

Primer Coloquio de Divulgación de la Comunidad de Ingeniería en Sistemas

Aprendizaje Automático en Python





**José Clemente Hernández
Hernández**



Gustavo Adolfo Vargas Hákim



COVNNEC - App

Research Group on Computer Vision, Neural Networks,
Evolutionary Computation and their Applications



Agenda

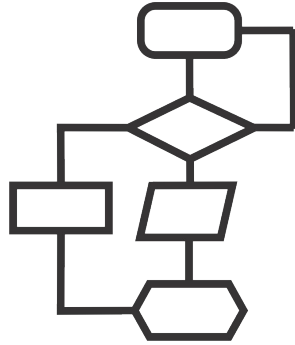
1. ¿Qué es el Aprendizaje Automático?
2. Tipos de Aprendizaje Automático
3. ¿Cuáles son las partes del Aprendizaje Automático?
4. ¿Cuándo utilizar Aprendizaje Automático?
5. Preprocesamiento de datos en Python



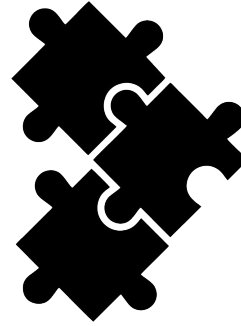
Aprendizaje Automático



