

MACHINE LEARNING

Ph.D Ana Rocío
Cárdenas Maita

 **TaReC Da**


PhawAI

CONTENIDO

- Introducción: Intuición detrás del aprendizaje de máquina
- Preparación de datos
- Aprendizaje supervisado
- Algoritmo de regresión lineal
- Aprendizaje no supervisado
- Algoritmo de k-means

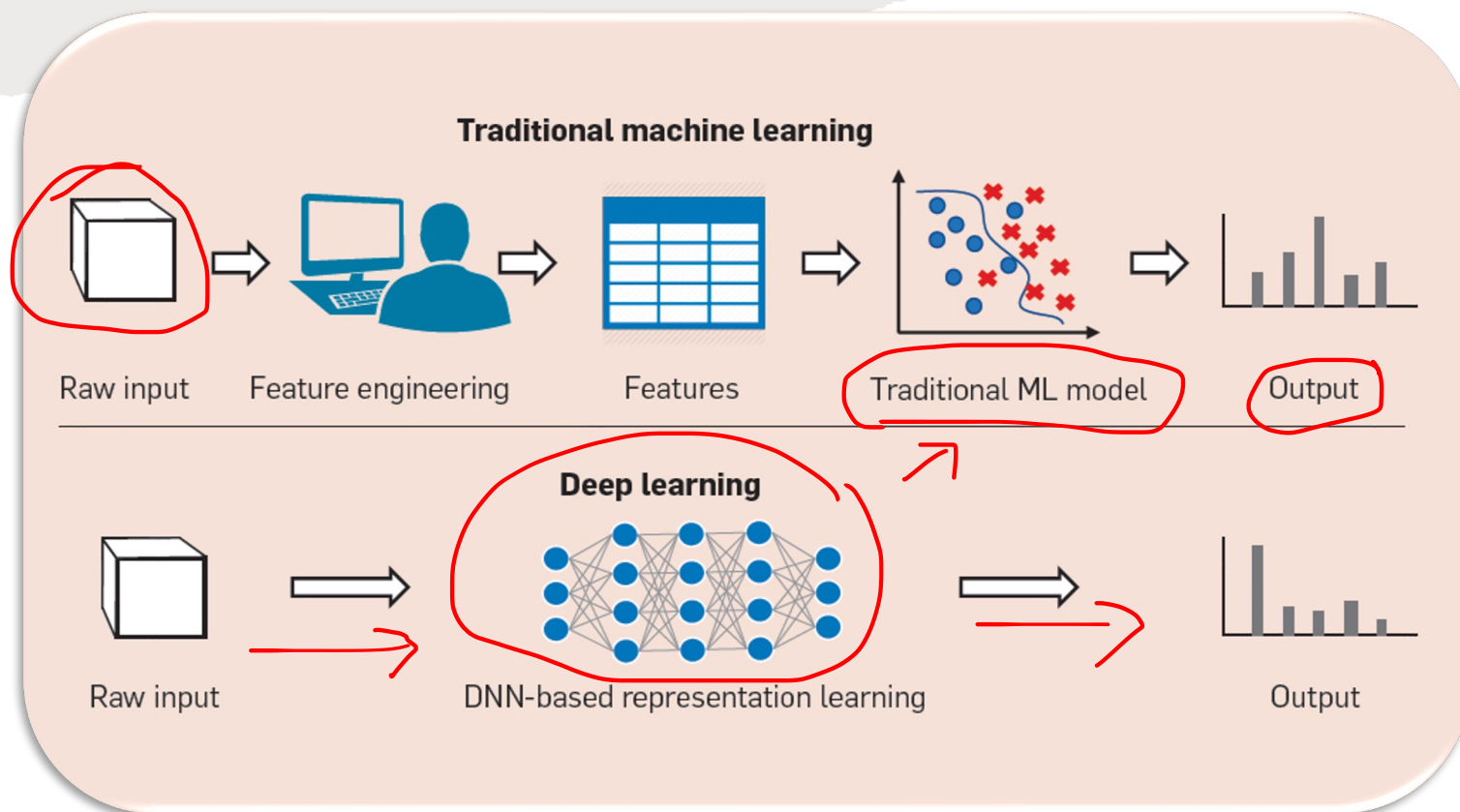
OBJETIVO

Aprender la esencia del aprendizaje de máquina.
Aprender los conceptos clave para que el alumno
tenga la capacidad de aprender por cuenta propia
nuevos algoritmos

