

# Apuntes

Isaac Araya Solano

2018151703

8 de Febrero de 2024

## Introducción a la Inteligencia Artificial

### Arthur Samuel

Arthur Samuel fue un pionero en el campo de la inteligencia artificial. Trabajó para IBM y se le reconoce especialmente por haber acuñado el término "aprendizaje automático" (*machine learning*) en 1959. En ese mismo año, Arthur Samuel fue capaz de desarrollar el primer programa que podía jugar a un juego de tablero a un alto nivel, concretamente se trataba del checkers, un juego bastante parecido a las damas españolas. Lo más interesante de este programa es que era capaz de mejorar su nivel de juego a medida que jugaba, ajustando los pesos de las variables importantes para la toma de decisiones [1]. Samuel nació el 5 de diciembre de 1901 y falleció el 29 de julio de 1990.

## Machine Learning

¿Qué es machine learning?

- Se puede ver desde la parte de la ciencia y de la ingeniería
  - Ciencia: entender a profundidad cómo funcionan las cosas, crear nuevos algoritmos, etc.
  - Ingeniería: Tengo este conocimiento, ¿cómo hago que se use en producción y se implemente en la industria y producción?
- Se busca: "Construir máquinas que puedan hacer tareas sin ser explícitamente programadas, solo infiriendo de los datos". Actualmente nosotros le decimos qué hacer a la máquina, pero se busca que la máquina pueda aprender a hacer una función infiriendo a partir de una serie de datos.
- Creamos un algoritmo que va a aproximar una función y estas aproximaciones son útiles pero no perfectas.

# Aristas del Machine Learning

## Ciencia:

- Generamos conocimientos
- Métricas
- Data Scientist
- Research Scientist

Los data y research scientists analizan artículos para mejorarlos o crean los suyos propios con el objetivo de obtener mejores métricas. La métrica más famosa es el accuracy que es la precisión que tiene el modelo al completar su función. Ellos no se dedican al software sino se dedican a la mejora de métricas.

## Ingeniería

- Puesta en producción de un modelo
- Transformar el modelo
- ONNX: es un formato abierto para modelos ML, que te permite intercambiar modelos entre varios marcos y herramientas de ML [2]
- MLOps

Esto sucede cuando ya se tiene un modelo y se busca al machine learning engineer para que pueda poner el modelo a producción. Para llevar a un modelo a producción se debe tener en cuenta la escalabilidad y otros factores.

Se usa Python para generar los modelos de machine learning pero se suele migrar a otros lenguajes que son más eficientes usando programas como ONNX.

Se usan técnicas como destilación que, por ejemplo, pueden disminuir el tamaño de una red neuronal, tal vez haciéndolo más eficiente pero reduciendo algunas métricas. Debe tenerse en cuenta que a veces vale la pena perder eficiencia para disminuir el tamaño del modelo y así beneficiar otros aspectos.

— Se dice que la Inteligencia Artificial engloba el machine learning que a su vez, abarca el deep learning.

### 0.0.1 Inteligencia Artificial (AI)

Puede simular sentidos humanos. Se incluyen algoritmos generativos, simbolismo y definir comportamientos.

### 0.0.2 Machine Learning (ML)

Se usan métodos estadísticos como regresión lineal, logística (clasificación binaria) y árboles de decisión. Son algoritmos que mejoran si son expuestos a más datos en el tiempo.

### 0.0.3 Deep Learning (DL)

Es una parte del machine learning con redes neuronales con muchas capas aprenden de una cantidad enorme de datos.

#### Tipos de aprendizaje:

- Supervisado: se supervisa el aprendizaje del algoritmo.  
Se le da una serie de tuplas y el algoritmo aprende a ver los resultados y aprender de ellos. Se llama supervisado porque yo a mi modelo le doy una serie de datos ya clasificados (con el label) y luego se le pide al modelo que haga la predicción a partir de los pares ordenados provistos.
- No supervisado  
Se provee una lista de datos con únicamente las x, sin proveer la etiqueta. Se necesita clasificarlos pero no se le enseña cómo.
- Semi-supervisado  
Es muy similar al supervisado pero la cantidad de datos que se le dan clasificados son muy pocos. Se usa cuando tenemos etiquetas o clases pero tenemos pocos ejemplos de esas clases.
- Auto-supervisado  
Se usa para lenguaje natural. Se usa para la predicción de la siguiente palabra.
- Refuerzo  
Se usa con feedback positivo para hacer que el modelo vaya aprendiendo qué está bien y qué está mal conforme va generando respuestas.
- Few-shot  
Le doy ciertos ejemplos al modelo para que aprenda a hacer la tarea.
- One-shot  
Le doy un único ejemplo al modelo para que haga la tarea.
- Zero-shot  
La inteligencia artificial puede hacer una tarea sin ningún ejemplo. Se cree que en el futuro se avanzará hacia esto.

## Machine Learning Pipeline

#### Pasos:

- Data Acquisition
- Data Preparation
- Feature Engineering

- Model Selection
- Model Training
- Model Deployment

## Data Acquisition

- Los datos deben ser relevantes al problema a resolver
- La calidad de los datos importa
- ¿De dónde sacamos los datos?

## Data Preparation

- ¿Cómo le presento los datos a mi algoritmo?
- Valores duplicados o faltantes
  - Si hay un dato faltante se puede colocar el promedio para reemplazarlo y algunas veces se agrega el mínimo o el máximo
- Transformación de datos (normalización, escalamiento)
- Objetivo: Brindar formato adecuado para su análisis

## Feature Engineering

- Feature: atributo que proporciona información de cada instancia
- Creación de nuevos features a partir de existentes
- Selección de features relevantes
- Correlación de datos
- Objetivo: Proporcionar features más informativos y relevantes

## Model Selection

- Seleccionar el mejor modelo que se ajuste a los requisitos del problema
- Explicabilidad (Debo tener en cuenta esto a la hora de escoger un modelo)
- Ejemplo: si se trata de un crédito bancario y se usa una red neuronal para definir si una persona puede acceder al crédito o no, no podríamos explicar a la persona por qué, por eso una red neuronal no sería buena para este caso.

## Model Training

- Hyper-parameters tuning (GridSearch)
- Los algoritmos hacen optimización basada en los datos de entrenamiento

## Model Deployment

- Puesta en producción del modelo
- Machine learning: ingeniería
- Se usa ONNX, técnicas de destilación y demás...

## References

- [1] M. B. D. e Inteligencia Artificial Málaga, “Los primeros programas de ia capaces de aprender,” Sin fecha. [En línea]. Disponible el 16 de febrero de 2024.
- [2] Microsoft Learn: Build skills that open doors in your career, “Modelos de onnx,” Sin fecha. [En línea]. Accedido el 16 de febrero de 2024.