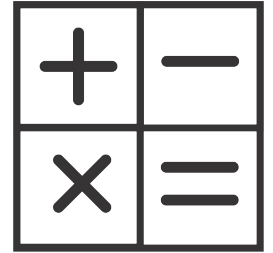
**Ciencias de la
Computación**



**Algoritmos
y técnicas**



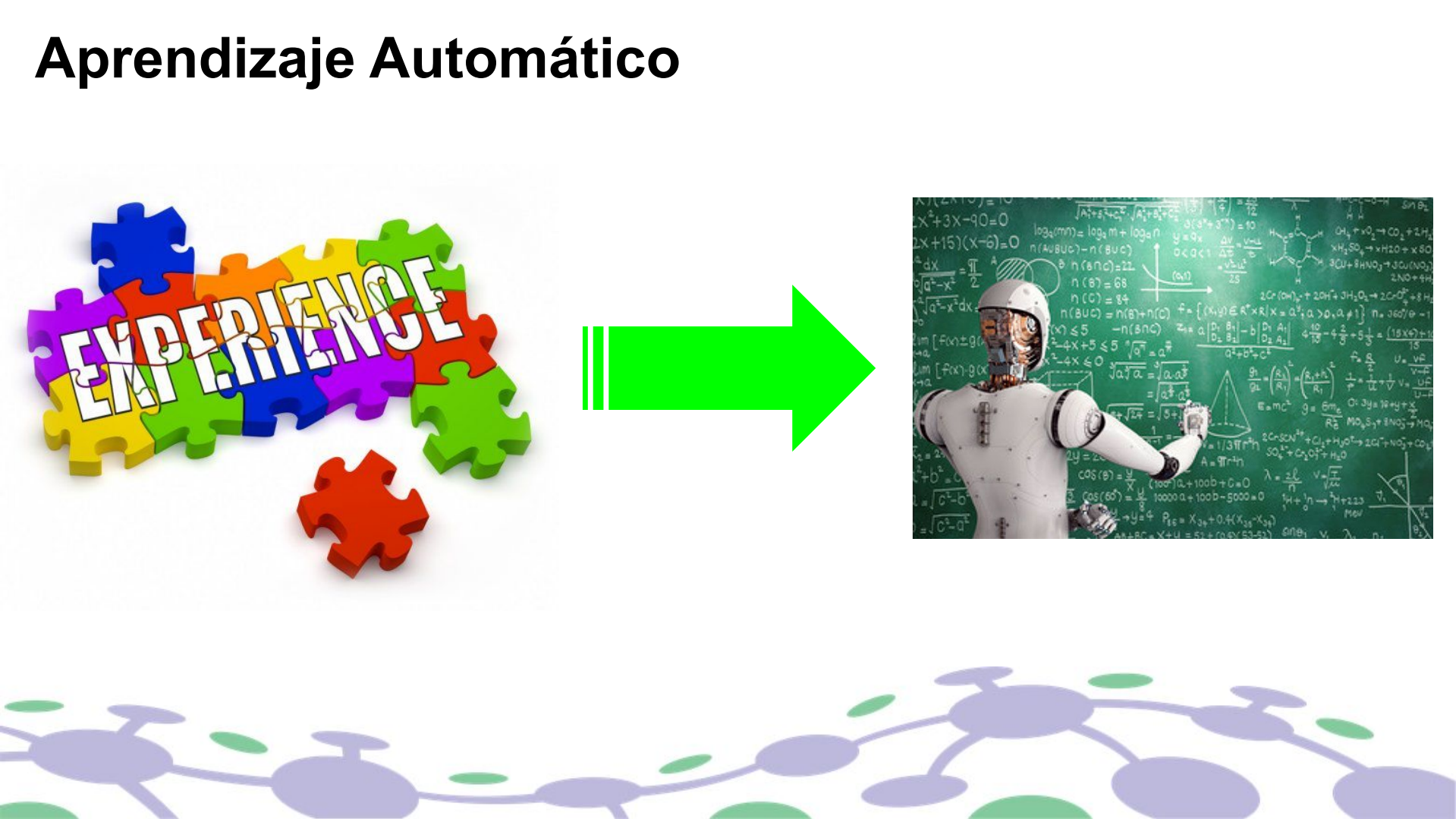
**Resolución de
problemas
complejos**



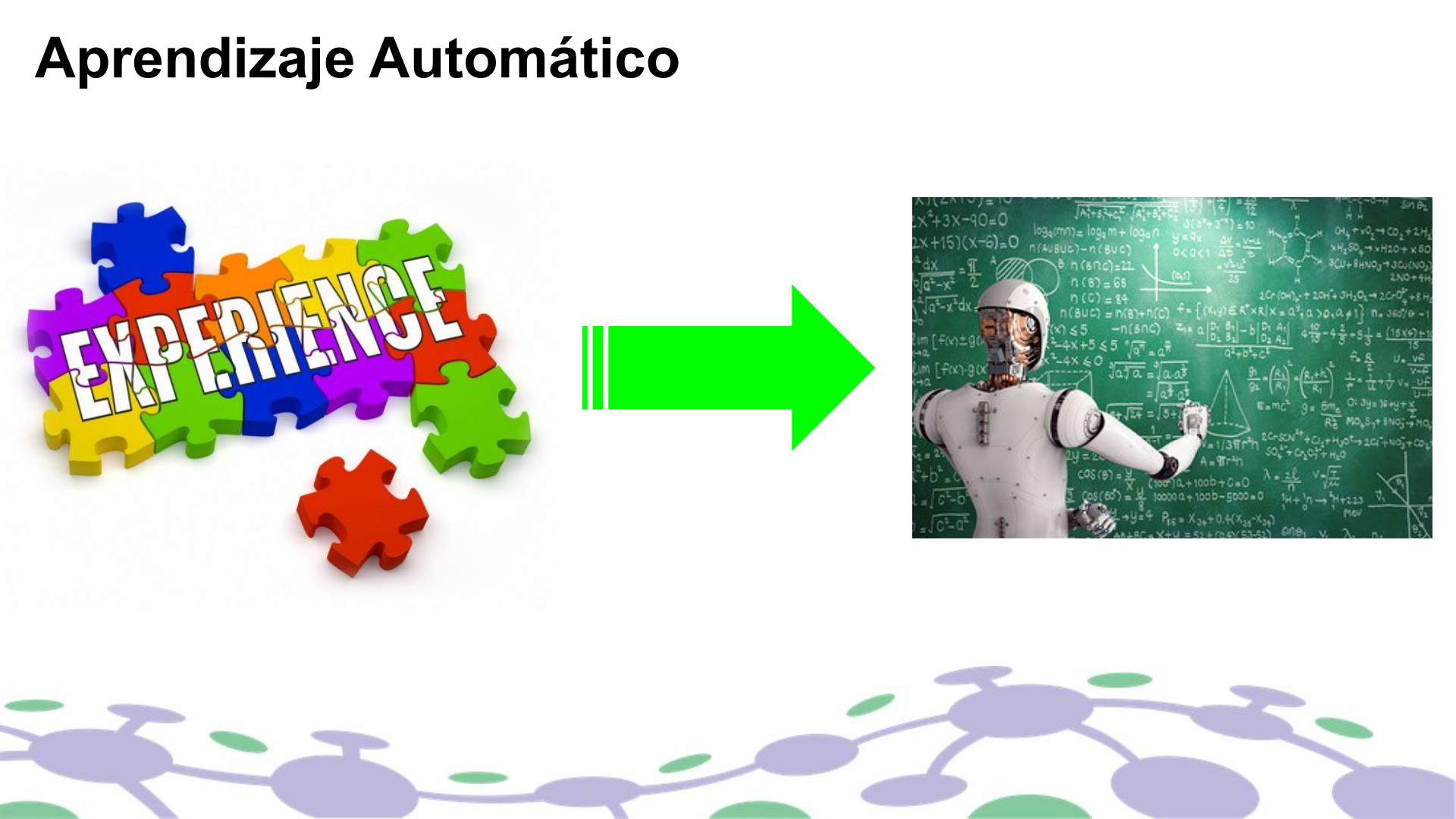
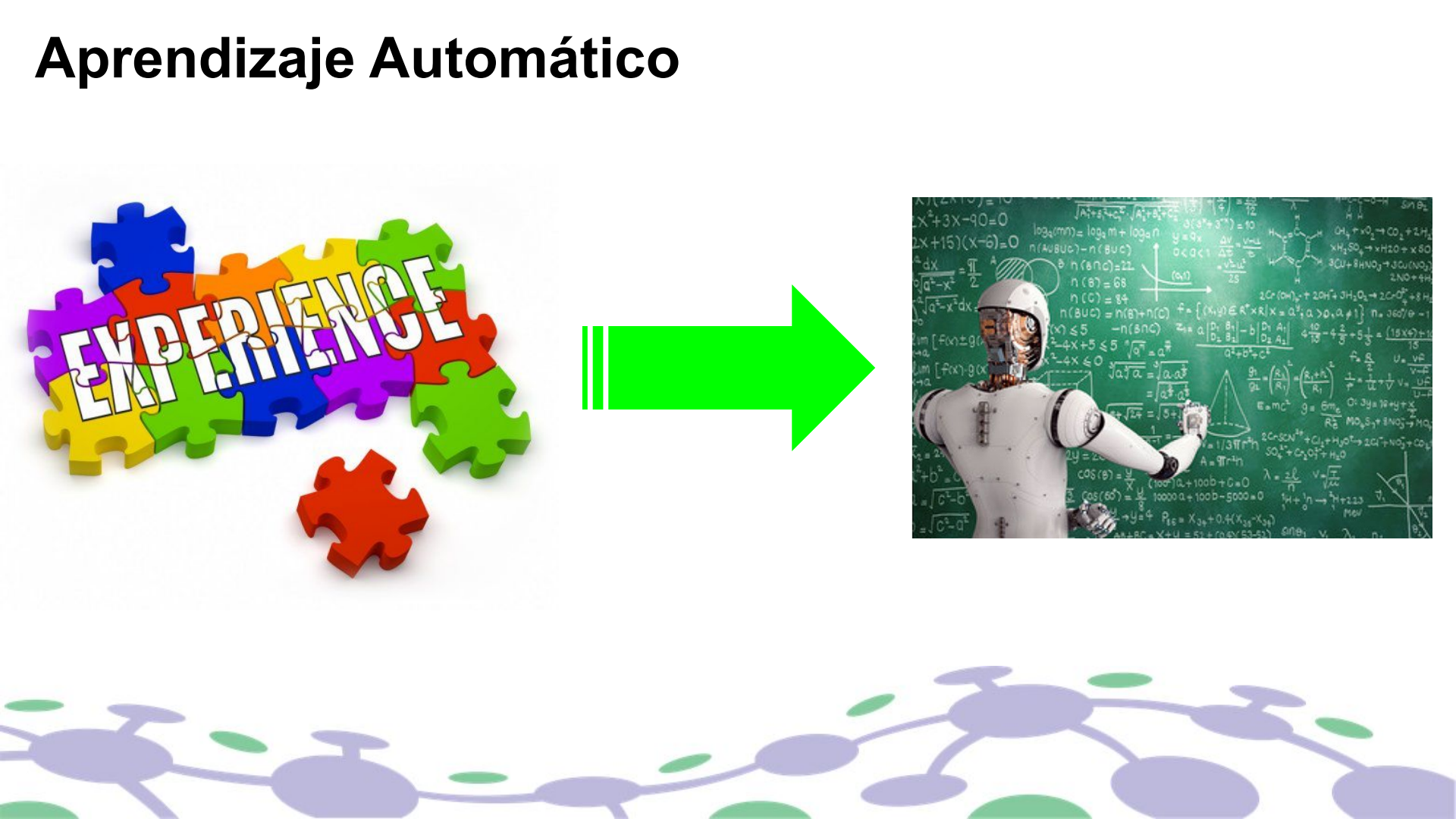
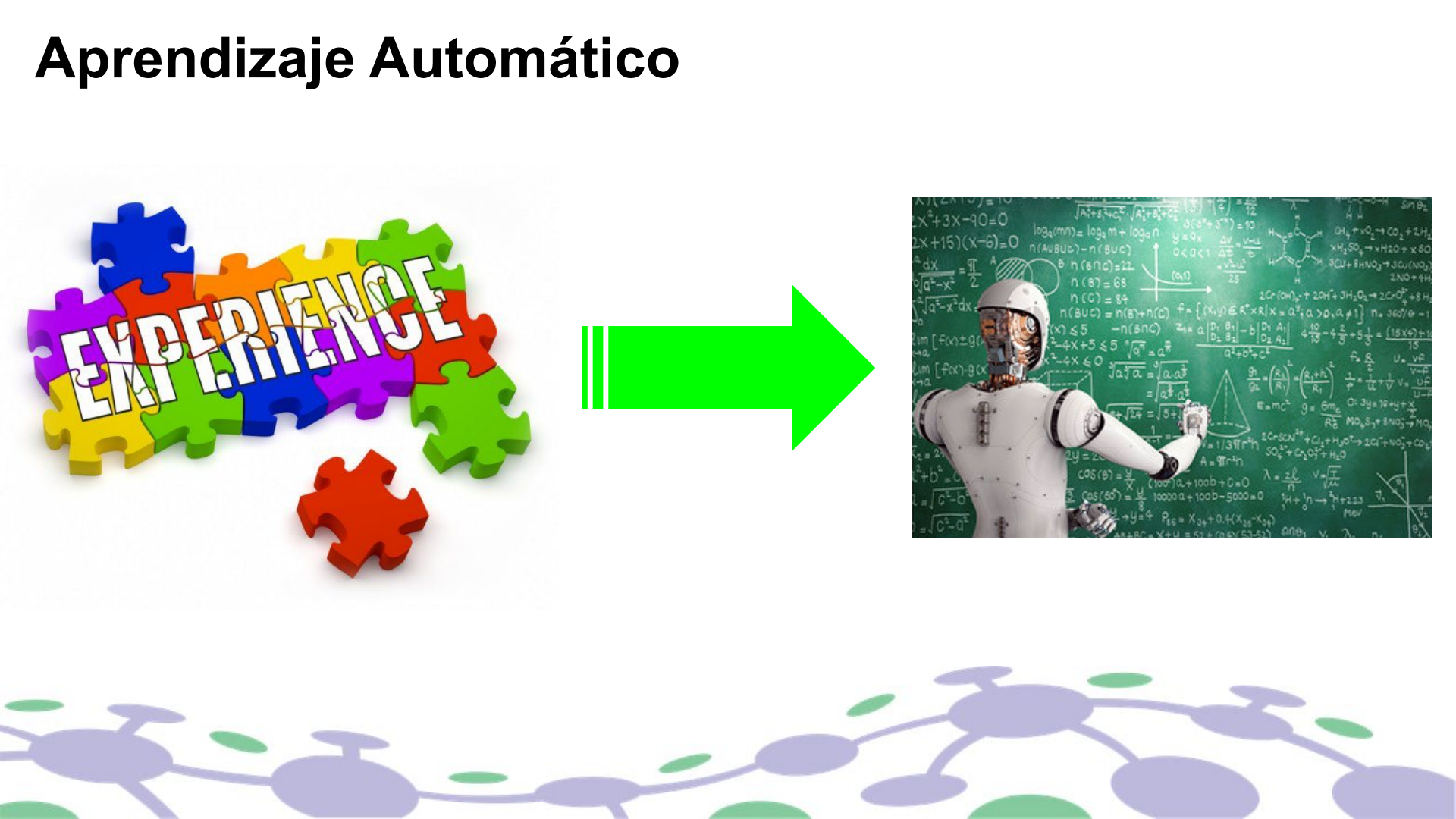
**Matemáticas
y Estadística**