TRABAJO FINAL

- Implementar un algoritmo de aprendizaje supervisado y no supervisado diferente al usado en clases.
- Entregar estos algoritmos con los mismos datos
- Realizar comparaciones con el resultado que obtuvimos en clases, usar criterios de tiempos, métricas de calidad, error, etc.
- Elaborar un informe discursivo sobre esta comparación.

INTRODUCCIÓN



APRENDIZAJE HUMANO



S. Russel, P. Norvig. Artificial Intelligence: A modern approach. 2003, 2a ed - Capítulo 18, seção 1, página 650.



CÓMO APRENDEMOS LOS CONCEPTOS DEL MUNDO
QUE NOS RODEA?



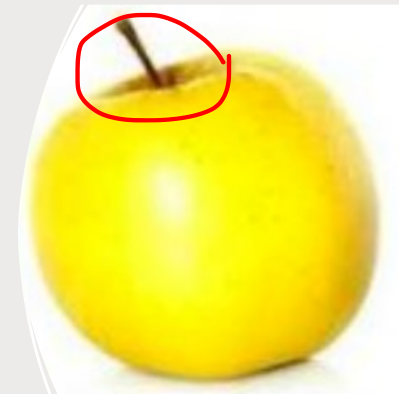
CÓMO APRENDEMOS LOS CONCEPTOS DEL MUNDO
QUE NOS RODEA?



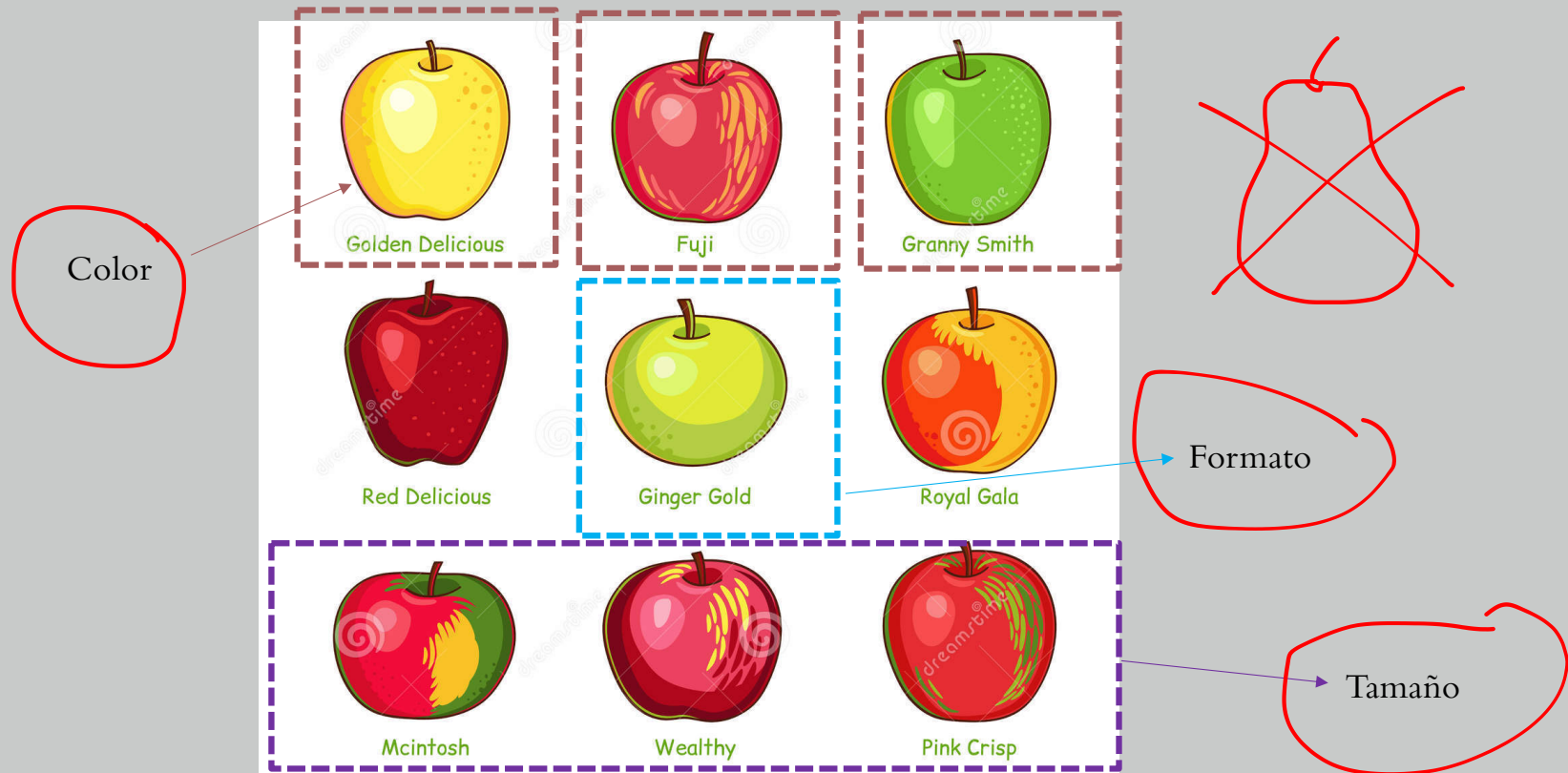
CÓMO APRENDEMOS LOS CONCEPTOS DEL MUNDO
QUE NOS RODEA?



