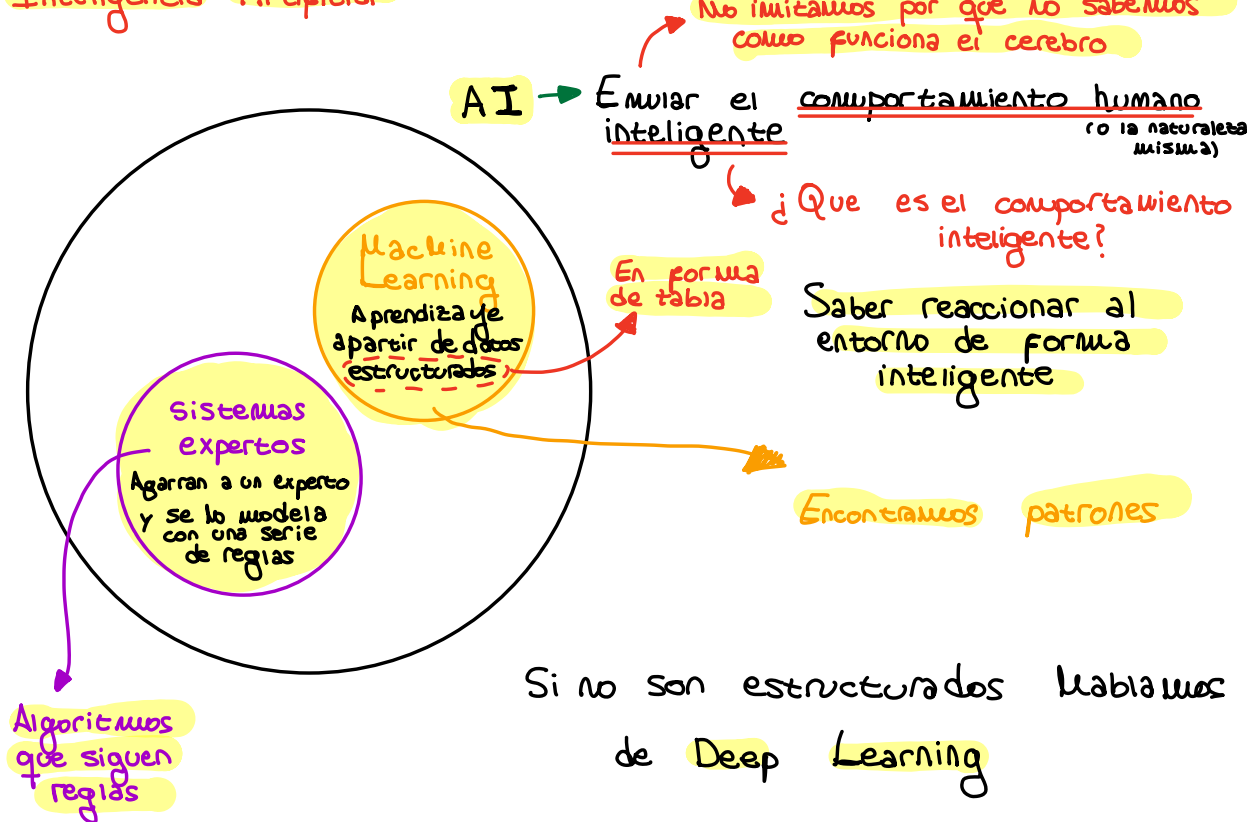
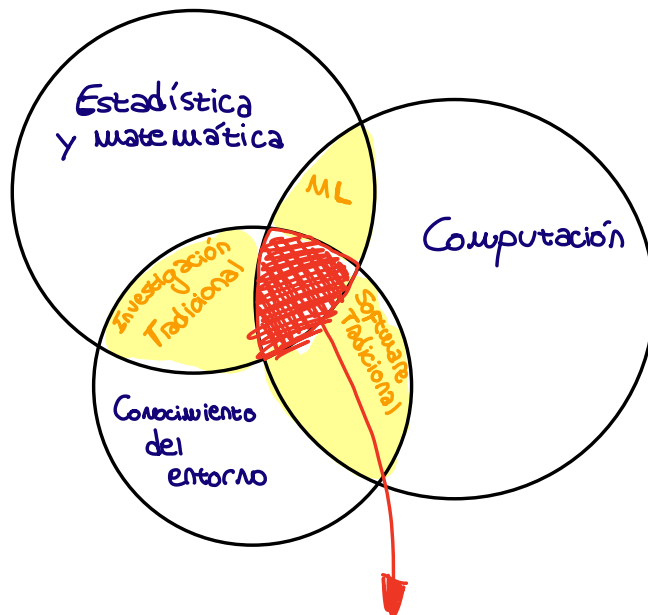


