

Machine Learning

Introducción

Dr. Mauricio Toledo-Acosta

Diplomado Ciencia de Datos con Python

Table of Contents

1 Introducción

- Inteligencia Artificial
- Machine Learning

2 Componentes del Machine Learning

3 Ciencia de Datos

¿Qué es la Inteligencia Artificial?

La **Inteligencia Artificial (IA)** es el conjunto de sistemas o algoritmos, cuyo propósito es crear máquinas que imiten la inteligencia humana para realizar tareas y pueden mejorar conforme la información que recopilan.

“... la capacidad de un sistema para interpretar correctamente datos externos, y así aprender y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible”

A. Kaplan, M. Haenlein, 2019.

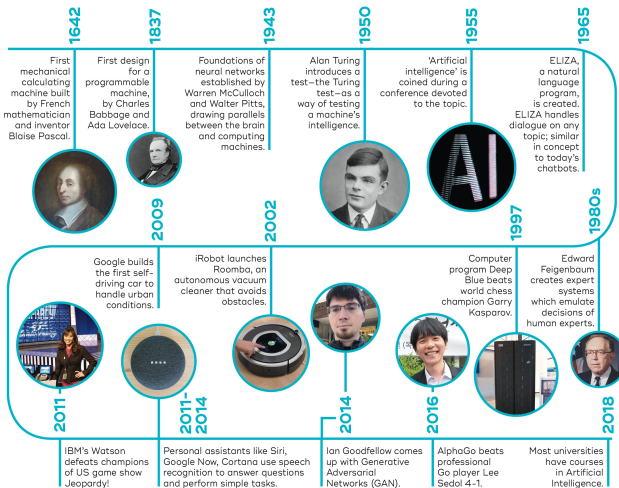
Las definiciones dependen de relaciones con la inteligencia y cognición humana: *aprender, interpretar, aprender,*

Tipos de Inteligencia Artificial

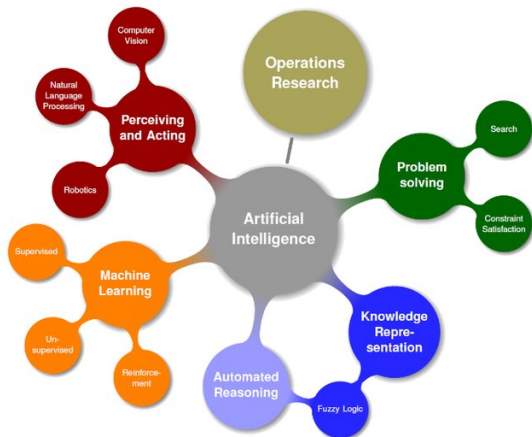
Clasificación de acuerdo a Russell & Norvig, *Artificial Intelligence: A modern approach*, (2002).

- **Los sistemas que piensan como humanos:** Estos sistemas tratan de emular el pensamiento humano automatizando actividades como la toma de decisiones, resolución de problemas y aprendizaje.
- **Los sistemas que actúan como humanos:** Estos sistemas tratan de actuar como humanos; es decir, imitan el comportamiento humano.

Evolución de la IA

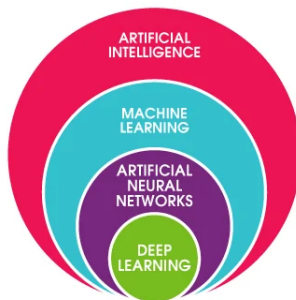


Algunas áreas de la IA



doi:10.13140/RG.2.2.23097.80485

Algunas áreas de la IA



¿Qué es el Machine Learning?

Machine Learning

El **Machine Learning** es una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan con el objetivo de resolver ciertas tareas.

¿Qué es el Machine Learning?

Machine Learning

El **Machine Learning** es una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan con el objetivo de resolver ciertas tareas.