CÓMO APRENDEMOS LOS CONCEPTOS DEL MUNDO
QUE NOS RODEA?



CÓMO APRENDEMOS LOS CONCEPTOS DEL MUNDO
QUE NOS RODEA?



CREAMOS UN **PATRON** MENTAL DE LAS **CARACTERÍSTICAS** QUE REPRESENTAN AL OBJETO OBSERVADO Y QUE GUARDAMOS EN LA MEMORIA

..... Y LOS
ALGORITMOS DE
APRENDIZAJE DE
MÁQUINA?????



DEFINICIÓN DE APRENDIZAJE DE MÁQUINA

“El proceso de aproximar una función a partir de un conjunto de datos descritos por atributos y relacionados con etiquetas.”

- Los atributos (descripciones de los datos) son los valores de entrada de la función.
- Los datos pueden estar relacionados con etiquetas específicas.
- Las etiquetas son los valores de salida de la función.

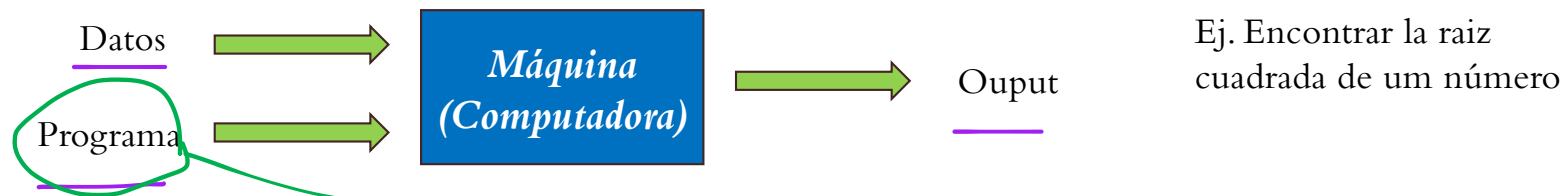
Un algoritmo de aprendizaje automático realiza este proceso revisando los datos y ajustando los parámetros que determinan la función requerida.

