



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Laboratorio de datos

Cierre - Aprendizaje Automático

Verano 2025

Modelos

Modelo: Cómo planeo modelar el proceso.

Por ejemplo, ¿árbol de decisión o knn? ¿Regresión lineal o polinomial?

Hiperparámetros: Valores que se especifican previo a realizar el aprendizaje, es decir determinan el proceso de aprendizaje automático.

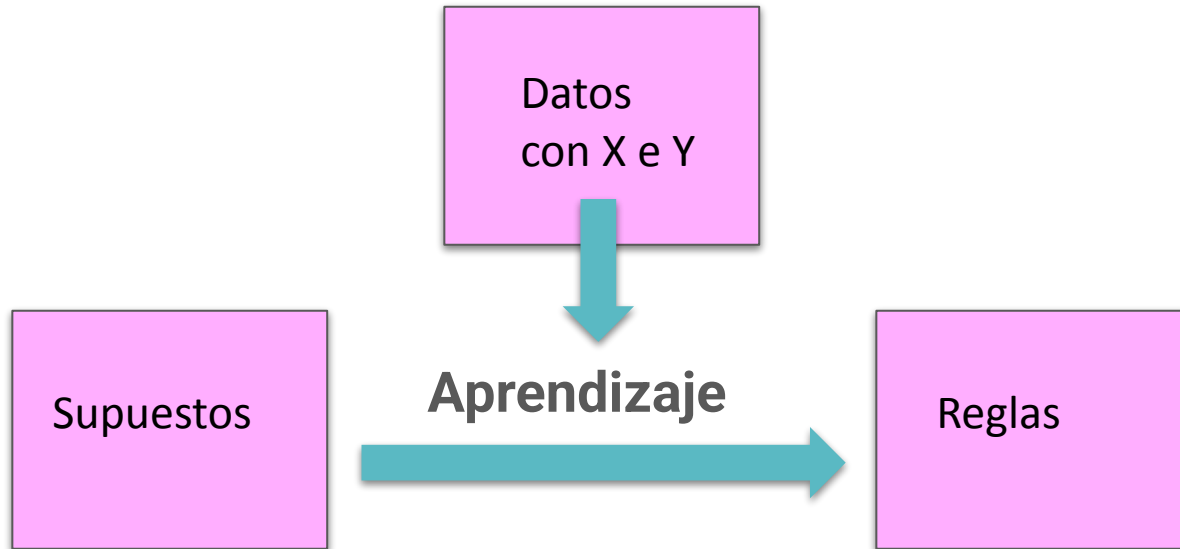
Por ejemplo, en árboles: máxima profundidad permitida, criterio de elección de atributos en cada nodo (Gini Gain, Information Gain), etc. O el valor de k en knn.

Modelo entrenado o Parámetros: Valores internos del modelo resultante **luego del aprendizaje**. Representan las **reglas aprendidas**. Representan el algoritmo final, que es un producto obtenido en el aprendizaje automático.

Por ejemplo, en árboles: las preguntas a realizar en cada nodo del árbol y la clase asignada a cada hoja del árbol.

Aprendizaje supervisado

A partir de un modelo basado en algunos supuestos, se entrena un algoritmo de manera de construir las reglas a partir de la observación de datos etiquetados.



Clasificación vs. Regresión

En clasificación buscamos explicar una variable que es categórica.

- True - False
- Setosa - Versicolor - Virginica
- Sobrevivió - No sobrevivió
- Ceibo - Pindó - Eucaliptus - Jacarandá

En regresión buscamos explicar una variable que es continua (puede tomar valores en \mathbb{R} o en \mathbb{Z})