Se dice que un programa de computadora aprende de la Experiencia E con respecto a alguna Tarea T y una medida de desempeño P , si su desempeño sobre T , medido por P , mejora con la experiencia E .

Machine Learning vs Algoritmos tradicionales

Un algoritmo tradicional toma una entrada y una lógica en forma de código y genera una salida para resolver un problema.

Machine Learning vs Algoritmos tradicionales

Un algoritmo tradicional toma una entrada y una lógica en forma de código y genera una salida para resolver un problema.

Por el contrario, un algoritmo de aprendizaje automático toma una entrada y una salida y aprende una lógica que puede utilizarse para trabajar con nuevas entradas y obtener una salida. Esta lógica se obtiene a partir de los patrones presentes en los datos.

Machine Learning vs Algoritmos tradicionales

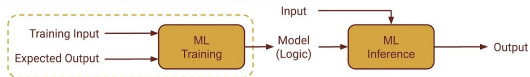
Un algoritmo tradicional toma una entrada y una lógica en forma de código y genera una salida para resolver un problema.

Por el contrario, un algoritmo de aprendizaje automático toma una entrada y una salida y aprende una lógica que puede utilizarse para trabajar con nuevas entradas y obtener una salida. Esta lógica se obtiene a partir de los patrones presentes en los datos.

Traditional Programs: Define algo/logic to compute output



Machine Learning: Learn model/logic from data



<https://www.linkedin.com/pulse/machine-learning-vs-traditional-software-development-ml4devs-gupta>

Fases de un programa de Machine Learning

Los programas de aprendizaje automático tienen dos fases distintas:

- 1 **Entrenamiento:** La entrada y la salida esperada se utilizan para entrenar y probar varios modelos. Se selecciona el modelo más adecuado.
- 2 **Inferencia o predicción:** El modelo se aplica a nuevos datos de entrada para predecir nuevas salidas

¿Por qué necesitamos el Machine Learning?

En el enfoque clásico, antes del ML, se usaban algoritmos que procesaban los datos con base en reglas lógicas (if, else, ...).

¿Por qué necesitamos el Machine Learning?

En el enfoque clásico, antes del ML, se usaban algoritmos que procesaban los datos con base en reglas lógicas (if, else, ...).

- La lógica para tomar las decisiones es específica de acuerdo al dominio y a la tarea. Pequeños cambios en la tarea requieren rediseñar el sistema.
- El diseño de las reglas requiere un entendimiento profundo del dominio por parte de un experto. Estas reglas pueden ser muy complicadas.

¿Por qué necesitamos el Machine Learning?

En el enfoque clásico, antes del ML, se usaban algoritmos que procesaban los datos con base en reglas lógicas (if, else, ...).

- La lógica para tomar las decisiones es específica de acuerdo al dominio y a la tarea. Pequeños cambios en la tarea requieren rediseñar el sistema.
- El diseño de las reglas requiere un entendimiento profundo del dominio por parte de un experto. Estas reglas pueden ser muy complicadas.

Ejemplos: Detectar correo SPAM, reconocer caras en una imagen.

Un ejemplo: Detección de SPAM

Enfoque Tradicional

Enfoque Machine Learning

¿Qué tareas puede resolver el Machine Learning?

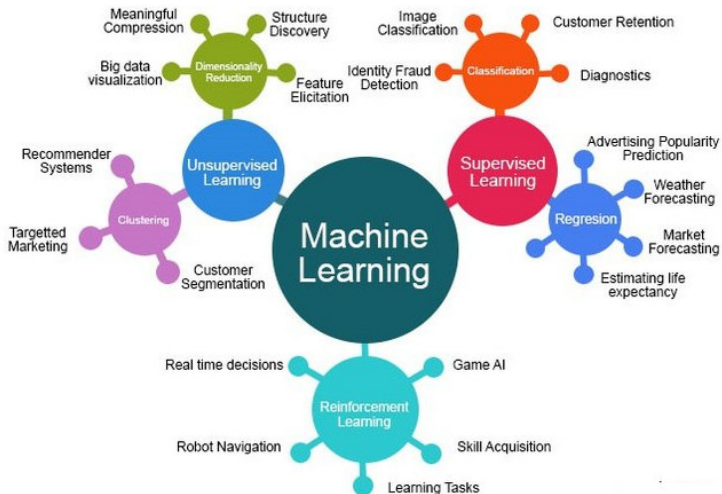
Algunas de las tareas que pueden resolver los métodos de Machine Learning son