- Proceso repetitivo.
- Búsqueda por el error mínimo

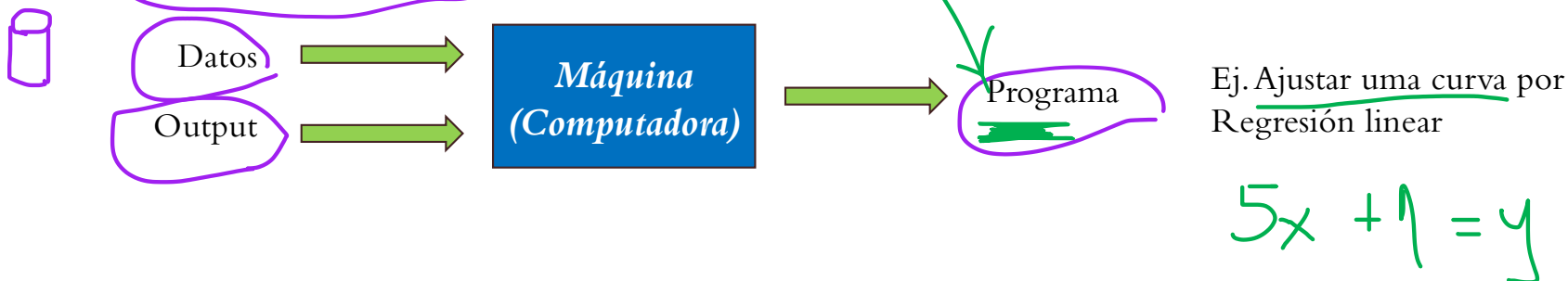


ENTONCES QUÉ ES LO UNA MÁQUINA ESTÁ APRENDIENDO

Programación tradicional



Aprendizaje de máquina



[<https://www.youtube.com/watch?v=h0e2HAPTGF4&t=539s>]

DATASET

- Un conjunto de datos es simplemente el conjunto de datos tabulados en cualquier sistema de almacenamiento de datos estructurado.
- Este conjunto de datos puede haber sido o no sometido a ingeniería de datos previa.
- Cada columna del conjunto de datos representa una variable y cada fila corresponde a los registros de datos que se procesan.

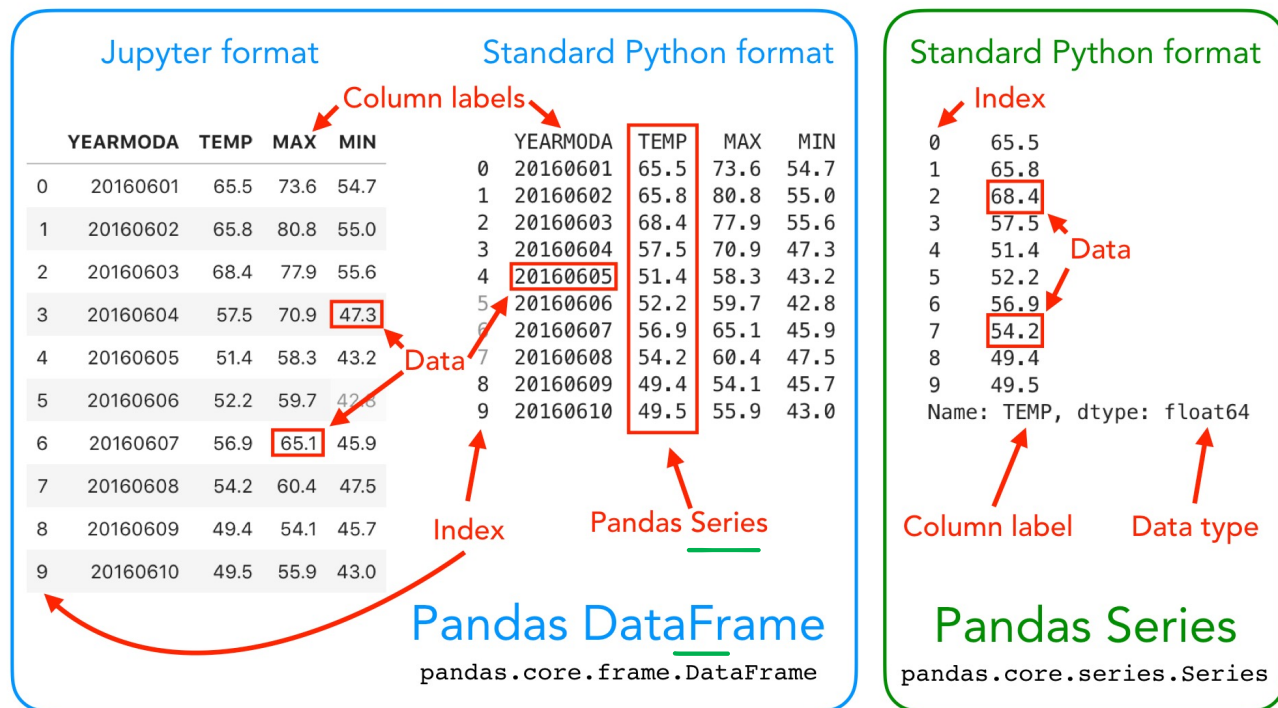
	<i>Color</i>	<i>Tamaño</i>	<i>Formato</i>	<i>Rótulo (Label)</i>
x_1	Rojo	2	alargado	0
x_2	Verde	15	plano	0
x_3	Amarillo	5	arredondado	1

