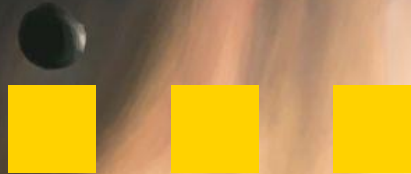


Dylan Vian

Sesión 2

Desafío CITT Duoc



Roberto Muñoz
Senior Data Scientist
EY MetricArts



■ The better the question. ■ The better the answer. ■ The better the world works.



Machine Learning

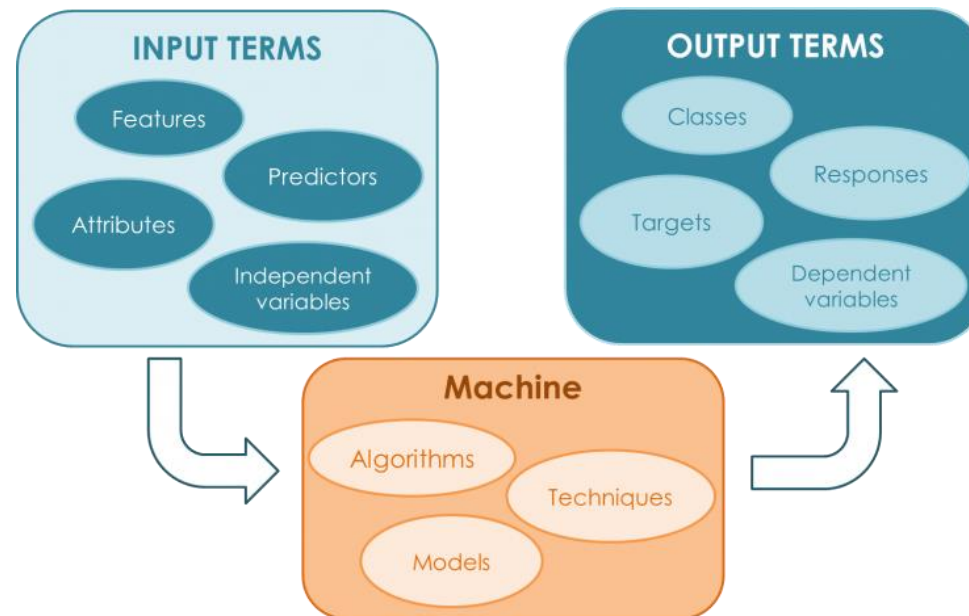


Links de interés

- Awesome Machine learning
 - <https://github.com/josephmisiti/awesome-machine-learning>
- Kaggle fake news
 - <https://www.kaggle.com/c/fake-news/data>
- Google datasearch
 - <https://datasetsearch.research.google.com/>

Machine learning

- *Machine learning* es un campo interdisciplinario basado fuertemente en la programación y análisis estadístico.
- Su principal objetivo es desarrollar y aplicar técnicas que permitan a los computadores aprender o imitar habilidades cognitivas propias de nosotros.



Usual tasks

- **Regresión**

"A Entel le interesa saber cuántos GB consumirán sus clientes el próximo mes. Deberá construir un modelo de regresión."

- **Clasificación**

"Falabella quiere promocionar un nuevo producto. ¿En qué segmento debería poner las fichas y promocionar este producto?"