- 1 Regresión.
- 2 Clasificación.
- 3 Clustering.
- 4 Traducción automática.
- 5 Detección de anomalías.
- 6 Síntesis y muestreo.
- 7 Estimación de la densidad de probabilidad y de la función de masa de probabilidad.

Los tres paradigmas del Machine Learning

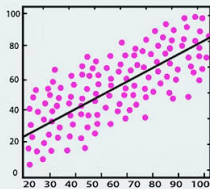
- ➊ Aprendizaje supervisado (Supervised Learning). El modelo aprende de ejemplos etiquetados para hacer predicciones de nuevos datos.
- ➋ Aprendizaje no supervisado (Unsupervised Learning). El modelo encuentra patrones o estructuras intrínsecas en los datos sin etiquetar.
- ➌ Aprendizaje por refuerzo (Reinforcement Learning). El modelo aprende las acciones a tomar en un entorno para maximizar una noción de recompensa acumulativa.

Los tres paradigmas del Machine Learning



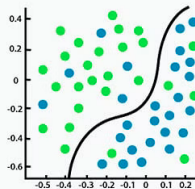
Aprendizaje Supervisado

- 1 Regresión.
- 2 Clasificación.



Regression

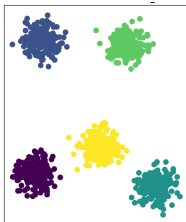
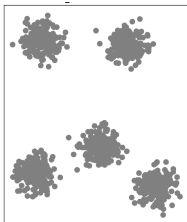
versus



Classification

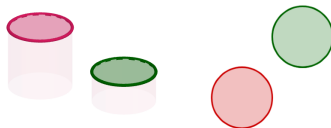
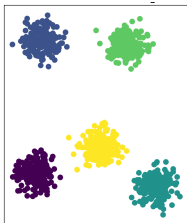
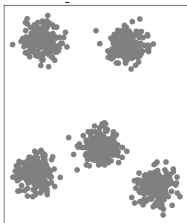
Aprendizaje No Supervisado

1 Clustering.



Aprendizaje No Supervisado

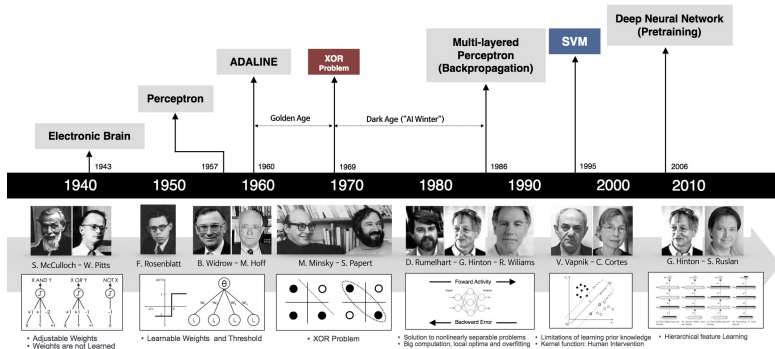
- 1 Clustering.
- 2 Reducción de dimensionalidad.



Aprendizaje por Refuerzo

- 1 Navegación en vehículos autónomos.
- 2 Texto predictivo.
- 3 Sistemas de recomendación.
- 4 Videojuegos.
- 5 Conservación y eficiencia energética.

