

Introducción al Aprendizaje de Máquinas

Jocelyn Dunstan Escudero

jdunstan@uc.cl

Departamento de Ciencia de la Computación
& Instituto de Matemática Computacional
Pontificia Universidad Católica de Chile

Santiago, Chile



- Introducir los conceptos fundamentales del Aprendizaje de Máquinas
- Diferenciar entre aprendizaje supervisado y no supervisado



■ Controles

- Control 7: 14/05
- Control 8: 28/05
- Control 9: 04/06
- Control 10: 11/06
- Control 11: 18/06
- Control 12: 02/07

■ Tareas:

- Tarea 4: 30/05 → 13/06
- Tarea 5: 27/06 → 11/07



Definiciones



¿Qué es aprendizaje de máquinas?

- El aprendizaje de máquinas (o Machine Learning - ML) consiste en programar un computador para que pueda aprender a partir de los datos
- De acuerdo a Arthur Samuel (1959), "Machine Learning es el campo de estudio que da a los computadores la capacidad de aprender sin ser programados explícitamente"
- Se dice que un programa computacional aprende de la experiencia E con respecto a alguna tarea T y alguna medida de rendimiento P , si su rendimiento en T , medido por P , mejora con la experiencia. (E. Tom Mitchell, 1997)
- Machine Learning y su variante más popular "Deep Learning" son las áreas de mayor impacto en la Inteligencia Artificial.



INTELIGENCIA ARTIFICIAL

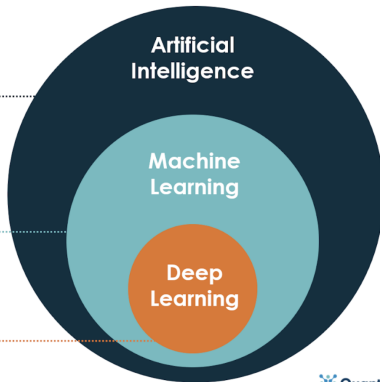
Programas computacionales que hacen tareas que usualmente requerirían inteligencia humana

APRENDIZAJE DE MÁQUINAS

Entrenar algoritmos para resolver tareas de reconocimiento de patrones en vez de programar reglas

APRENDIZAJE PROFUNDO

Entrenar algoritmos que usan redes neuronales profundas

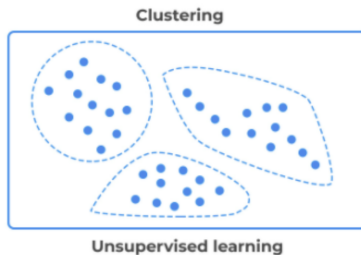
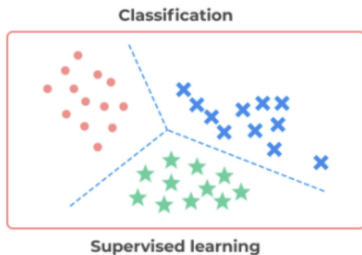


 Quantib

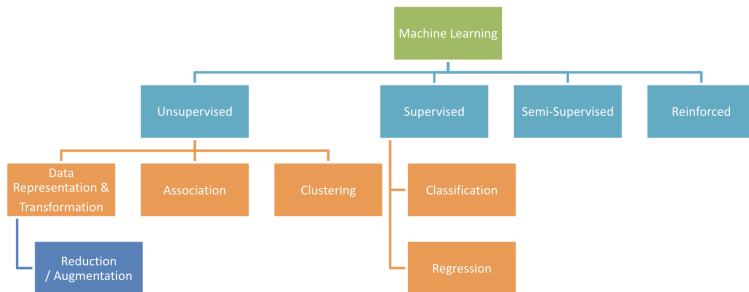
<https://www.quantib.com/the-ultimate-guide-to-ai-in-radiology>