$$D = \begin{pmatrix} \mathbf{A_1} & \mathbf{A_2} & \dots & \mathbf{A_d} & | & \mathbf{C} \\ x_{11} & x_{12} & \dots & x_{1d} & | & c_1 \\ x_{12} & x_{22} & \dots & x_{2d} & | & c_2 \\ \vdots & \vdots & \ddots & \vdots & | & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} & | & c_n \end{pmatrix}$$

[<https://geo-python.github.io/site/notebooks/L5/exploring-data-using-pandas.html>]
[<https://resumos.leic.pt/apre/classification/>]



ESTRUCTURA DE DATOS EN PANDAS PYTHON



[<https://geo-python.github.io/site/notebooks/L5/exploring-data-using-pandas.html>]

PREPARACIÓN DE DATOS

- Exploración de datos

Realizar estadísticas por fila

- Número de NaN
- Número de ceros
- Número de valores negativos
- Promedio, máximo, mínimo, etc.

Redondeo de variables numéricas

Escalado de variables numéricas dentro de un rango determinado

- Reducción de dimensionalidad.

Eliminación de columnas que no aportan información al conjunto de datos.

- Nombres.
- ID de columna.
- Columnas resultantes de otras columnas.

PREPARACIÓN DE DATOS

- Asignación de categorías a números, proceso llamado: Encoding.
- Categorías a números ordenados (LabelEncoder).

<i>Tamaño (Categórico)</i>	<i>Tamaño (numérico)</i>
Pequeño	1
Mediano	2
Grande	3

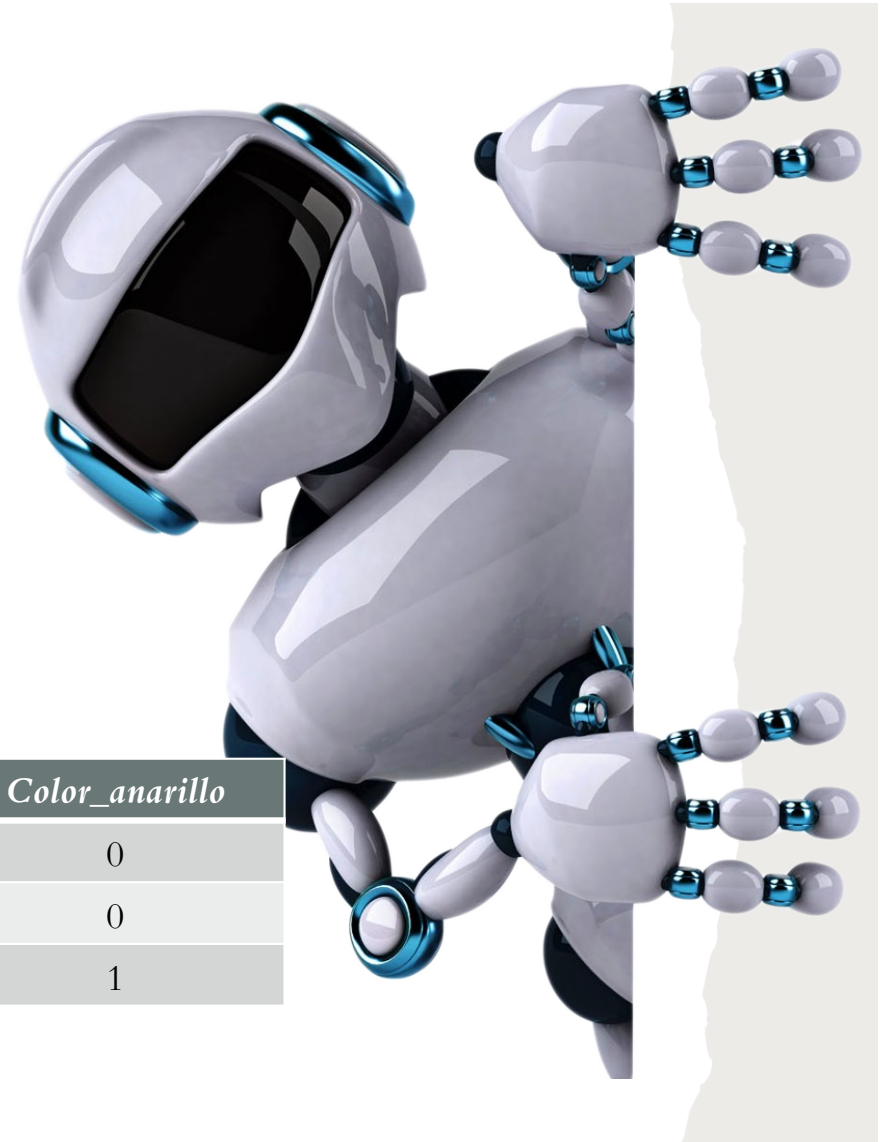
- Categorías a nuevas columnas binarias o dummies (OneHotEncoder).

