



Aprendizaje Automático y Análisis de Datos

Introducción

Julián Gil González

julian.gil@javerianacali.edu.co (Periodo 2023-I)

27 de enero de 2023

Metodología de la clase I

- Curso de tres créditos: 3 horas de clase por semana y 3 horas de trabajo individual.
- El trabajo individual puede consistir de:
 - Lecturas en inglés.
 - Talleres de ejercicios.
 - Revisión de videos y/o tutoriales.
 - Tareas de implementación.

Metodología de la clase II

- El trabajo individual estará disponible en la plataforma Brightspace junto con el material de la clase.
- Los trabajos de implementación se realizarán exclusivamente en Python (Colab).
- El horario de atención a estudiantes será el viernes de 8:00am a 10:00am. Edificio central, oficina 2-41.

Manual de Convivencia

- Se toma lista pasando una hoja durante los primeros 15 minutos de la clase, donde cada persona debe firmar. Quien llegue luego de esos 15 minutos puede entrar a clase pero no debe firmar la asistencia.
- Se les solicita a todos llegar puntualmente, entrar tarde puede interrumpir la clase.
- Los celulares deben permanecer en silencio durante toda la clase.
- Si deben responder mensajes o llamadas urgentes, lo deben hacer fuera del salón.
- Solo se puede interrumpir la clase para resolver dudas.

Agenda

Temas:

- Introducción al curso.
- Concepto de aprendizaje.
- Definición de una tarea de aprendizaje

Objetivos del aprendizaje: Al final de esta clase los estudiantes estarán en la capacidad de:

- Explicar los componentes de un sistema de aprendizaje.
- Plantear una tarea de aprendizaje.
- Identificar las principales aplicaciones del aprendizaje automático.



Table of Contents

- ▶ **Introducción**
- ▶ Tareas de aprendizaje
- ▶ Aspectos prácticos del aprendizaje de máquina

Campos relacionados

Data Science Is Multidisciplinary

By Brendan Tierney, 2012

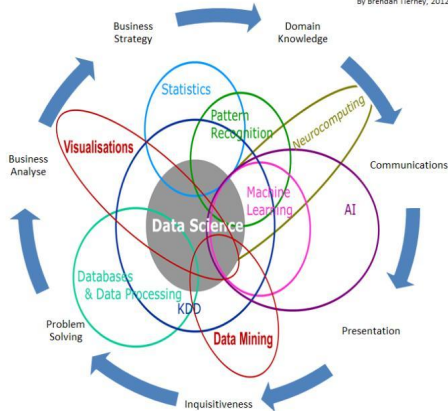


Figura: Tomada de: <https://www.datasciencecentral.com/difference-of-data-science-machine-learning-and-data-mining/>

El renacer de la inteligencia artificial – Aprendizaje profundo I

TECH / ARTIFICIAL INTELLIGENCE / SCIENCE

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing



From left to right: Yann LeCun | Photo: Facebook. Geoffrey Hinton | Photo: Google. Yoshua Bengio | Photo: Botler AI

/ Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By JAMES VINCENT

Mar 27, 2019, 9:02 AM GMT-5 | [0 Comments](#) / [0 View](#)



Figura: Tomada de: <https://www.theverge.com/2019/3/27/18280665/ai-godfathers-turing-award-2018-yoshua-bengio-geoffrey-hinton-yann-lecun>

Yann Lecun (Chief AI Scientist at Facebook), Geoffrey Hinton (Google), Joshua Bengio (Universidad de Mostreal).

El renacer de la inteligencia artificial – Aprendizaje profundo II

- En 2006, Hinton et. al demostraron que un algoritmo de aprendizaje profundo podía reconocer dígitos escritos a mano con una precisión $> 98 \%$.
- El aprendizaje profundo rápidamente empezó a superar el rendimiento de algoritmos clásicos.
- El factor clave: Poder de cómputo (GPUs) y grandes cantidades de datos.
- El aprendizaje de máquina es la base de muchos productos tecnológicos (Reconocimiento de voz, autos que se conducen de forma autónoma).

¿Qué es el aprendizaje de máquina?