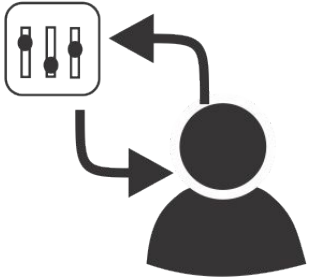
Aprendizaje Automático



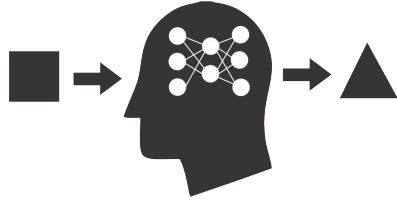
The image is a collage illustrating the components of machine learning. On the left, colorful puzzle pieces form the word "EXPERIENCE". In the center, a large green arrow points right. On the right, a robot stands before a chalkboard filled with mathematical and scientific formulas. At the bottom, a purple network diagram with green nodes is visible.



Tipos de Aprendizaje Automático



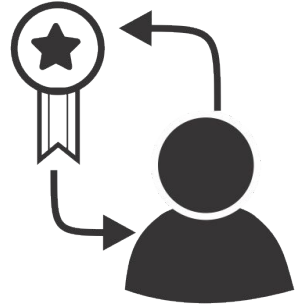
Supervisado



No supervisado



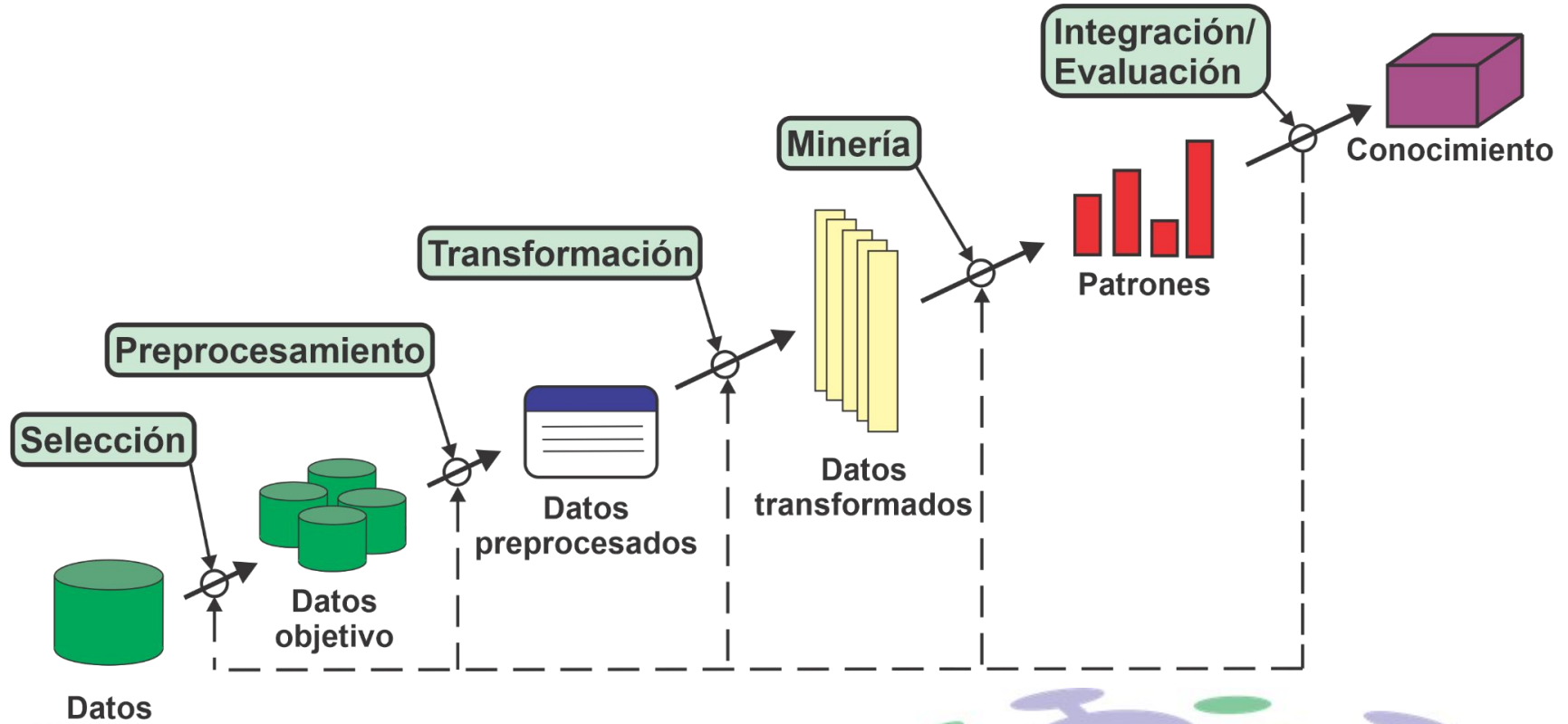
Semi-supervisado



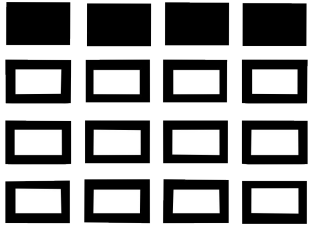
Por refuerzo



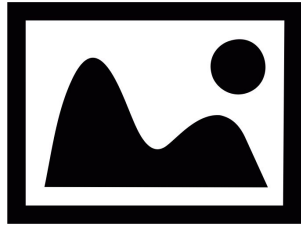
Proceso del Aprendizaje Automático



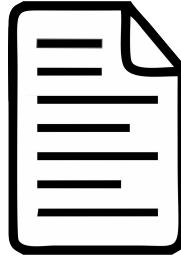
Los datos



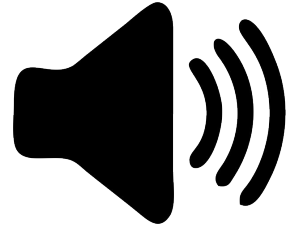
Tablas



Imágenes



Texto



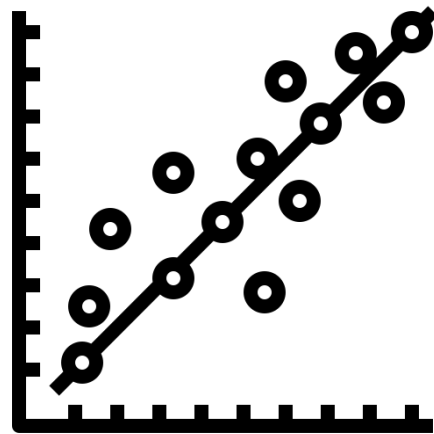
Audio



¿Cuándo **No** utilizar aprendizaje automático?