- **Clustering**

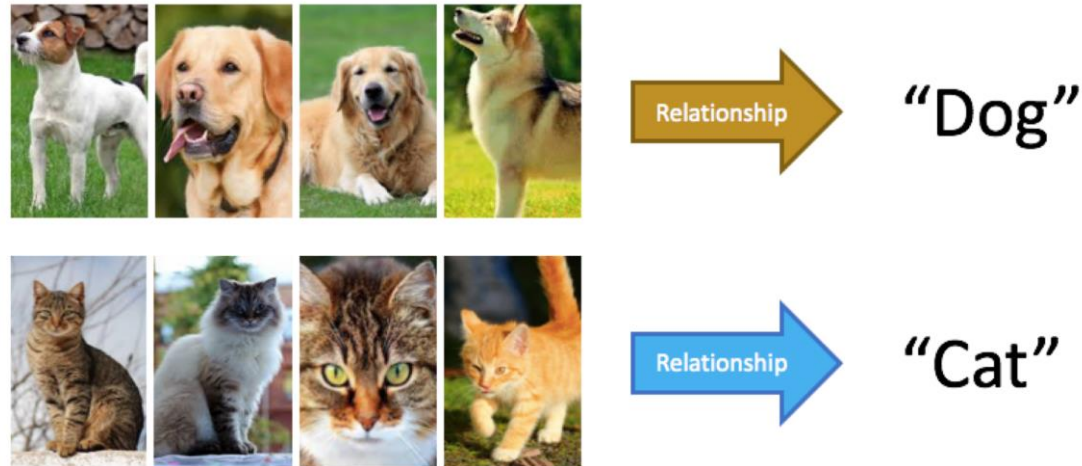
"Los observatories de la ESO necesitan procesar las imágenes que toman cada noche e identificar de manera automática las imágenes anómalas. Los datos no están etiquetados."

Tipos de ML

- Supervised learning

El sistema aprende utilizando datos previamente clasificados. Los datos pueden estar estructurados o no estructurados.

El algoritmo genera un modelo que establece la correspondencia entre los datos de entrada y la salida esperada del sistema.



Tipos de ML

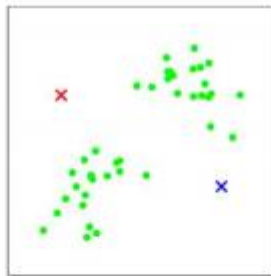
- Unsupervised learning

Los datos no han sido clasificados previamente.

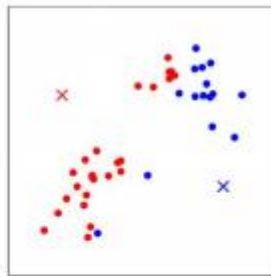
El sistema debe poder reconocer patrones y generar sus propias etiquetas. El modelo debe clasificar los nuevos datos de entrada.



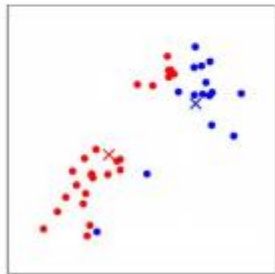
(a)



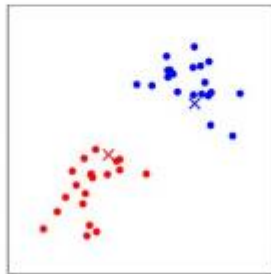
(b)



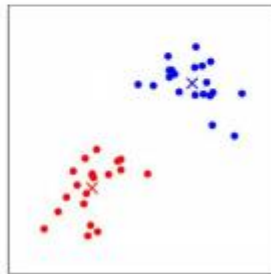
(c)



