



iimas

7ª Escuela de Invierno en Ciencia de Datos y Sistemas Complejos

Unidad Académica del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas en el Estado de Yucatán

Modelos de aprendizaje de máquinas y su interpretación para la toma decisiones

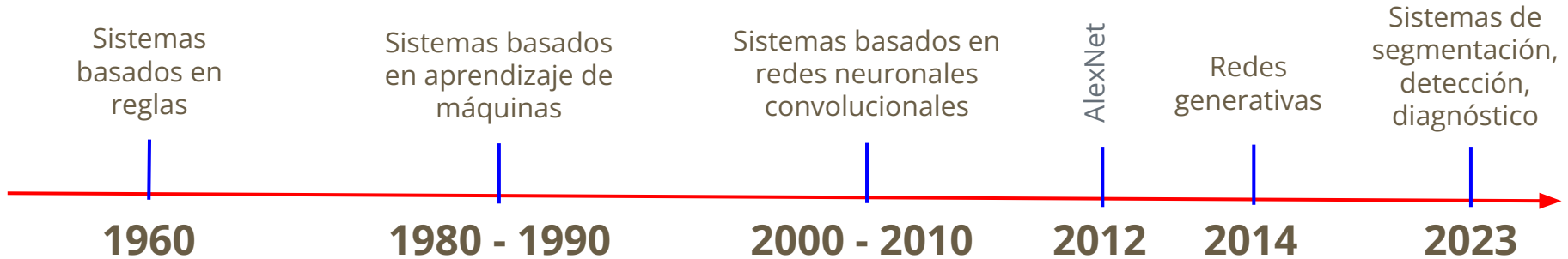
Presenta

Blanca Vázquez

Artificial Intelligence in Biomedicine Group (ArBio)

18 de enero de 2024

Evolución de la Inteligencia Artificial (IA)



Aprendizaje de máquinas

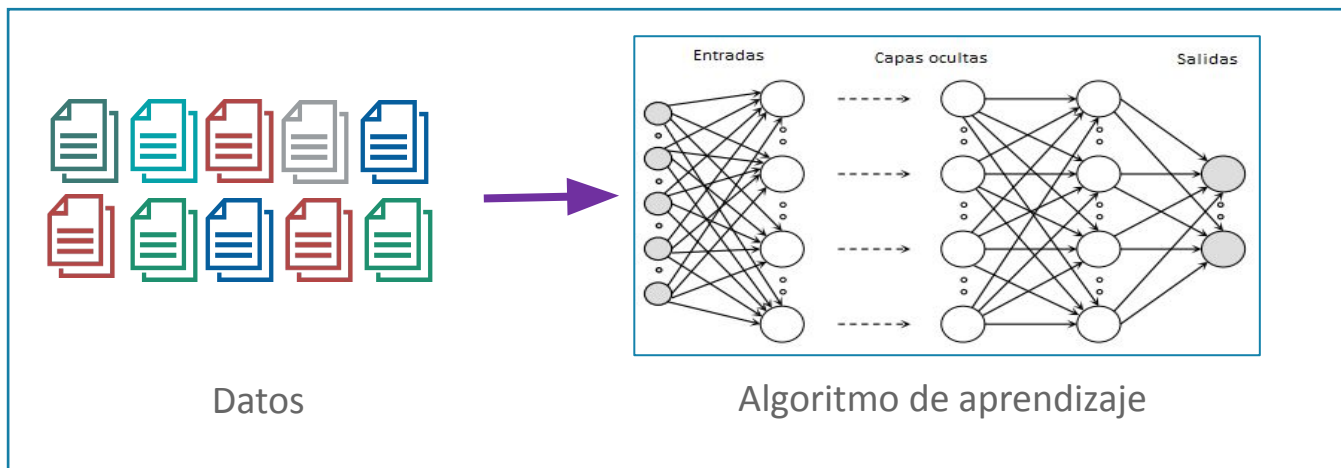
- ❖ Es el estudio de programas que aprenden a partir de ejemplos para estimar comportamientos futuros.
- ❖ El resultado de un programa es la probabilidad de que un evento ocurra.