- Altura de una persona
- Precio de una propiedad
- Temperatura

Clasificación vs. Regresión

Algunos modelos para clasificación

- Umbral
- Árboles de decisión
- knn

Métricas para clasificación

- Exactitud/accuracy
- Precisión/precision
- Exhaustividad/Recall
- F1

Algunos modelos para regresión

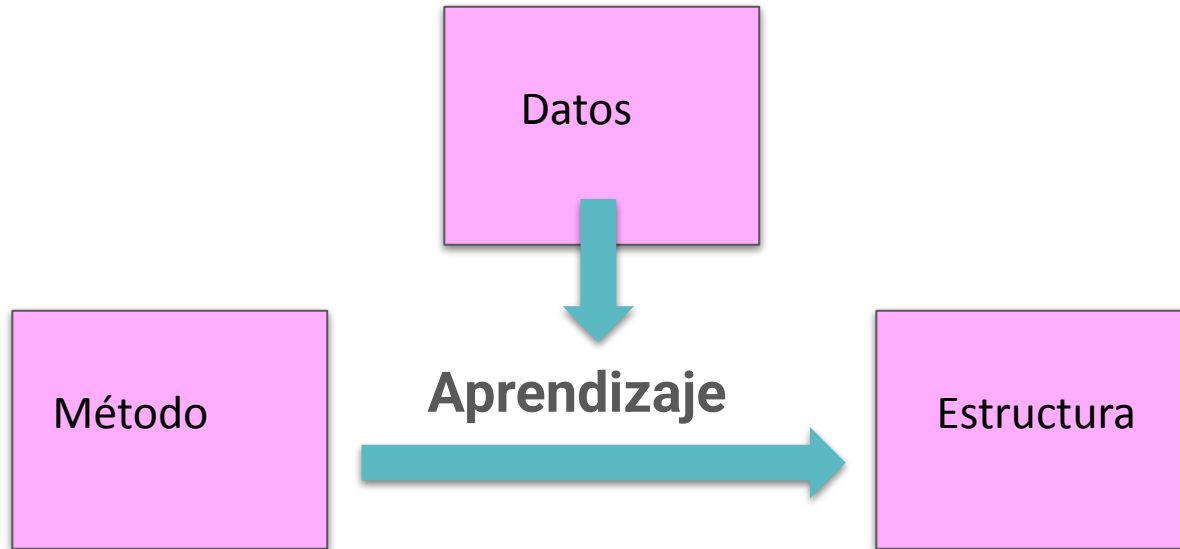
- Modelos lineales
- knn

Métricas para regresión

- R^2
- MSE

Aprendizaje no supervisado

A partir de un modelo o método, se buscan patrones o estructura en los datos.
No se cuenta con datos etiquetados.



Aprendizaje no supervisado

En clustering buscamos agrupar los datos en fragmentos homogéneos.

Métodos:

- K-means
- DBSCAN
- Jerárquico

Medidas:

- WCSS

En reducción de la dimensión buscamos resumir información de los datos.

Métodos:

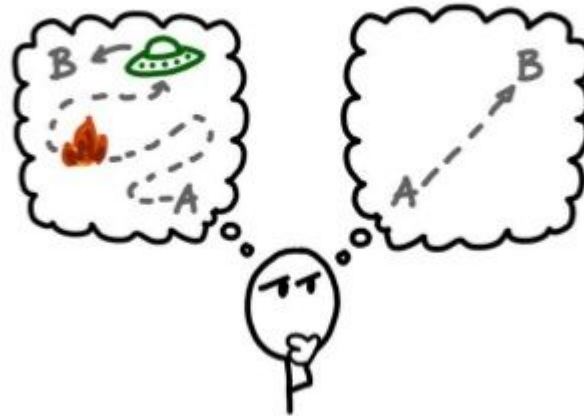
- PCA

Medidas:

- Variabilidad explicada

Principio de la Navaja de Ockham

Occam's Razor



"When faced with two equally good hypotheses, always choose the simpler."

choose the simpler one

Principio de la Navaja de Ockham

Ante dos explicaciones posibles, elegiremos la más sencilla.

Cuando tenemos dos modelos que compiten, y que tienen una predicción similar, debemos elegir el más simple. Es decir, con menos parámetros o menor complejidad.

- Evita el sobreajuste y por lo tanto mejora el rendimiento en datos no vistos.
- Mejora la interpretación: Los modelos más simples son más fáciles de interpretar y entender.
- Reduce los costos computacionales.