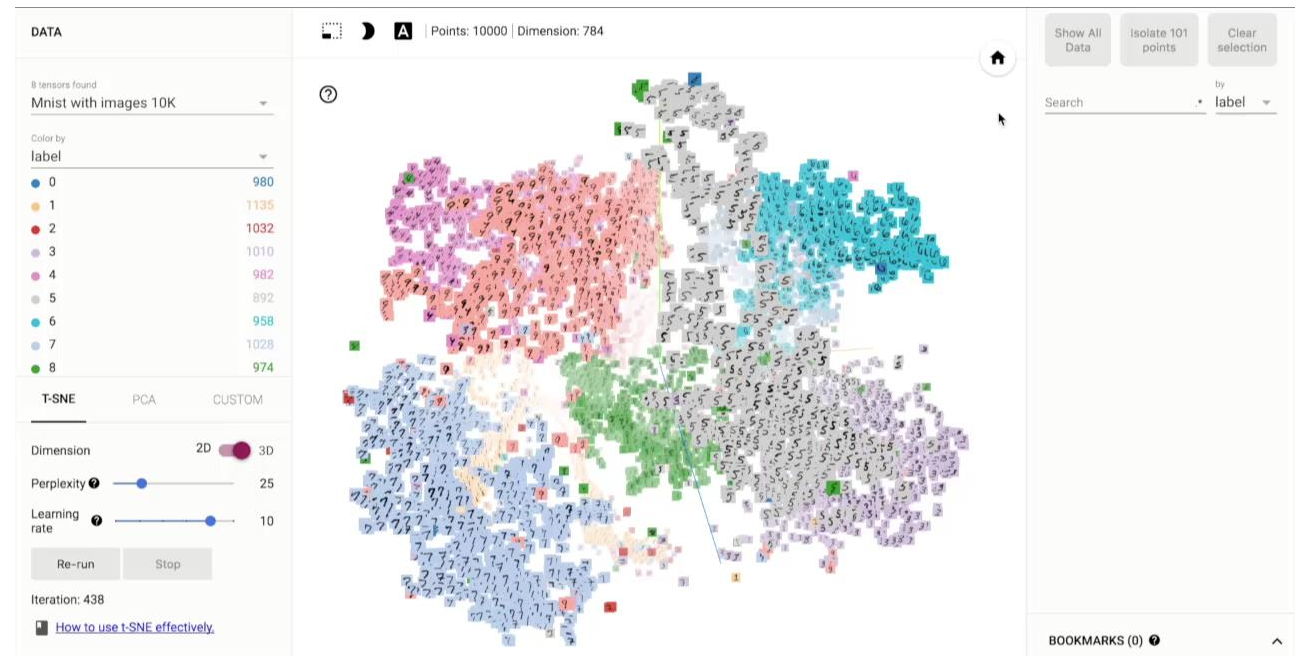
(d)



(e)



(f)