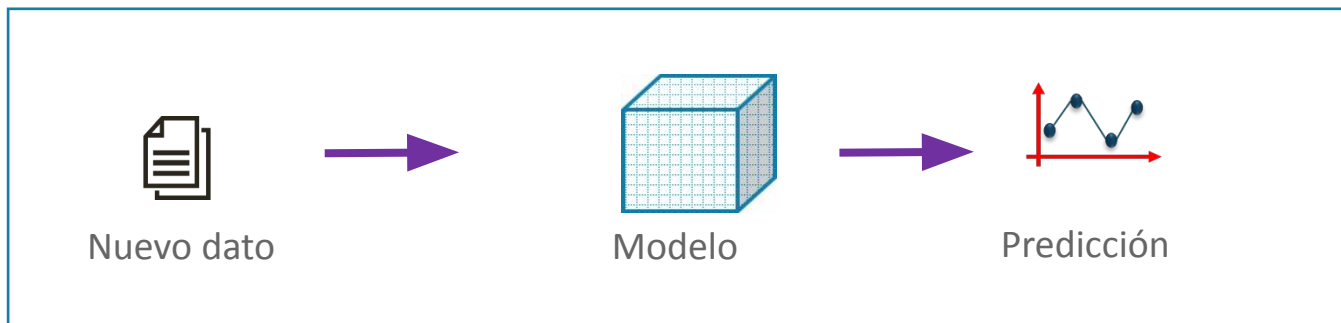
La idea: ¡aprender de los datos!

¿Cómo funciona?