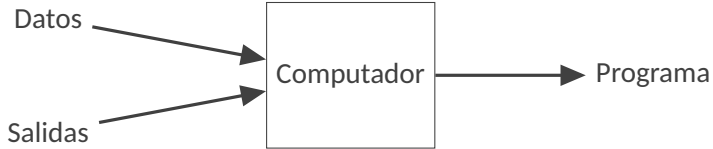
- Según **Arthur Samuel** (1959) el aprendizaje de máquina-(AM) es el campo de estudio que le da a los computadores la habilidad de aprender sin ser programados explícitamente.
- Según **Tom Mitchell** (1997), el aprendizaje automático es el estudio de algoritmos que:
 - Dada una tarea **T** (Detectar Spam).
 - Aprovechan las experiencias (datos) **E** (Correos catalogados como Spam o no).
 - Mejoran su rendimiento **P** (¿Cómo?).
- En otras palabras, el aprendizaje de máquina consiste en programar sistemas computacionales que aprendan de los datos.

¿Por qué usar aprendizaje de máquina? I



Aprendizaje por reglas impuestas. Reglas difíciles de mantener y escalar a grandes cantidades de datos.

¿Por qué usar aprendizaje de máquina? II



Con base en un modelo o algoritmo, se aprende de los datos. Permite tener un mayor nivel de automatización.

¿Por qué usar aprendizaje de máquina? III

En resumen, el aprendizaje de máquina es útil para:

- Problemas donde las soluciones requieren de grandes listados de reglas: **El aprendizaje de máquina permite tener códigos más simples que usualmente funcionan bien.**
- Problemas complejos que no pueden solucionarse con técnicas de programación.
- Entornos que cambian: **Los sistemas de aprendizaje automático se pueden adaptar a nuevos datos.**
- Obtener información acerca de problemas complejos con grandes cantidades de datos. **El aprendizaje de máquina puede ayudarnos a conocer los problemas.**



Table of Contents

► Introducción

► Tareas de aprendizaje

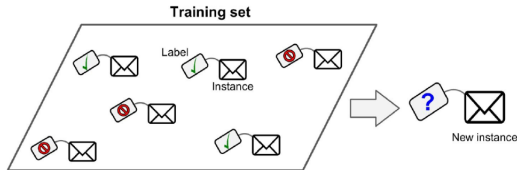
► Aspectos prácticos del aprendizaje de máquina

Tipos de aprendizaje

Se reconocen al menos tres tipos de aprendizaje de máquina:

- Aprendizaje supervisado.
 - Disponible: Datos de entrenamiento + Salidas deseadas (etiquetas).
- Aprendizaje no supervisado.
 - Disponible: Datos de entrenamiento (no se dispone de salidas deseadas).
- Aprendizaje por refuerzo.
 - Recompensas o castigos luego de una secuencia de acciones.

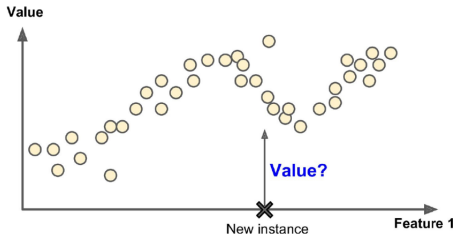
Aprendizaje supervisado: Clasificación¹



- Instancia: muestra del fenómeno.
- Atributo: propiedad que codifica la instancia.
- Etiqueta: Número entero que codifica la clase. (Spam=1, No Spam=0)

¹Tomada de Hands on machine learning

Aprendizaje supervisado: Regresión²



Se mantienen los mismos conceptos que en clasificación; sin embargo, ahora la etiqueta corresponde a un valor real.

²Tomada de Hands on machine learning

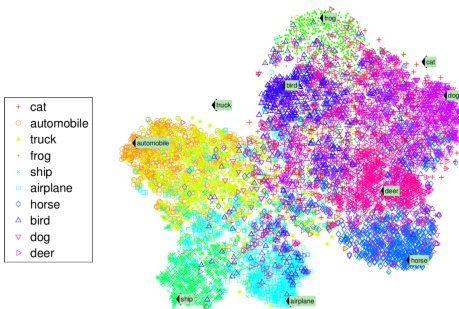
Aprendizaje No supervisado: Agrupamiento ³



Se buscan grupos a partir de las relaciones entre las instancias (Regularidades entre los datos). **Segmentación de mercado, Análisis de redes sociales.**