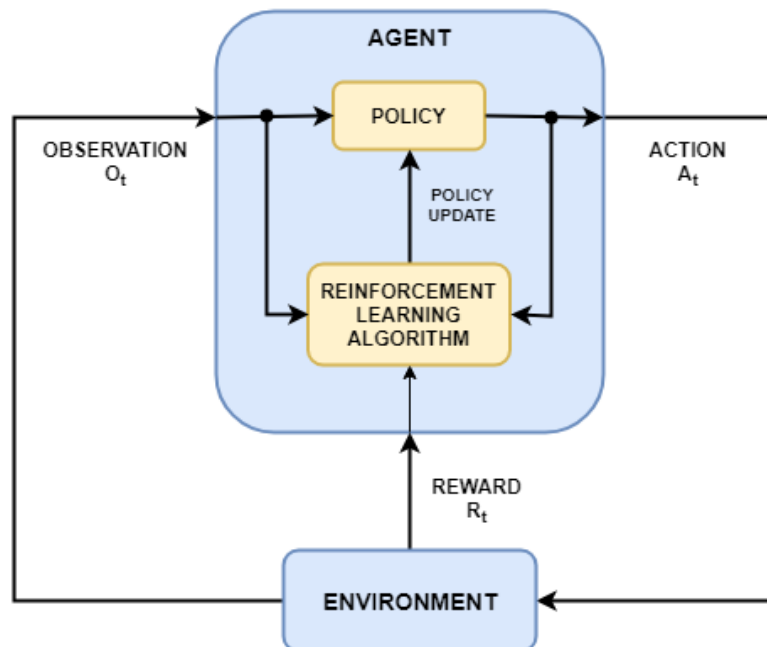


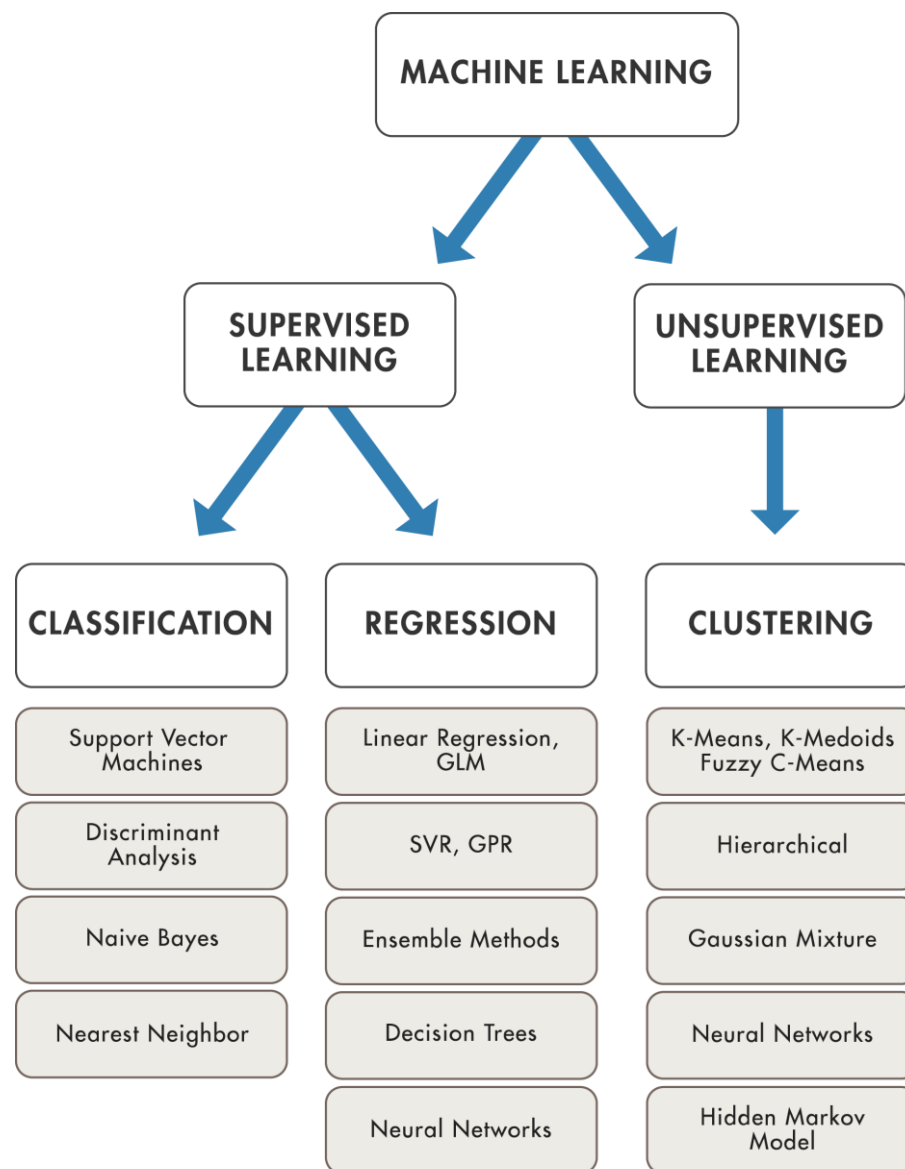
Tipos de ML

- Reinforcement learning

Tenemos un entorno que representa el mundo exterior para el agente y un agente que realiza acciones.