<i>Color</i>	<i>Color_rojo</i>	<i>Color_verde</i>	<i>Color_anarillo</i>
Rojo	1	0	0
Verde	0	1	0
Amarillo	0	0	1

.... CUÁL ES
MEJOR?????

<i>Tamaño (Categórico)</i>	<i>Tamaño (numérico)</i>
Pequeño	1
Mediano	2
Grande	3

<i>Color</i>	<i>Color_rojo</i>	<i>Color_verde</i>	<i>Color_anarillo</i>
Rojo	1	0	0
Verde	0	1	0
Amarillo	0	0	1



TIPOS DE APRENDIZAJE DE MÁQUINA



TIPOS DE APRENDIZAJE AUTOMÁTICO

Deductive, Inductive, and Abductive Syllogisms

Deductive	Inductive	Abductive
All men are mortal;	Socrates is a man;	All men are mortal;
Socrates is a man;	Socrates is mortal;	Socrates is mortal;
∴ Socrates is mortal.	∴ All men are mortal.	∴ Socrates is a man.

Adapted from: Hui, J., Cashman, T. and T. Deacon. 2008. Bateson's Method: Double Description. What is It? How Does It Work? What Do We Learn? in J. Hoffmeyer (ed.) A Legacy for Living Systems: Gregory Bateson As Precursor to Biosemiotics.



TIPOS DE APRENDIZAJE AUTOMÁTICO

Deductive, Inductive, and Abductive Syllogisms

Deductive	Inductive	Abductive
All men are mortal;	Socrates is a man;	All men are mortal;
Socrates is a man;	Socrates is mortal;	Socrates is mortal;
∴ Socrates is mortal.	∴ All men are mortal.	∴ Socrates is a man.

Aprendizaje no supervisado

and T. Deacon, 2008. Bateson's Method: Double Description. What is It? How Does It Work? What Do we learn? in J. Hoffmeyer (ed.) A Legacy for Living Systems: Gregory Bateson As Precursor to Biosemiotics.



TIPOS DE APRENDIZAJE AUTOMÁTICO

Deductive, Inductive, and Abductive Syllogisms

Deductive	Inductive	Abductive
All men are mortal;	Socrates is a man;	All men are mortal;
Socrates is a man;	Socrates is mortal;	Socrates is mortal;
∴ Socrates is mortal.	∴ All men are mortal.	∴ Socrates is a man.