Un ejemplo muy intuitivo:

Masa	Aceleración	Fuerza
16 Kg	5 m/s ²	80 N
10 Kg	1 m/s ²	10 N
48 Kg	3 m/s ²	144 N
5 Kg	6 m/s ²	x



Regresión lineal



$$F = 29.87 \text{ N}$$

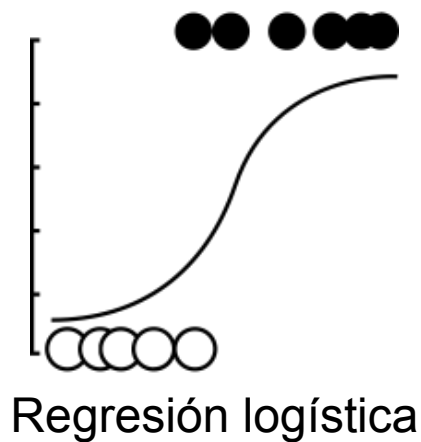
$$F = m \cdot a$$



¿Cuándo **No** utilizar aprendizaje automático?

Otro ejemplo

Sensor 1	Sensor 2	Salida
1	1	1
1	0	0
0	1	0
0	0	x



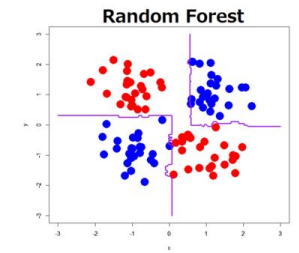
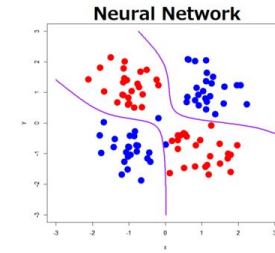
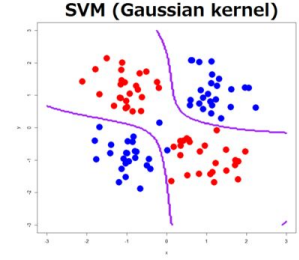
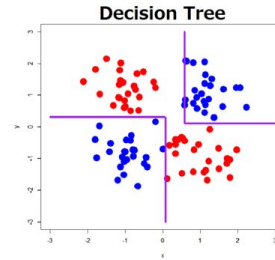
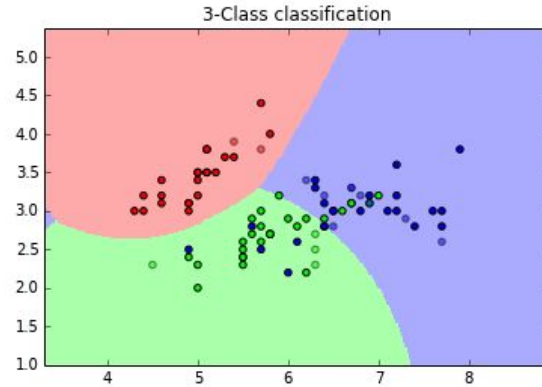
$x = 0$

AND

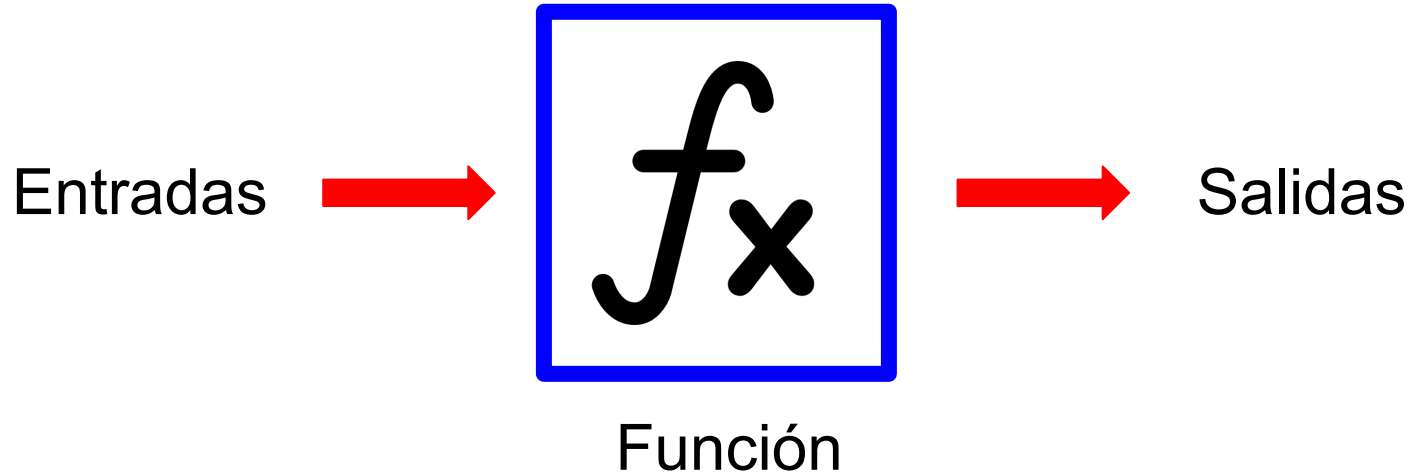
Entonces... ¿cuándo utilizarlo?

 f_x

Aproximador de
funciones



Entonces...¿cuándo utilizarlo?