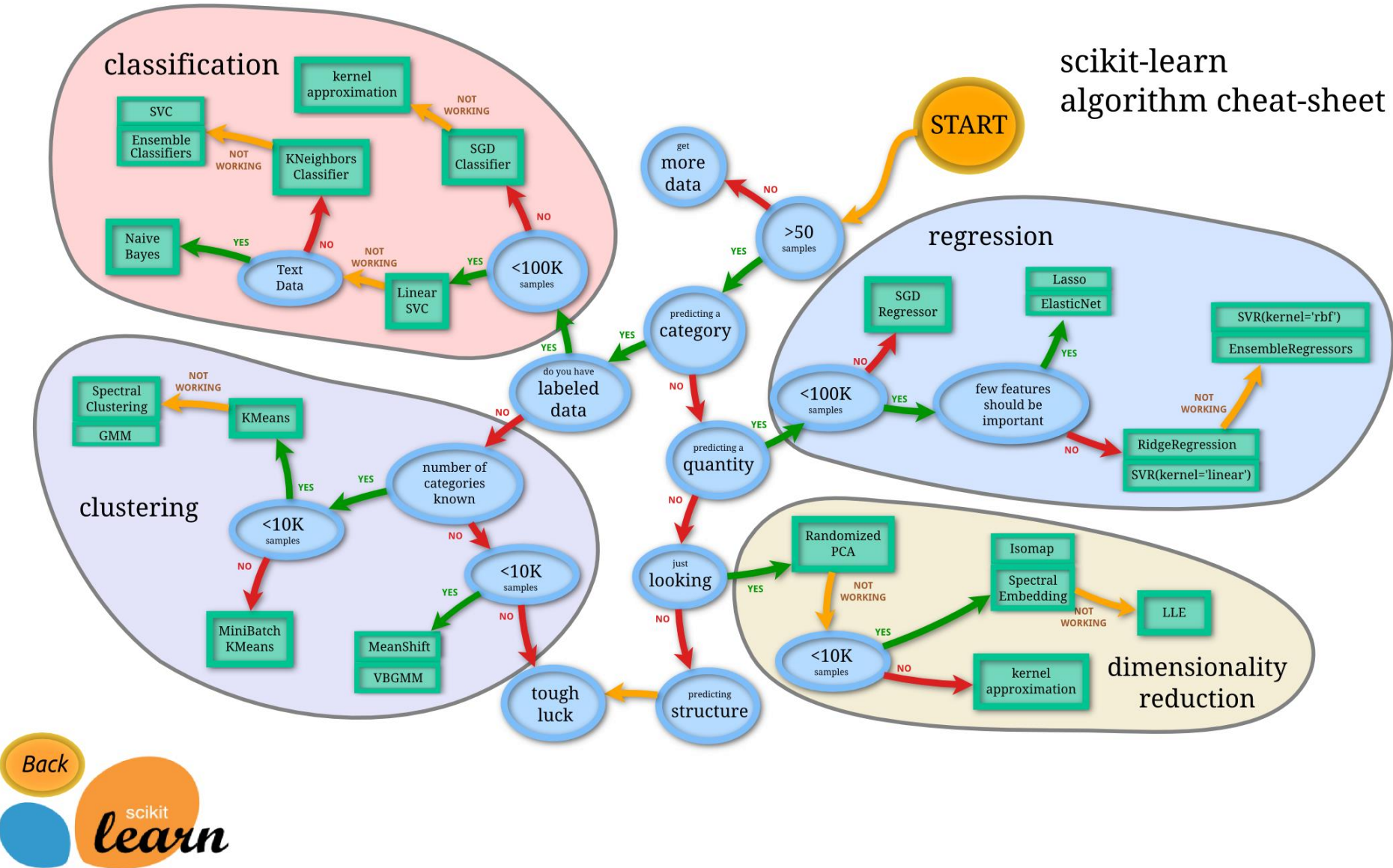
El agente recibe observaciones del entorno que consisten en una recompensa por su acción e información de su nuevo estado.