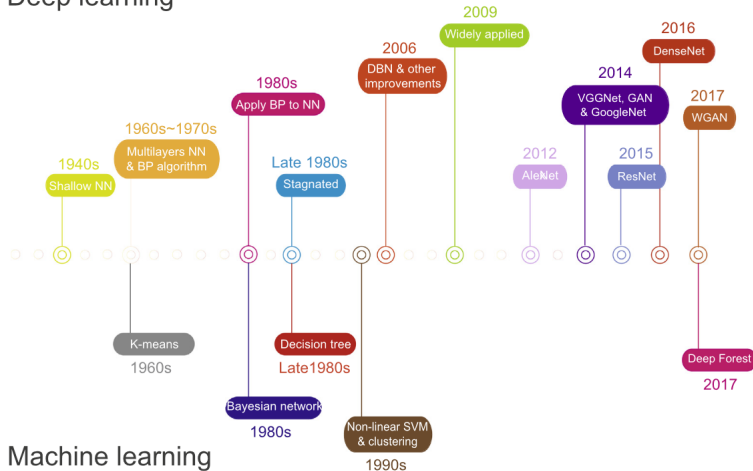
Timeline de los algoritmos de Machine Learning



<http://beamlab.org/deeplearning/2017/>

Timeline de los algoritmos de Machine Learning

Deep learning



Machine learning

Referencias

- Müller, A. C., & Guido, S., 2016. *Introduction to Machine Learning with Python: a Guide for Data Scientists*. O'Reilly Media, Inc..
- Flach, P. A., 2012. *Machine Learning : the Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press.

Table of Contents

- 1 Introducción
 - Inteligencia Artificial
 - Machine Learning
- 2 Componentes del Machine Learning
- 3 Ciencia de Datos

Componentes del Machine Learning

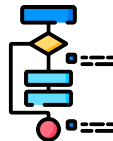
Datos



Variables (features)



Algoritmos



Datos

Los datos pueden tener muchas formas diferentes:

- Tablas estructuradas.
- Imágenes.
- Texto.
- Archivos de audio.
- Archivos de video.

A un conjunto de datos, se le llama **dataset**. Kaggle tiene una colección grande, al igual que Scikit-Learn.

Algunos datasets famosos

- MNIST
- Iris Flowers Dataset.
- Boston House Price Dataset.
- Wine Quality Dataset.
- Pima Indians Diabetes Dataset.
- 20newsgroups.

Features

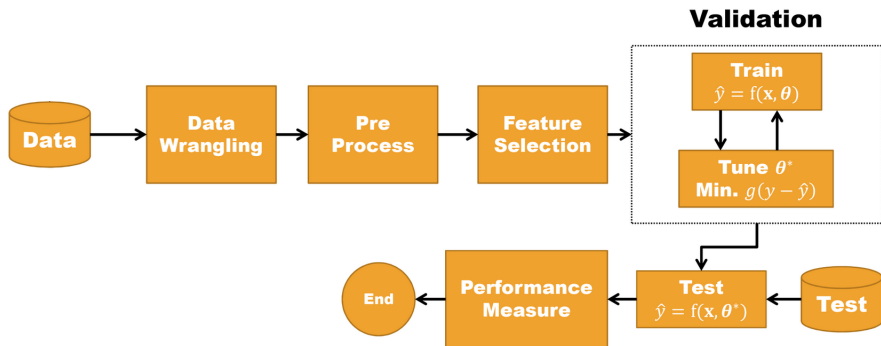
Hay varios tipos de variables

- Numéricas
 - Continuas
 - Discretas
- Categóricas
 - Ordinales: escalas numéricas.
 - Nominales: género, medio de transporte, tipos.