¿Podemos programar esta función?



Ejemplos interesantes



[i1]



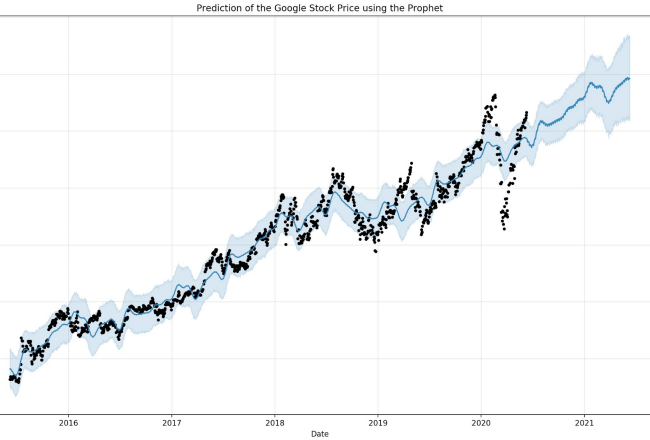
[i2]



[i3]



Ejemplos interesantes



[i4]



[i5]



[i6]



Manejo y preprocesamiento de datos



NumPy



matplotlib





Los datos

Variables/características/descriptores

Ejemplos

X^D



x_1	x_2	x_3	x_4	Etiqueta
0.56	1.98	0.74	1.03	A
0.98	1.42	0.88	0.76	B
...

$$x^2 = (0.98, 1.42, 0.88, 0.76)$$

$$y_2 = B$$

Etiquetas/salidas/clases

y



Manos a la obra



Discretización de variables

x_1	x_2	y
0.7	0.8	1
0.6	0.4	0
0.1	0.9	0
0.2	0.3	0

Variables continuas

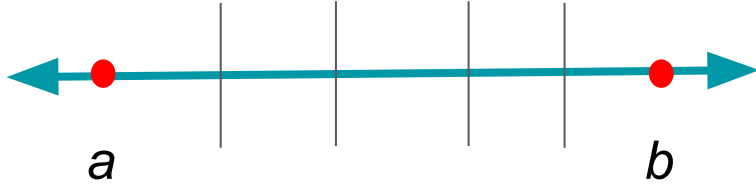


x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	0

Variables discretas



¿Cómo se hace?



Mediante intervalos

***Class-Attribute
interdependence
Maximization***

Algoritmo CAIM



Intervalos con límites conocidos

x_1	x_2	y
0.7	0.8	1
0.6	0.4	0
0.1	0.9	0
0.2	0.3	1

Sujeto a:
 $0 \leq x_1, x_2 \leq 1$

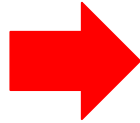


Intervalos con límites conocidos

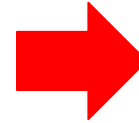
Intervalos a considerar:

$$i_0 = [0, 0.5]$$

$$i_1 = (0.5, 1]$$



x_1	x_2	y
i_1	i_1	1
i_1	i_0	0
i_0	i_1	0
i_0	i_0	0



x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	0

Discretizado



Intervalos con límites desconocidos

x_1	x_2	y
0.7	0.8	1
0.6	0.4	0
0.1	0.9	0
0.2	0.3	1

Sujeto a:
 $? \leq x_1, x_2 \leq ?$



Intervalos con límites desconocidos

Límites a
considerar:

$$0.1 \leq x_1 \leq 0.7$$
$$0.3 \leq x_2 \leq 0.8$$



Intervalos a considerar

Para x_1 :

$$i_0 = [0.1, 0.4]$$
$$i_1 = (0.4, 0.7]$$

Para x_2 :

$$i_0 = [0.3, 0.55]$$
$$i_1 = (0.55, 0.8]$$



Intervalos con límites desconocidos

x_1	x_2	y
0.7	0.8	1
0.6	0.4	0
0.1	0.9	0
0.2	0.3	1



x_1	x_2	y
i_1	i_1	1
i_1	i_0	0
i_0	i_1	0
i_0	i_0	1



x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	1

Discretizado



Algoritmo CAIM

Class-Attribute interdependence Maximization (CAIM)

L. A. Kurgan and K. J. Cios, "CAIM discretization algorithm," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2, pp. 145-153, Feb. 2004, doi: 10.1109/TKDE.2004.1269594.



Aumentación de datos

Los datos son nuestra materia prima.

x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	0

2 variables binarias



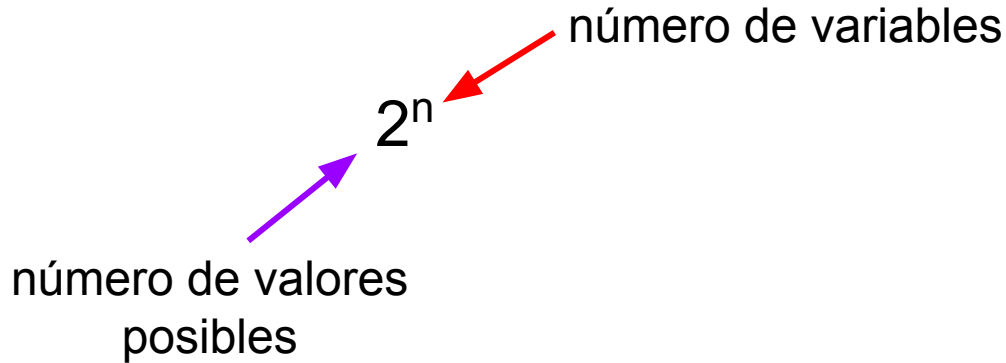
x_1	x_2	x_3	y
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

3 variables binarias



Aumentación de datos

El número de ejemplos necesarios para representar todos los casos crece de forma exponencial!