Métodos de Scikit-learn

scikit-learn algorithm cheat-sheet



Librerías en Python

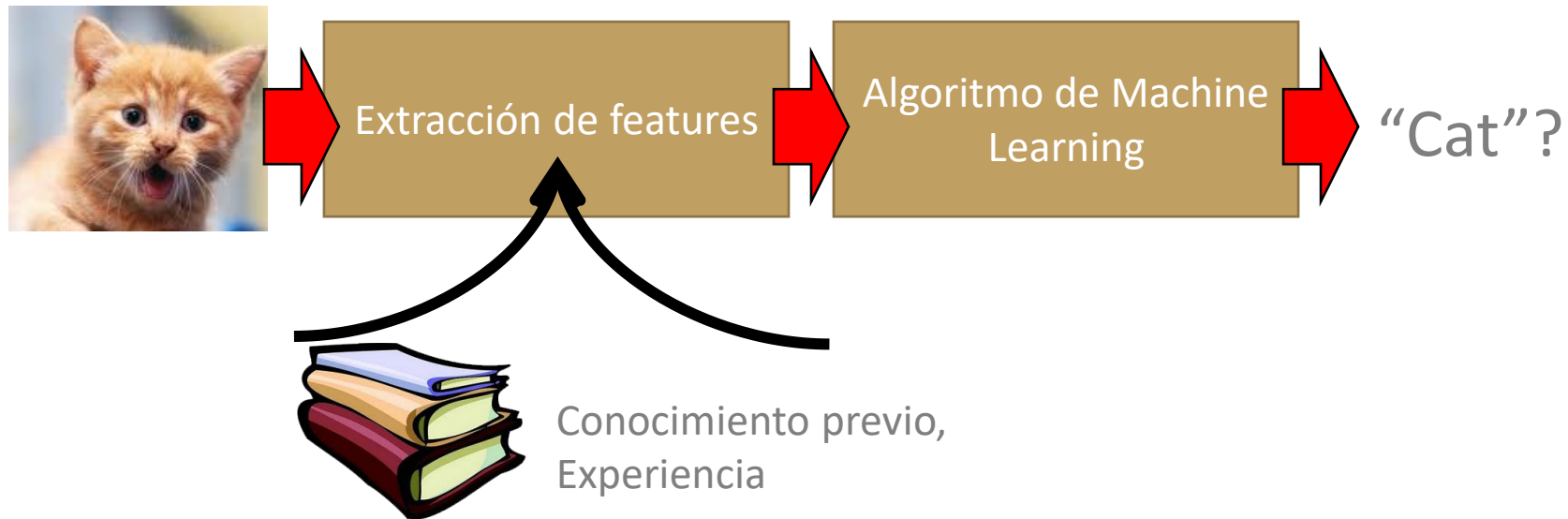
- Pandas
 - <https://pandas.pydata.org/>
 - Lectura de datos, limpieza, ingeniería de features, normalización
- Scikit-learn
 - https://scikit-learn.org/stable/user_guide.html
 - Entrenamiento de modelos supervisados y no supervisados
- Spacy
 - <https://spacy.io>
 - Procesamiento de lenguaje natural

The background is a dark blue field filled with a complex network of glowing nodes and connecting lines. The nodes are small spheres in white, cyan, and orange, while the lines are thin and translucent in shades of red and orange. A bright, multi-colored starburst of light is positioned in the center of the image. Faint, vertical columns of binary code (0s and 1s) are visible in the background, adding a digital aesthetic.

Deep Learning

ML con features manuales

- El approach clásico es hacer machine learning con ingeniería de features manual
 - Algoritmo de ML aprende a hacer predicciones a partir de la representación de mayor nivel
 - Se usa conocimiento del dominio y experiencia previa