Adapted from: Hui, J., Cash

Aprendizaje supervisado

scription. What is It? How Does It Work? What Do
ns: Gregory Bateson As Precursor to Biosemiotics.



TIPOS DE APRENDIZAJE AUTOMÁTICO

Deductive, Inductive, and Abductive Syllogisms

Deductive	Inductive	Abductive
All men are mortal;	Socrates is a man;	All men are mortal;
Socrates is a man;	Socrates is mortal;	Socrates is mortal;
∴ Socrates is mortal.	∴ All men are mortal.	∴ Socrates is a man.

Adapted from: Hui, J., Cashman, T. and T. Deacon. 2008. Bateson's Method: Double We Learn? in J. Hoffmeyer (ed.) A Legacy for Living Systems

Aprendizaje por refuerzo





APRENDIZADO SUPERVISIONADO

- El aprendizaje supervisado aprende de datos "etiquetados".
- Esto implica que algunos datos ya están marcados con la respuesta correcta.
- El aprendizaje supervisado se clasifica en dos categorías de algoritmos:

Clasificación: Un problema de clasificación ocurre cuando la variable de salida es una categoría, como "Rojo" o "azul", "enfermedad" o "sin enfermedad".

Regresión: Un problema de regresión es cuando la variable de salida es un valor real, como "dólares" o "peso".





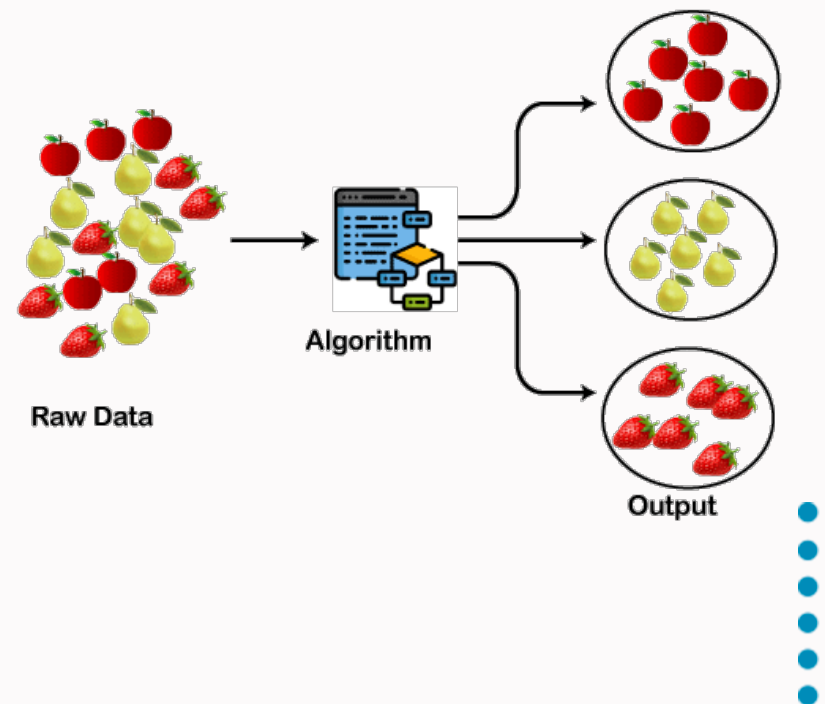
EJEMPLOS DE ALGORITMOS DE APRENDIZAJE SUPERVISADO

- Regresión lineal
- Regresión logística
- KNN - Vecino más cercano
- Bayes ingenuo gaussiano
- Árboles de decisión
- Máquina de vectores de soporte (SVM)
- Flores aleatorias



APRENDIZAJE NO SUPERVISADO

- El aprendizaje no supervisado es el entrenamiento de una máquina utilizando información que no está clasificada ni etiquetada y permitiendo que el algoritmo actúe sobre esa información sin orientación.
- La tarea de la máquina es agrupar información no clasificada de acuerdo con similitudes, patrones y diferencias sin ningún entrenamiento previo de datos.
- Descubrimiento de relaciones existentes entre datos.





