observables)
- **Outputs:** porcentaje de supervivencia (nodos hojas, variables respuesta)

Por ejemplo, si un pasajero es hombre y es adulto, entonces tiene una probabilidad de sobrevivir del 20%.

Aprendizaje automático no supervisado: (datos no etiquetados)

EJEMPLO DE CLUSTERING NO JERÁRQUICO.

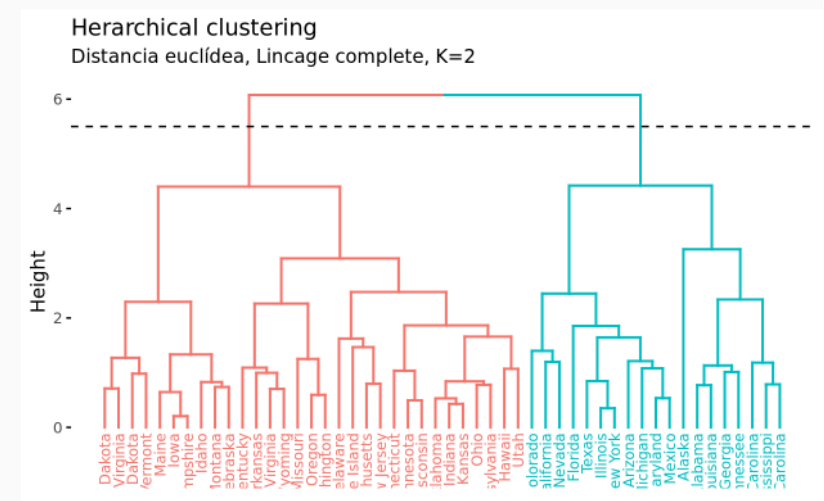
Se parte una base de datos que contiene información sobre el total de delitos (asaltos y asesinatos) en los 50 estados de EE.UU. Se quiere estimar cuales son los estados con características delictivas similares.

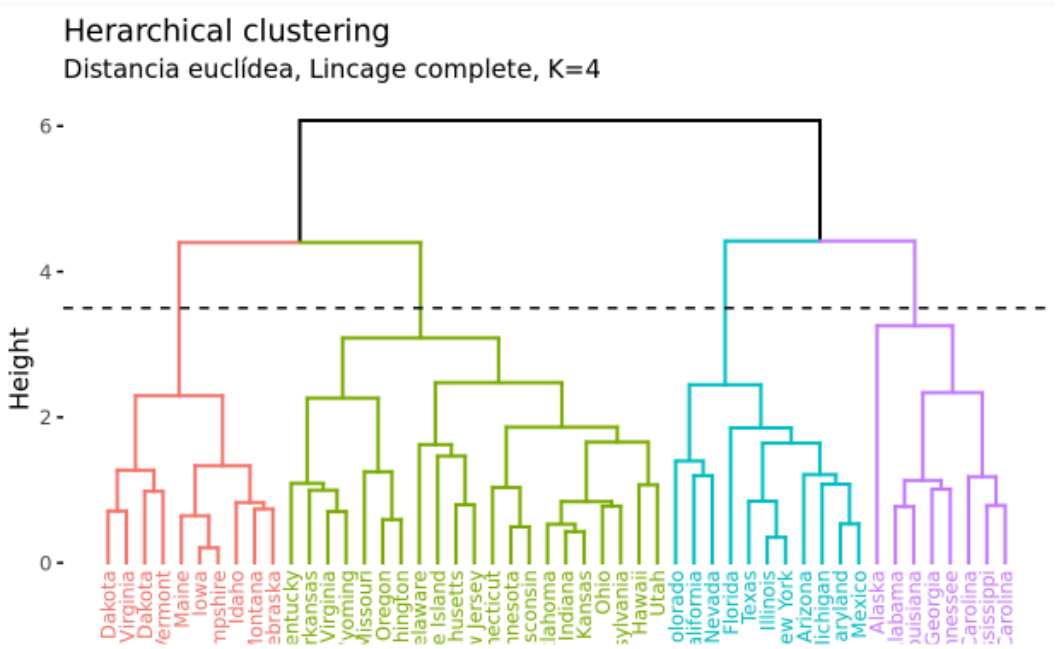


Eje x: cantidad de asesinatos.

Eje y: cantidad de asaltos.

EJEMPLO DE CLUSTERING JERÁRQUICO.





ACTIVIDAD:

1. ¿Cuál es el objetivo del aprendizaje automático?
2. ¿Qué diferencias existen entre el aprendizaje supervisado y no supervisado?
3. Dar ejemplos de aprendizaje automático reconociendo la variable respuesta y las variables observables.