Entrenamiento



Inferencia



Ejemplo de aplicaciones



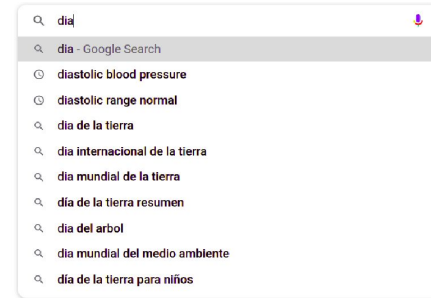
Reconocimiento de rostros



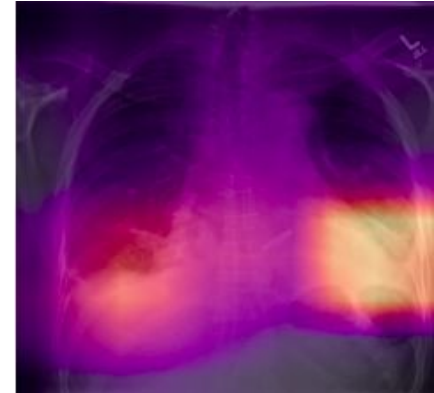
Reconocimiento de emociones



Sugerir palabras



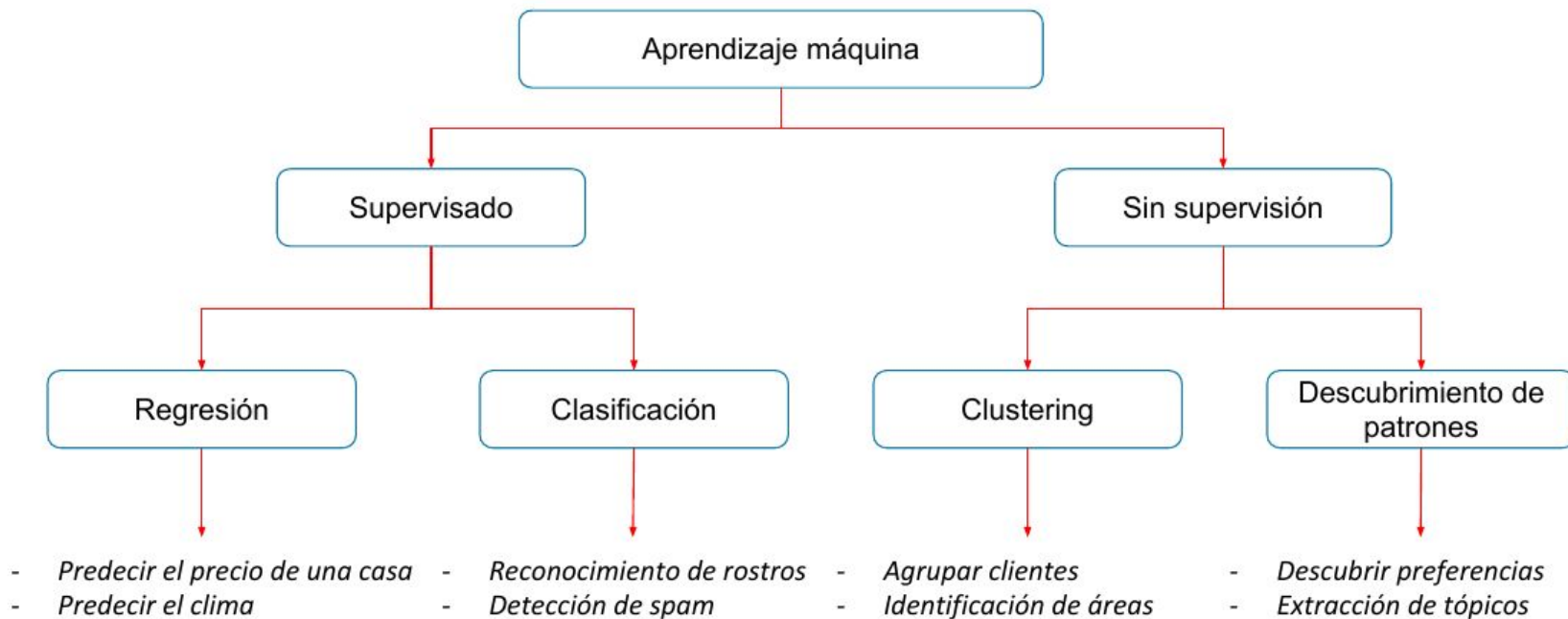
Diagnóstico y detección de enfermedades



Vista general del desarrollo y validación de modelos de aprendizaje de máquinas



Tipos de aprendizaje



Evaluación del rendimiento de los modelos

Una matriz de confusión es una representación matricial que describe el rendimiento de un modelo de clasificación binaria.

		Clase predicha	
		Positivo	Negativo
Clase real	Positivo	Verdadero positivo	Falso negativo
	Negativo	Falso positivo	Verdadero negativo

Matriz de confusión

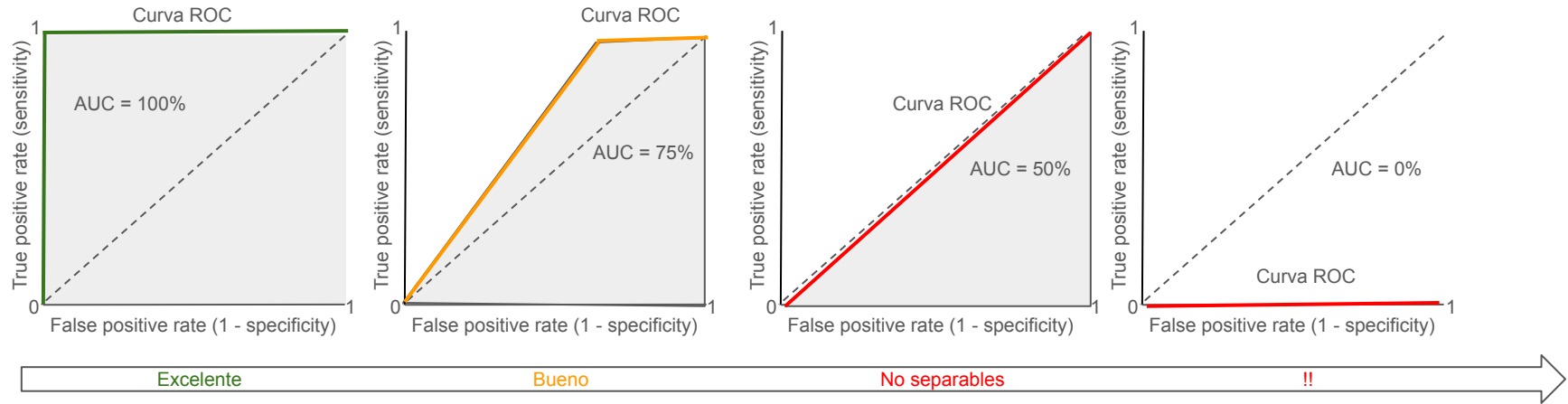
$$accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