Algoritmos

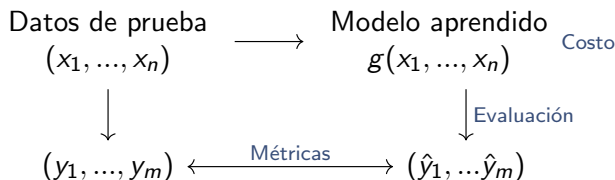


Componentes del Machine Learning



El aprendizaje en el caso supervisado

Dadas variables independientes x_1, \dots, x_n y variables dependientes y_1, \dots, y_m .
El problema consiste en encontrar una *mejor* función (modelo) g tal que $g(x_1, \dots, x_n) = (y_1, \dots, y_m)$.



Métricas de desempeño

Las **métricas de desempeño** varían de acuerdo al tipo de tarea, su función es dar cuenta del desempeño del modelo entrenado. Suelen ser funciones *fácilmente* interpretables.

- Regresión: MSE, MAE.
- Clasificación: Accuracy, precision, recall, F1-score, ROC-AUC.
- Clustering: AMI, MI, silhouette score.
- Métricas NLP: Perplexity, entropy, coherence.
- ...

Funciones de costo

Una **función de pérdida**, o **función de costo**, es una función que asigna un evento o los valores de una o más variables a un número real que representa intuitivamente algún *costo* asociado al evento. Un problema de optimización trata de minimizar una función de pérdida.

Un algoritmo de Machine Learning busca minimizar o maximizar esta función cambiando sus **parámetros internos**. Frecuentemente se usa el **descenso de gradiente** para este fin, por lo tanto, típicamente se requiere de una función de costo diferenciable o convexa.

- Regresión: MSE, RMSE, MAE.
- Clasificación: 0-1, binaria asétrica, entropía cruzada, Hinge loss.

Diferencia entre función de costo y métrica de desempeño

Típicamente son funciones diferentes, bajo ciertas condiciones se puede usar la misma.

- Usando la función de costo como métrica de desempeño: puede ser confusa de interpretar.
- Usando la métrica de desempeño como función de costo: puede no ser posible si no es diferenciable o convexa.

¿Por qué Python?

Ventajas

- Rápido de aprender.
- El código es claro y fácil de leer.
- Desarrollo rápido de modelos.
- Lenguaje orientado a objetos.
- Muchas librerías: Numpy, Pandas, Scipy, Matplotlib.

Machine Learning



Deep Learning



Table of Contents

- 1 Introducción
 - Inteligencia Artificial
 - Machine Learning
- 2 Componentes del Machine Learning
- 3 Ciencia de Datos

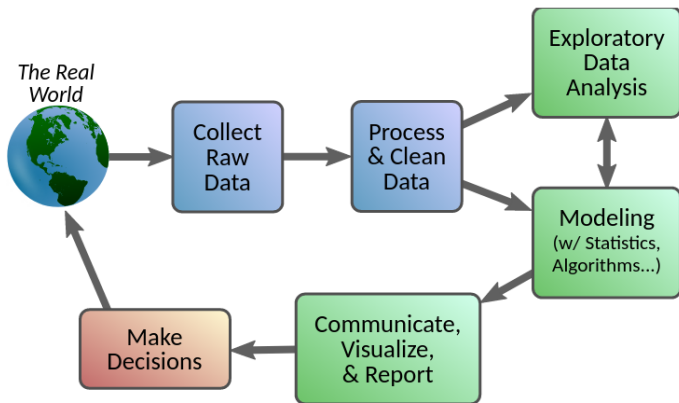
Ciencia de Datos

Ciencia de datos

La **ciencia de datos** es una disciplina que analiza grandes cantidades de datos para extraer información y patrones que sea útiles en la creación de estrategias que permitan aumentar la eficiencia, reconocer nuevas oportunidades de mercado y aumentar la ventaja competitiva de una organización.

La ciencia de datos emplea las disciplinas de las matemáticas, estadística y las ciencias de la computación. Además, incorpora técnicas del Machine Learning, la minería de datos y la visualización, entre otras.

El proceso de la ciencia de datos



<https://snakebear.science/01-Introduction/WhatIsDS.html>