³Tomada de <https://analyticsjobs.in/blog/clustering-in-machine-learning/>

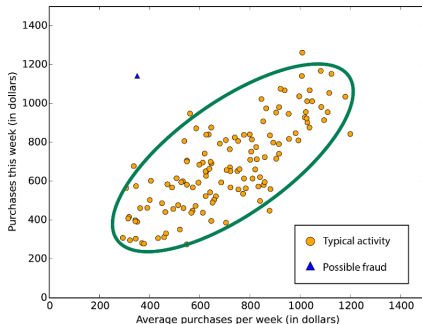
Aprendizaje No supervisado: Reducción de la dimensión ⁴



Se busca representar los datos en un espacio de baja dimensión que represente los datos originales (Compresión)

⁴Tomada de Hands on machine learning

Aprendizaje No supervisado: Detección de anomalías⁵

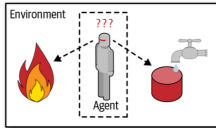


¿La nueva muestra/instancia sigue el comportamiento de los datos de entrenamiento?

⁵Tomada de

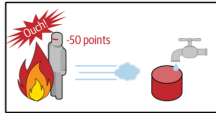
https://www.researchgate.net/publication/321682378_Pattern_Recognition_Anomaly_Detection_Challenges

Aprendizaje por refuerzo.⁶



1 Observe

2 Select action using policy



3 Action!

4 Get reward or penalty



5 Update policy (learning step)

6 Iterate until an optimal policy is found

- El sistema (agente) observa el entorno.
- Toma decisiones obteniendo recompensas o penalizaciones.
- Modifica su estrategia para maximizar las recompensas.
- Ver ejemplo 1. **Mario Bross**
- Ver ejemplo 2. **Robótica**

⁶Tomada de Tomada de Hands on machine learning

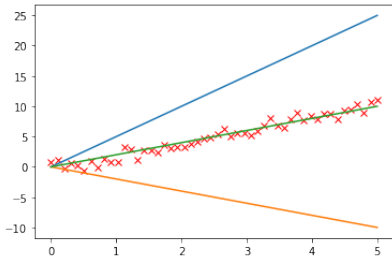


Table of Contents

- ▶ Introducción
- ▶ Tareas de aprendizaje
- ▶ Aspectos prácticos del aprendizaje de máquina

El aprendizaje como una búsqueda I

- Una vez se tienen los datos y una representación (modelo), el aprendizaje se convierte en un problema de búsqueda.
- Se busca la configuración de la representación que mejor describe los datos



El aprendizaje como una búsqueda II

- Los espacios de búsqueda suelen ser considerablemente grandes, por lo cual no se puede usar búsquedas sistemáticas.
- Se suelen usar métodos de optimización basados en gradiente.
- Un algoritmo de entrenamiento hace uso de un espacio de búsqueda, una función de evaluación y un método de búsqueda.

Datos

- Los datos no son perfectos:
 - Los atributos pueden ser inadecuados para la tarea de aprendizaje.
 - Existen instancias con datos faltantes.
 - Algunos atributos tienen valores erróneos.
 - Tener pocos datos conlleva a problemas en el aprendizaje.
- El procesamiento de los datos es fundamental al momento de entrenar un modelo.
El mejor modelo puede fallar sino se tiene un buen procesamiento de datos.

Aprendizaje automático en la práctica

Corresponde a un proceso iterativo que incluye las siguientes etapas:

- Entender el contexto, conocimiento previo y objetivos.
- Integración de los datos, selección, limpieza, pre-procesamiento.
- Aprender los modelos.
- Interpretar resultados.
- Consolidar y desplegar, descubrir el conocimiento.

Cuatro claves del éxito

- Cómputo de alto rendimiento (Algunos gratis: Colab, Kaggle).
- Diferenciación automática. (No más derivadas complicadas)
- Optimización por mini-lotes (Resuelve problemas con muchos datos de forma eficiente).
- Transferencia del conocimiento. Usar modelos pre-entrenados sobre grandes bases de datos. Se ajusta a los datos de nuestro problema.

Preguntas acerca del Video y la clase

Ver el siguiente **Video**

- Cuáles son las principales características del aprendizaje de máquina.
- Cuáles son los pasos para construir un sistemas de aprendizaje de máquina.
- Definir aprendizaje supervisado.
- Definir aprendizaje no supervisado.
- Definir aprendizaje por refuerzo.