



Rodolfo Ferro

ferro@cimat.mx

<https://rodolfoferro.xyz>

Clubes de Ciencia México 2025

"Satélites y Neuronas: Explorando lo Invisible con AI"

Julio, 2025

Introducción a ML

Día 2

Rodolfo Ferro (ferro@ciimat.mx)

- › Sr. SWE (Data Engineer) @ Bisonic México
- › Tutor de Ciencia de Datos en Código Facilito y en el Diplomado en Ciencia de Datos de la ENES UNAM León
- › Profesor de AI & Coordinador del TomorrowLab @ EdgeHub School of Innovation (Ags.)
- › **Formación:** BMath @ UG, CSysEng @ UVEG, StatMethods Specialist @ CIMAT
- › **Experiencia:** ML Engineer @ Vindoo.ai (España), Sherpa Digital en IA @ Microsoft México, AI Research Assistant @ CIMAT & AI Research Intern @ Harvard.



ferro@ciimat.mx

Tabla de contenidos

1 Intro al aprendizaje de máquina

2 Tipos de IA

3 Creación de datasets para modelos de aprendizaje automático

- 1 Intro al aprendizaje de máquina
 - Motivación
 - ¿Qué es la IA?
 - ¿Por qué el ML?
- 2 Tipos de IA
 - Introducción a modelos generativos
- 3 Creación de datasets para modelos de aprendizaje automático
 - Definición de dataset
 - Consideraciones al crear un dataset
 - ¿Cómo se divide un dataset?



Section 1: **Intro al aprendizaje de máquina**



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```

Motivación

Intro al aprendizaje de máquina



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else if(speed<12){  
    status=RUNNING;  
} else {  
    status=BIKING;  
}
```

Motivación

Intro al aprendizaje de máquina

Motivación

Intro al aprendizaje de máquina



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else if(speed<12){  
    status=RUNNING;  
} else {  
    status=BIKING;  
}
```



```
// ????
```

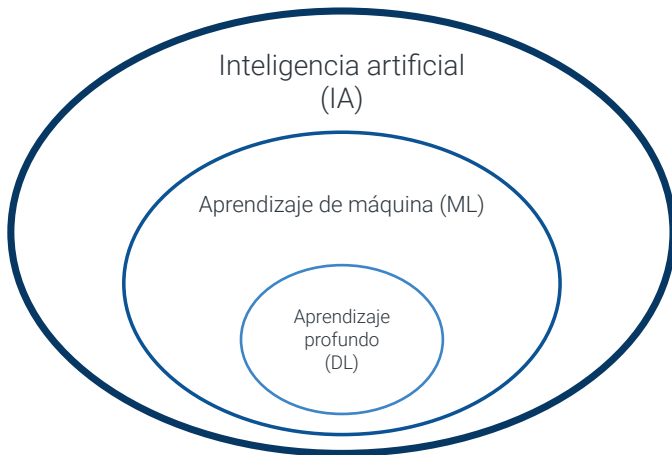
¿Qué es el la IA?

Intro al aprendizaje de máquina

- **Inteligencia artificial:** Cualquier técnica que permita a las computadoras emular o imitar el comportamiento humano.
- **Aprendizaje de máquina:** Capacidad de aprender sin ser programado explícitamente, enfoque en los algoritmos y la matemática.
- **Aprendizaje profundo:** Extrae patrones de datos utilizando redes neuronales.

¿Qué es el la IA?

Intro al aprendizaje de máquina



ferro@ciimat.mx

¿Por qué el *Machine Learning*?

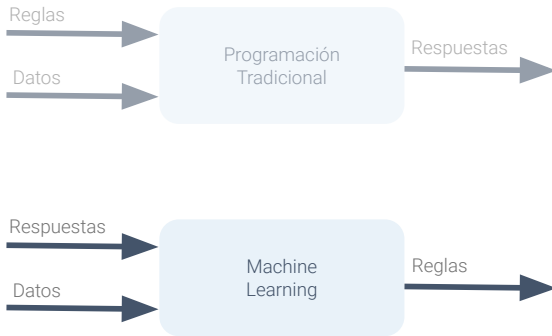
Intro al aprendizaje de máquina



ferro@ciimat.mx

¿Por qué el *Machine Learning*?

Intro al aprendizaje de máquina



ferro@ciimat.mx

EJEMPLOS

Pregunta: ¿Qué IAs conoces? ¿Con qué IAs has interactuado?



Section 2: **Tipos de IA**

› Modelos discriminativos:

- › Se enfocan en la predicción de etiquetas.
- › Ejemplos: Regresión logística, SVM, redes neuronales feedforward.

› Modelos generativos:

- › Modelan la distribución conjunta de los datos.
- › Ejemplos: Redes Bayesianas, GANs, VAEs.