A diagram illustrating the exponential growth of data points. The expression 2^n is centered. A red arrow points from the text "número de variables" to the variable n in the expression. A purple arrow points from the text "número de valores posibles" to the base 2 of the expression.

$$2^n$$

número de variables

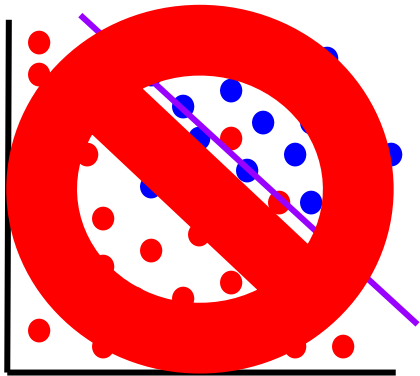
número de valores posibles

Esto se conoce como la **maldición de la dimensionalidad**.

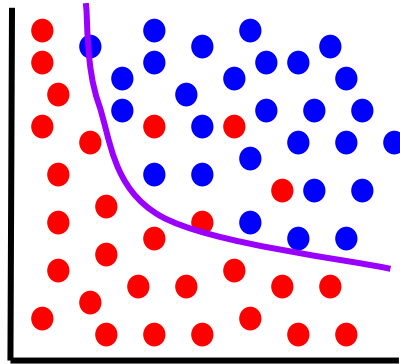


Aumentación de datos

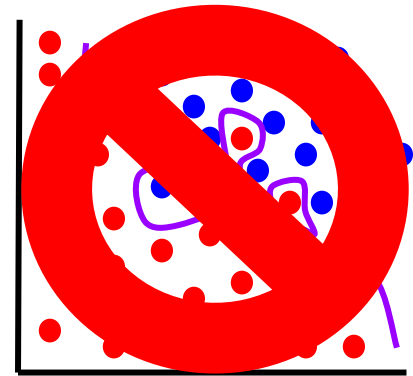
¿Qué pasa si no tenemos suficientes datos?



Desajuste



Ajuste



Sobreajuste



Aumentación de datos

¿Qué hacemos si no hay más datos? **Aumentarlos**

Synthetic Minority Over-Sampling Technique (SMOTE)

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16(1), pp. 321-357



Aumentación de datos

¿De qué trata SMOTE?



$y_1 = \text{dog}$

$N_1 = 80$



$y_2 = \text{cat}$

$N_2 = 20$



$$\mathbf{x}_{dog} = (x_1, x_2, \dots,$$

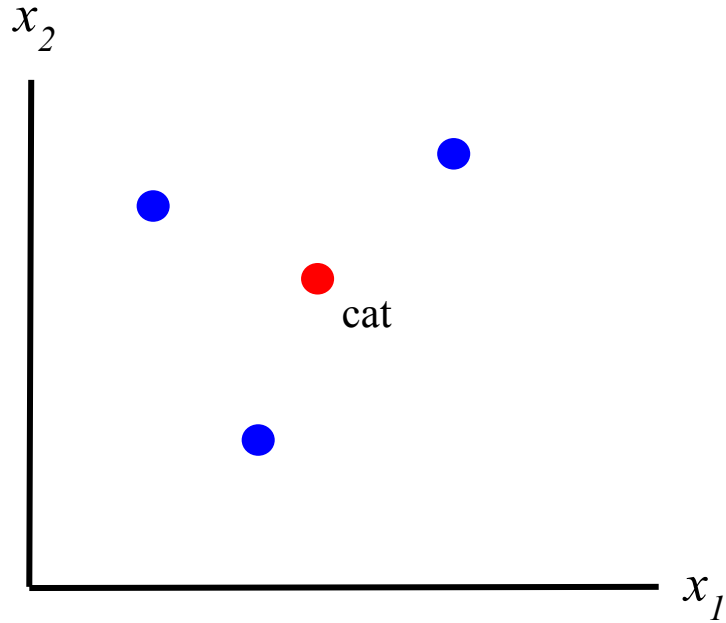
$x_D)$

$$\mathbf{x}_{cat} = (x_1, x_2, \dots, x_D)$$



Aumentación de datos

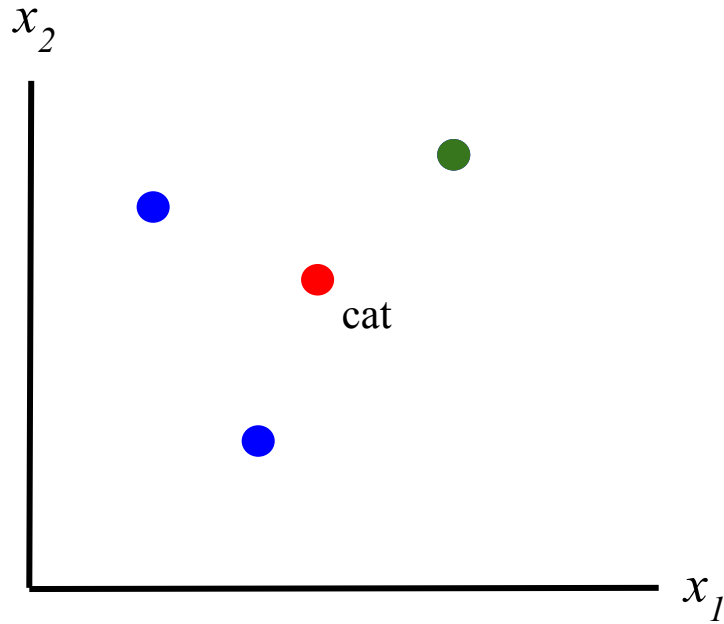
Elegimos a los k vecinos más cercanos de la misma clase