Inteligencia Artificial





"Mas grande la muestra, mejor"

La Big Data te permite aumentar "el N"

Aprendizaje estadístico →

Conjunto de herramientas para entender los datos

No supervisado

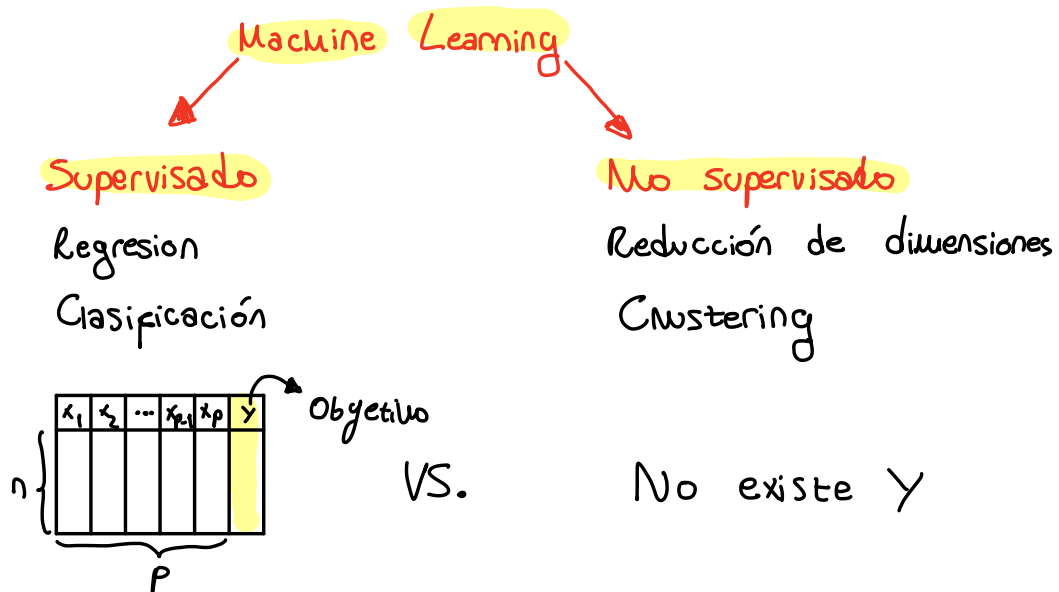
Hay entradas pero no salidas "no supervisadas"
Buscamos relaciones y estructuras de tales datos

Supervisado → Modelos estadísticos para predecir, estimar, una salida basado en uno o mas entradas

- Cuando queremos predecir datos continuos o cuantitativos hablamos de problemas de regresión.
- Cuando queremos predecir datos que no son numéricos (que son categoricos o cualitativos) hablamos de problemas de clasificación
 $\in \{s_1, s_2, \dots, s_n\} \rightarrow$ Conjunto finito
 Ambos problemas son supervisados

En los no supervisados (tenemos las entradas pero no las salidas) tenemos

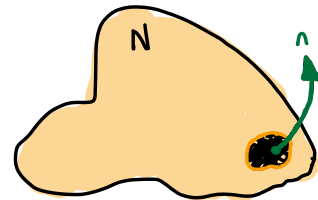
- Cuando queremos agrupar las entradas segun las características observadas hablamos de **clustering**
- Cuando queremos reducir las dimensiones de los datos



Correlación espuria → Conexión entre dos variables que aparentan casualidad cuando en realidad no es así

Inferir → Sacar estimaciones de parámetros

Predicción → Deducir una salida

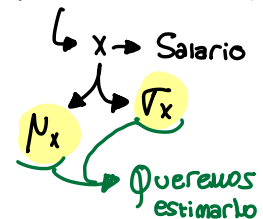


Población de tamaño N → De la población extraemos n muestras ($n \ll N$)

Estimador de μ_x : $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

Estimador de σ_x : $s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$

↳ Es bueno si cumple con los supuestos de estadísticas



TCM → Teorema Central del limite → Suma de variables aleatorias sin importar la distribución que siguen, la suma es normal

Las distribuciones reales son feas. Trabajamos sobre los supuestos
Predecir a secas, es muy difícil.