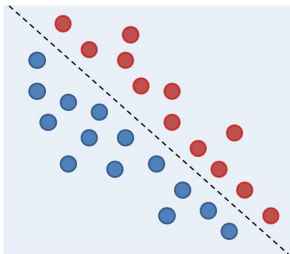
- › **En esencia...** Los modelos discriminativos responden a preguntas como "¿A qué clase pertenece esta muestra?", mientras que los generativos responden a "¿Cómo se ve una muestra típica de esta clase?".

¿Qué son los Modelos Generativos?

Tipos de IA

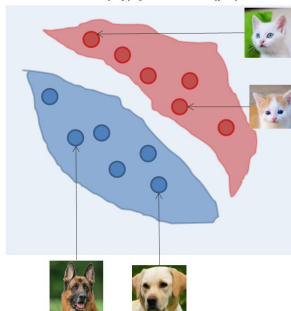
Discriminativo

Estimar directamente $P(y/x)$



Generativo

Estima $P(x/y)$ y deduce $P(y/x)$



ferro@ciimat.mx

¿Qué son los Modelos Generativos?

Tipos de IA

- Un modelo generativo es un tipo de modelo que puede generar nuevos datos a partir de un conjunto de datos de entrenamiento.
- A diferencia de los modelos discriminativos, que se centran en predecir etiquetas a partir de características, los modelos generativos buscan aprender la distribución subyacente de los datos para poder generar nuevas muestras.
- **En esencia...** Los modelos generativos se enfocan en modelar cómo se distribuyen los datos de forma que puedan generar datos que "parecen" reales.

› **Imágenes:**

- › Generación de rostros humanos (e.g., StyleGAN).
- › Generación de arte abstracto.

› **Texto:**

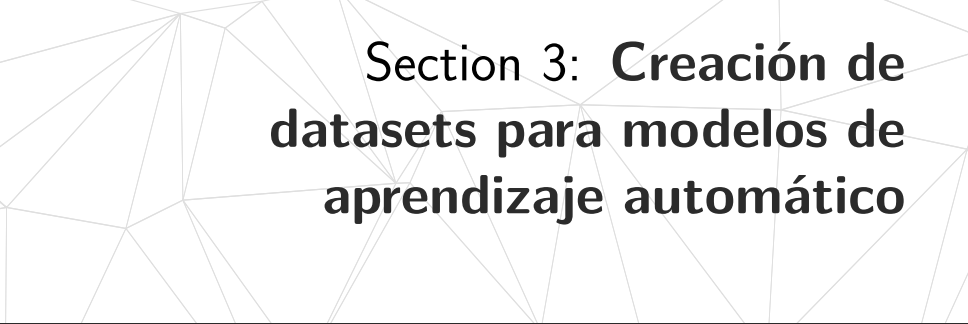
- › Composición automática de párrafos o historias.
- › Traducción automática y generación de poesía.

› **Música:**

- › Composición de melodías.
- › Generación de pistas basadas en un estilo dado.

› **Videojuegos:**

- › Generación de mundos y personajes.
- › Creación de niveles procedurales.



Section 3: Creación de datasets para modelos de aprendizaje automático

Un dataset D es un conjunto de observaciones estructuradas que se utilizan para entrenar, validar o evaluar modelos de aprendizaje automático.

Matemáticamente, lo podemos representar como:

$$D = \{(x_i, y_i)\}_{i=1}^n$$

donde:

- $x_i \in \mathbb{R}^d$ es el vector de características (o entrada) del ejemplo i -ésimo
- y_i es la etiqueta o salida correspondiente (puede ser una clase, un valor continuo, una máscara, etc.)
- n es el número total de ejemplos/muestras en el dataset

Tipos de datasets:

- › **Datos etiquetados:** (x_i, y_i) – aprendizaje supervisado
- › **Datos no etiquetados:** sólo x_i (sin y_i) – aprendizaje no supervisado
- › **Datos parcialmente etiquetados:** aprendizaje semi-supervisado.

Consideraciones al crear un dataset

Creación de datasets para ML

1 Relevancia del problema:

- » ¿Qué queremos que el modelo aprenda?
- » Definir claramente la variable objetivo y y las variables predictoras x

2 Calidad de los datos:

- » Datos limpios y representativos
- » Etiquetas en formato correcto
- » Evitar sesgos y duplicados

3 Tamaño del dataset:

- » Suficientes para capturar la variabilidad del problema
- » Pero no excesivo si se pueden extraer patrones antes (data augmentation o técnicas de balanceo)

4 Balance de clases (para clasificación):

- » Evitar clases desbalanceadas

ferro@ciimat.mx

¿Cómo se divide un dataset?

Creación de datasets para ML

$$D = D_{train} \cup D_{val} \cup D_{test}$$

- **Entrenamiento (train):** Para ajustar los parámetros del modelo.
- **Validación (validation):** Para ajustar hiperparámetros y prevenir sobreajuste.
- **Prueba (test):** Para evaluar el rendimiento final del modelo.

Ejemplo típico de partición:

- 70% **entrenamiento**
- 15% **validación**
- 15% **prueba**

EJEMPLOS

Idea intuitiva de los modelos de IA.