APRENDIZAJE NO SUPERVISADO

- Aprendizaje no supervisado
- Se clasifica en dos categorías de algoritmos:
- **Agrupación:** Un problema de agrupación es cuando desea descubrir las agrupaciones inherentes a los datos, como agrupar a los clientes por comportamiento de compra.
- **Asociación:** Un problema de aprendizaje de reglas de asociación es cuando desea descubrir reglas que describan grandes porciones de sus datos, como las personas que compran X también tienden a comprar Y.





EJEMPLOS DE APRENDIZAJE NO SUPERVISADO

- Formas de agrupación

- Único (partición)

- Aglomerativo

- Solapar

- Probabilístico

- Técnicas de agrupación

- Agrupación jerárquica

- Agrupación de K-medias

- Análisis de componentes principales



Parameters	Supervised machine learning	Unsupervised machine learning
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data that is not labeled
Computational Complexity	Simpler method	Computationally complex
Accuracy	Highly accurate	Less accurate
No. of classes	No. of classes is known	No. of classes is not known
Data Analysis	Uses offline analysis	Uses real-time analysis of data
Algorithms used	Linear and Logistics regression, Random forest, Support Vector Machine, Neural Network, etc.	K-Means clustering, Hierarchical clustering, Apriori algorithm, etc.
Output	Desired output is given.	Desired output is not given.
Training data	Use training data to infer model.	No training data is used.
Complex model	It is not possible to learn larger and more complex models than with supervised learning.	It is possible to learn larger and more complex models with unsupervised learning.
Model	We can test our model.	We can not test our model.
Called as	Supervised learning is also called classification.	Unsupervised learning is also called clustering.
Example	Example: Optical character recognition.	Example: Find a face in an image.

COMPARATIVO

[<https://www.geeksforgeeks.org/supervised-unsupervised-learning/>]

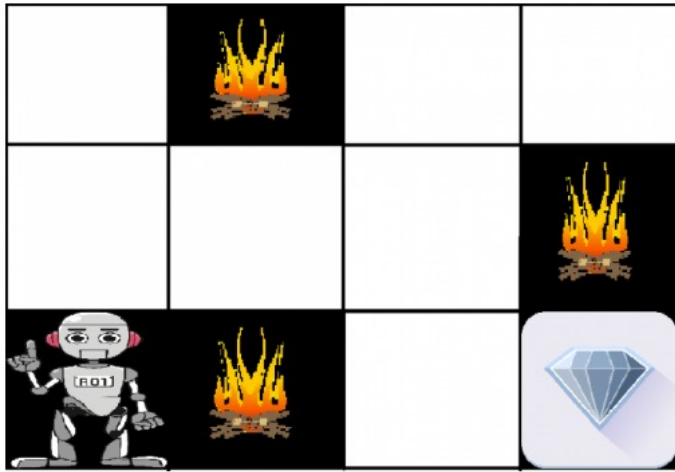


APRENDIZAJE POR REFUERZO O REINFORCEMENT LEARNING

- Es el entrenamiento de modelos de aprendizaje automático para tomar una secuencia de decisiones.
- Se caracteriza porque no hay una referencia -una plantilla- que diga: "esta es la mejor acción a tomar".
- El agente aprende a lograr un objetivo en un entorno incierto y potencialmente complejo.
- El sistema de inteligencia artificial se enfrenta a una situación.
La computadora usa prueba y error para encontrar una solución al problema.
Para que la máquina haga lo que el programador quiere, el sistema recibe recompensas o penalizaciones por las acciones que realiza.
Tu objetivo es maximizar la recompensa total.



EJEMPLO DE APLICACIÓN



- Tenemos un agente y una recompensa, con muchos obstáculos en el medio.
- El agente debe encontrar el mejor camino posible para llegar a la recompensa y cuando se encuentra con un obstáculo, debe ser penalizado, ya que debe elegir el camino sin obstáculos.
- Con el aprendizaje por refuerzo, podemos entrenar al agente para encontrar el mejor camino.

[<https://www.deeplearningbook.com.br/o-que-e-aprendizagem-por-reforco/>]



Google Research



PhawAI

TaReCDa



Universidad Católica
San Pablo

Departamento de Ciencia
de la Computación



RICE



University of
Pittsburgh



UTMACH

GRACIAS