Aprendizaje supervisado vs. no-supervisado



Aprendizaje supervisado vs. no-supervisado



Ejemplo: Detección de spam

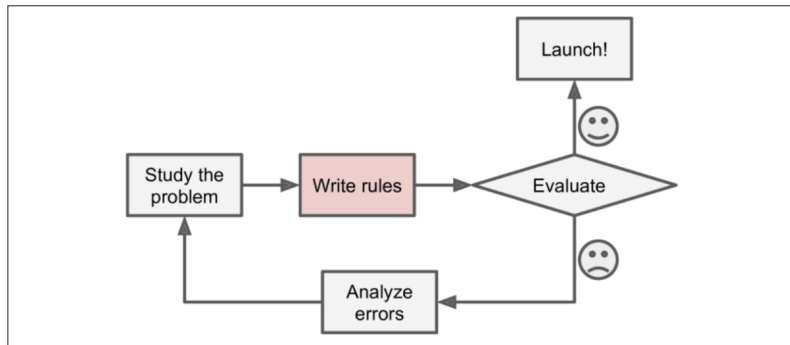


Enfoque tradicional para detectar spam

1. Primero se analizan las propiedades del spam. Aquí nos daremos cuenta que algunas palabras o frases ("lotería", "tarjeta de crédito", "gratis." "viagra") tienden a aparecer mucho en el asunto de los correos spam. Luego se analizan otros patrones en el nombre del remitente, el texto del correo electrónico y así sucesivamente.
2. Se diseña un algoritmo o programa que detecte cada uno de los patrones del paso anterior. El programa clasificará los correos electrónicos como spam cuando se detectan varios de estos patrones.
3. Se prueba el programa, y se repiten los pasos 1 y 2 hasta que sea lo suficientemente bueno.



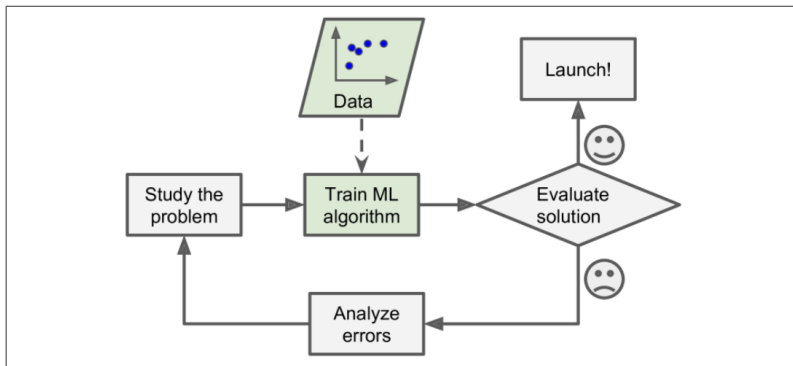
Aprendizaje supervisado vs. no-supervisado



1. Por otro lado, un filtro de spam basado en técnicas de ML aprende automáticamente qué palabras y frases son buenas para predecir spam en un correo.
2. Esto se hace detectando patrones de palabras que ocurren más frecuentemente en ejemplos de correos etiquetados por los usuarios como spam en comparación con ejemplos de correos etiquetados como normales.



Aprendizaje supervisado vs. no-supervisado



Enfoque de aprendizaje de máquinas

1. Supongamos que los spammers se dan cuenta que su correos con la palabra "lotería" están siendo bloqueados.
2. Éstos podrían reemplazar la palabra "lotería" por "premio" es sus nuevos spam.
3. Un filtro de spam basado técnicas de programación tradicionales necesitaría ser actualizado para darse cuenta del cambio.
4. Pero si los spammers siguen refinando constantemente sus correos, tendríamos que seguir escribiendo nuevas reglas para siempre.
5. En cambio, un filtro de spam basado en técnicas de Machine Learning puede seguir aprendiendo sobre las nuevas etiquetas de los usuarios de correo y adaptarse automáticamente a que la palabra "premio" se usa en el spam.



Machine learning vs. Deep learning



1. Antes del 2014 el estado del arte en análisis de texto era el uso de ingeniería de características + un modelo *shallow* de ML.
2. Construir las características requiere. conocimientos específicos del dominio.
3. Por otro lado, los sistemas de aprendizaje profundo se basan en redes neuronales para aprender automáticamente buenas representaciones.



Y hasta el 2019 creíamos que representation learning era el estado del arte... hasta la irrupción de los modelos generativos.

Pero eso es otro curso!



- Introducir los conceptos fundamentales del Aprendizaje de Máquinas
- Diferenciar entre aprendizaje supervisado y no supervisado