$$recall = \frac{VP}{VP + FN}$$

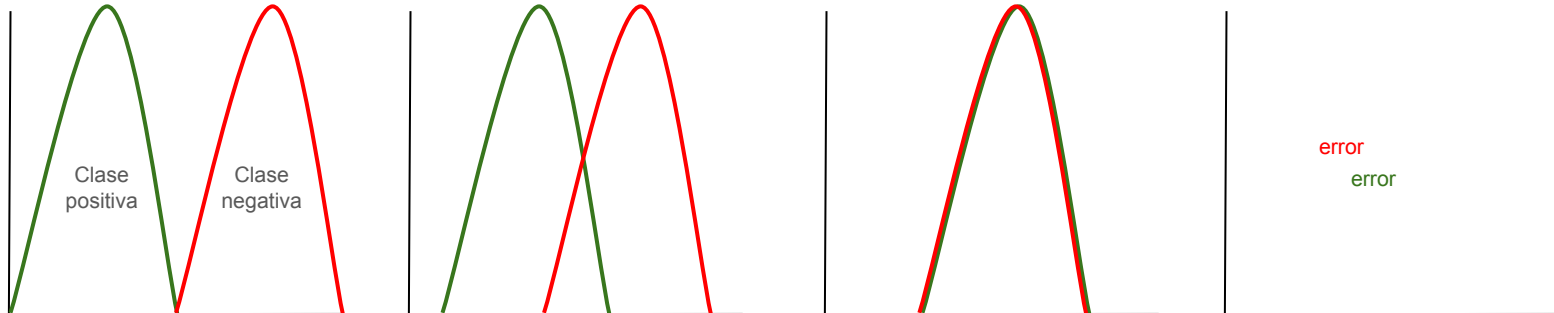
$$precisión = \frac{VP}{VP + FP}$$

$$especificidad = \frac{VN}{VN + FP}$$

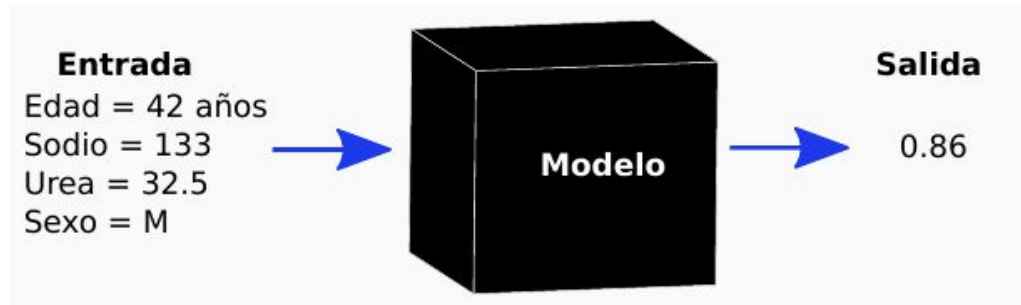
Curva ROC (Receiver Operating Characteristic)



Traslape: ¿qué tan bueno es el modelo para separar las clases positivas y las negativas?



Retos en los modelos de predicción



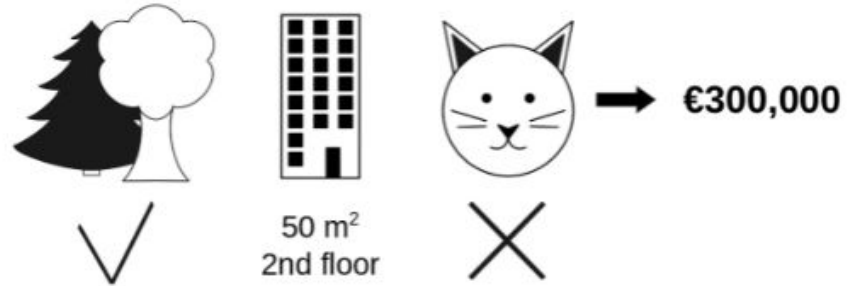
Interpretability is the degree to which a human can understand the cause of a decision (Miller, 2017).



Ejemplo

Supongamos que construimos un modelo para predecir precios de departamentos, las características que tomamos en cuenta para la predicción son:

- El número de pisos del edificio
- Si se permiten mascotas
- La cercanía a un parque
- El número de habitaciones
- Número de metros cuadrados



Tipos de interpretabilidad

Intrínsecos

Métodos de aprendizaje que por su naturaleza sencilla pueden ser explicados sin necesidad de cálculos adicionales.



- Árboles de decisión

Pos-adhoc

Métodos que se utilizan posterior al entrenamiento de los modelos.



- Específicos
 - Dependen del modelo
 - Interpretan los pesos de una regresión o de una red profunda.
- Agnósticos
 - Son independientes del modelo.
 - No tienen acceso a los pesos del modelo construido.

Shap values: modelo para interpretar la salida de los modelos

- Es un enfoque basado en los valores Shapley y la teoría de juegos.
- Donde el juego es la tarea de predicción para una instancia y los jugadores son los valores de las características
- El valor Shapley ϕ_j (val) es el pago justo que recibe un jugador j por el juego y se define de la siguiente manera:

$$\phi_j(val) = \sum_{S \subseteq \{1, \dots, p\} \setminus \{j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{j\}) - val(S))$$

donde la suma se basa en todos los subconjuntos posibles s de los demás jugadores; val es una función que devuelve la contribución de un subconjunto dado y p es el número total de jugadores.