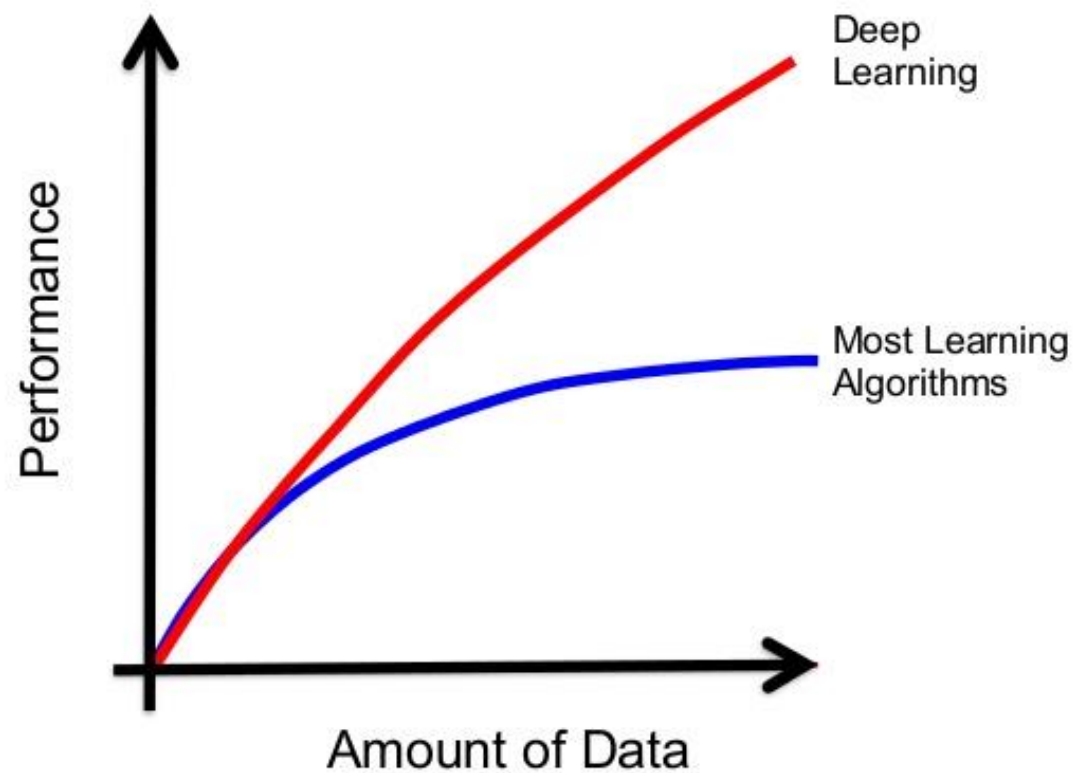


Proliferación datos no estructurados

- Datos no estructurados son datos crudos y que suelen no estar organizados.
- Deben ser procesados y transformados para luego ser almacenados en una base de datos.



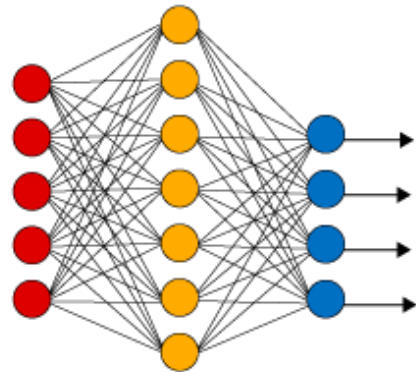
BIG DATA & DEEP LEARNING



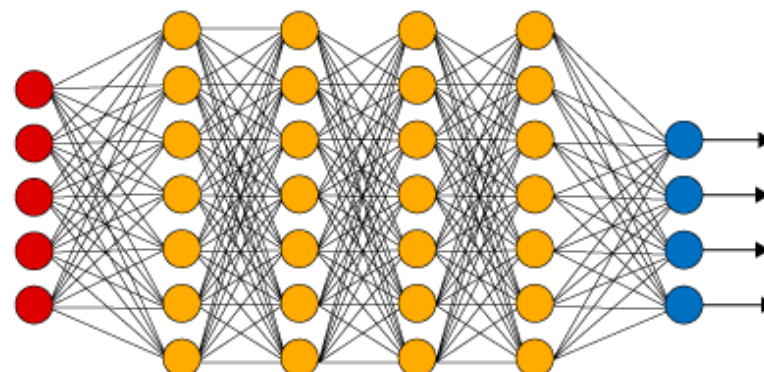
Deep Learning

- Deep learning is a collection of machine learning methods based in feature or representation learning.
- The most famous are Deep Neural Networks (DNN). They are based on Artificial Neural Networks and multiple layer architectures.