Aumentación de datos

Elegimos un número entero entre 1 y k al azar

$$p = 2$$



Aumentación de datos

Crear ejemplo sintética provisional:

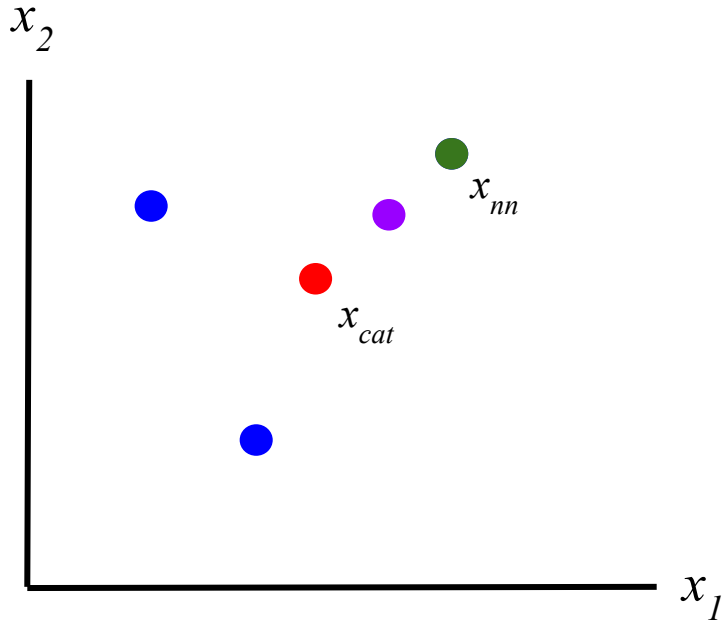
$$x_{new} = []$$

Por cada variable i en x_{cat} , calcular lo siguiente:

$$diff = x_{nn}^i - x_{cat}^i$$

$$gap = U(0,1)$$

$$x_{new}^i = x_{cat}^i + gap \cdot diff$$



Selección de características

x_1	x_2	x_3	x_4	x_5	y
1	1	0	1	0	1
1	1	1	0	0	0
0	1	1	1	0	0
0	1	0	0	0	1

5 características/variables



x_1	x_4	y
1	1	1
1	0	0
0	1	0
0	0	1

Dimensionalidad
reducida



Selección de características

Utilizado para:

- Eliminar características redundantes
- Reducir el ruido en los datos
- Mejorar el desempeño del aprendizaje
- Minimizar la “maldición de la dimensionalidad”



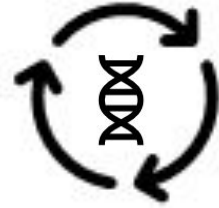
Selección de características



Selección manual



Fuerza bruta



**Computación
Evolutiva**



Algoritmo SFS

Sequential Forward Selection



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
-------	-------	-------	-------	-------	-------	-------	-------	-------

Seleccionar la mejor, con respecto a una métrica



Algoritmo SFS



Por
pares

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
-------	-------	-------	-------	-------	-------	-------	-------	-------



Por
ternas

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
-------	-------	-------	-------	-------	-------	-------	-------	-------



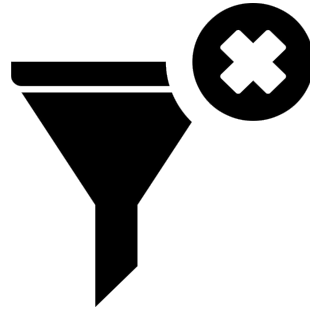
Hasta obtener las k características seleccionadas...



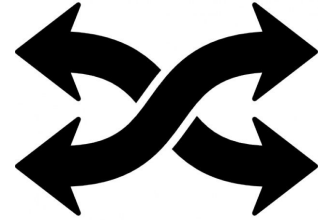
Algoritmos Evolutivos



Métodos *wrapper*



Métodos *filter*



Métodos *híbridos*



Imágenes

- i1. Analytics (s.f.). *Semantic Image Segmentation Annotation for Deep Learning Analytics*. Recuperado de:
https://docs.google.com/presentation/d/1urNzejQZE-69cyLbN5KDj_0t9feteLf8aCHedpuieE/edit#slide=id.gda9529ccee_0_0
- i2. Kuno, N. S. (2019). *Winners announced in multi-agent reinforcement learning challenge*. Recuperado de:
<https://www.microsoft.com/en-us/research/blog/winners-announced-in-multi-agent-reinforcement-learning-challenge/>
- i3. Gershgor, D. (2018). *AI can generate fake faces now. Here's how to spot them*. Recuperado de:
<https://qz.com/1510746/how-to-identify-fake-faces-generated-by-ai/>
- i4. Loukas, S. (2020). *Time-Series Forecasting: Predicting Stock Prices Using Facebook's Prophet Model*. Recuperado de:
<https://ai.plainenglish.io/predicting-stock-prices-using-facebooks-prophet-model-b527ea8e445>
- i5. Pegus digital (s.f.). *Top 5 Voice Recognition Apps*. Recuperado de:
<https://pegus.digital/top-5-voice-recognition-apps/>
- i6. Trouillet, B. (2019). *From Phantom Limb to Prosthetic Arm*. Recuperado de:
<https://news.cnr.fr/videos/from-phantom-limb-to-prosthetic-arm>



Imágenes

- i7. Cohut, M. (2018). *Dogs: Our best friends in sickness and in health*. Recuperado de: https://www.medicalnewstoday.com/articles/322868#_noHeaderPrefixedContent
- i8. n.a. (s.f.). s.t. Recuperado de: <http://www.imcorp.jp/jmss.asp?iid=162948521&cid=43>



Referencias

1. Rebala, G., Ajay, R. & Churiwala, S. (2019). *Introduction to machine learning*.
<https://doi.org/10.4018/978-1-7998-0414-7.ch003>