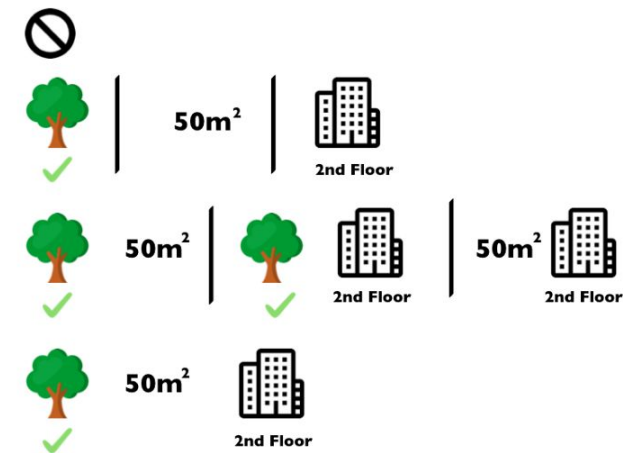


Imagen tomada de Matthew Stewart, 2020.

Shap values: modelo para interpretar la salida de los modelos

- Para calcular la importancia de las variables, se promedian los valores absolutos Shapley sobre todos los conjuntos de datos:

$$I_j = \sum_{i=1}^n |\phi_j^{(i)}|$$

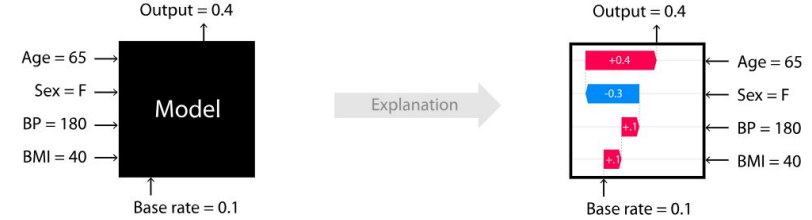


Imagen tomada de <https://github.com/shap/shap>

- Las variables con valores grandes Shapley se consideran importantes en la predicción.
- Se distingue por su capacidad interpretativa a nivel global, local e individual.

Inteligencia artificial en el área clínica

Diagnóstico	Pronóstico	Tratamientos
Predecir una enfermedad: <ul style="list-style-type: none">- Cáncer- Neuro-degenerativas- Respiratorias- Cardiacas- Autoinmunes	Predecir el riesgo de un evento: <ul style="list-style-type: none">- Mortalidad- Falla cardíaca- Accidente cerebro-vascular- Readmisión hospitalaria- Tamaño de la estancia- Intervención clínica- Estimar la supervivencia	Estimar los efectos de un tratamiento: <ul style="list-style-type: none">- Medicina personalizada

¿Por qué aplicar modelos predictivos en el área de salud?

- ❖ Recursos limitados
- ❖ Decisiones críticas
- ❖ Cada minuto cuenta



<https://southwesthealthcollaborative.org/workgroups/emergency-department-utilization/>

¿Quiénes están trabajando?

MIT Clinical ML



Google Research

Stanford ML



UNIVERSITY
OF
CALIFORNIA



IBM
Watson
Health



USC University of
Southern California

y muchos más

... pero ¿cuál es el objetivo en común?



Construir herramientas
computacionalmente viables y
médicamente relevantes que
contribuyan a la toma de decisiones.

Imagen tomada de <https://trends.levif.be>

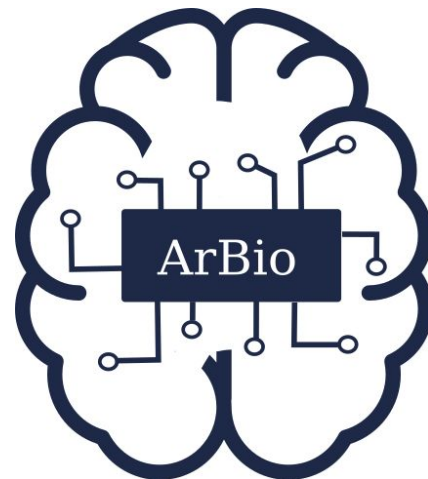
Gracias

Blanca Hilda Vázquez Gómez

Unidad Académica del IIMAS en el estado de Yucatán.

blancavazquez2013@gmail.com

<http://turing.iimas.unam.mx/~blancavg/>



Artificial Intelligence in Biomedicine Group