Simple Neural Network



Deep Learning Neural Network



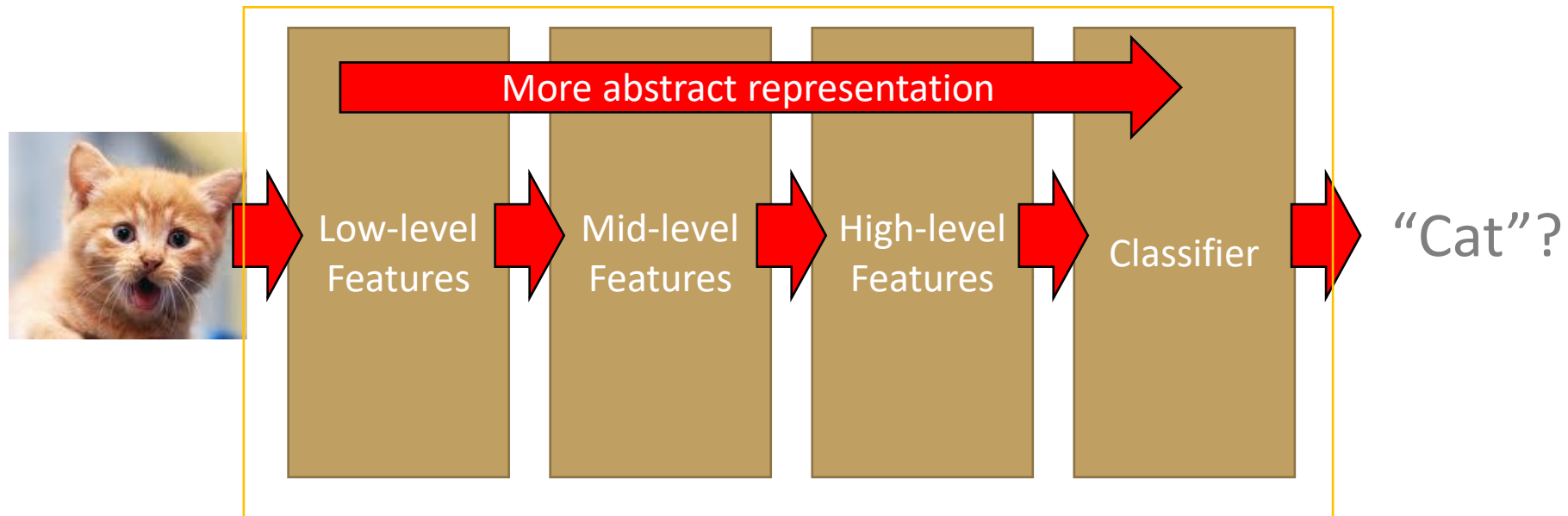
● Input Layer

● Hidden Layer

● Output Layer

Deep Learning

- Deep Learning
 - Entrenar múltiples capas de features y abstracciones a partir del dato no estructurado
 - Tratar de descubrir las representaciones que permitan hacer mejores predicciones



Librerías en Python

- Tensorflow
 - <https://www.tensorflow.org/>
 - Framework para entrenar modelos de Deep Learning
 - Desarrollado originalmente por Google
- Keras
 - <https://keras.io/>
 - Framework de alto nivel para entrenar modelos de Deep Learning
 - Puede usar otros frameworks en el backend
 - Desarrollado originalmente por François Chollet
- Pytorch
 - <https://pytorch.org/>
 - Framework para entrenar modelos de Deep Learning
 - Desarrollado originalmente por Facebook

Visit the website

<http://playground.tensorflow.org>

