

Inteligencia Artificial y BigData en VoIP

Elio Rojano
<http://sinologic.net/>

[Introducción]

Quién soy

```
{
  name: "Elio Rojano",
  email: "elio@sinologic.net",
  web: "https://www.sinologic.net/",
  work: "http://VOZ.com",
  twitter: "https://twitter.com/hellc2",
  linkedin: "https://www.linkedin.com/in/rojano/",
  interests: [
    "VoIP",
    "WebRTC",
    "SIP",
    "Development",
    "Javascript",
    "PHP",
    "Python",
    "Linux",
    "OpenSource"
  ],
  keynotes: "https://www.sinologic.net/portfolio"
}
```

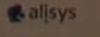
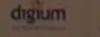
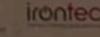
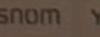
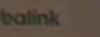


```
/etc/asterisk
/usr/sbin/asterisk
/usr/lib/asterisk
/var/lib/asterisk
/var/spool/asterisk
/var/log/asterisk
```

```
[logfiles]
;; hundreds of comments about how this file works...
;debug => debug
;console => notice,warning,error
;console => notice,warning,error,debug
;messages => notice,warning,error,security
;full => notice,warning,error,debug,verbose,dtmf,fax

:syslog.local0 => notice,warning,error

/var/log/asterisk/messages
/var/log/asterisk/full
```

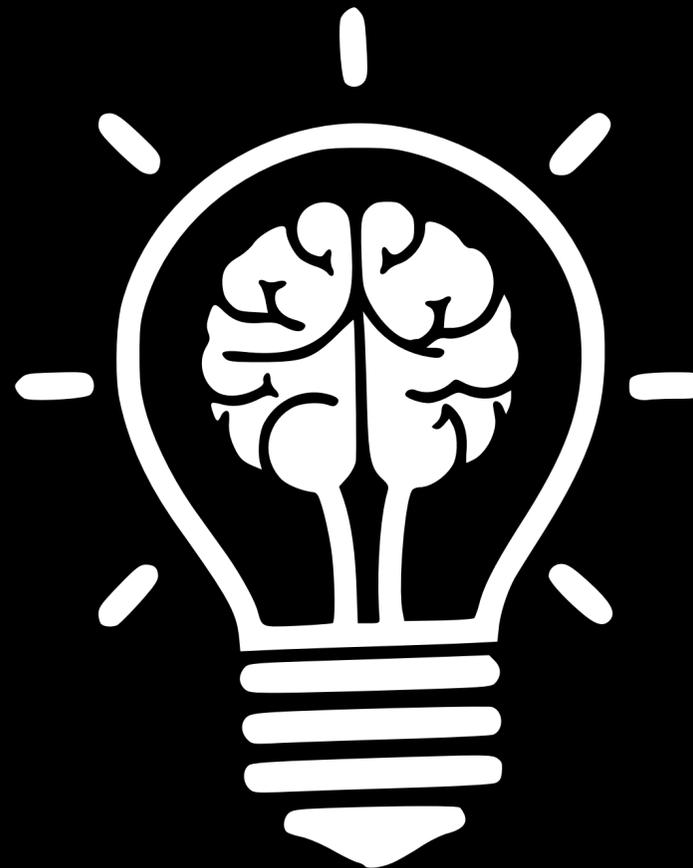
Interés en la IA y la VoIP

Curiosidad sobre IA

Conceptos matemáticos

+

Datos



Curiosidad sobre VoIP

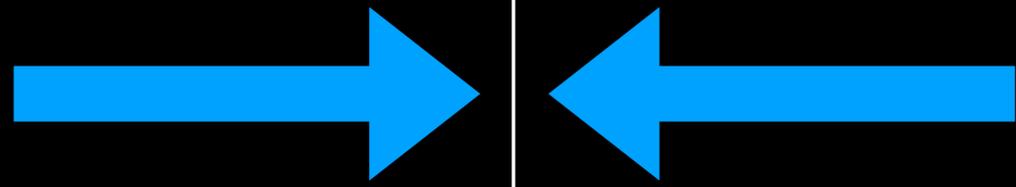
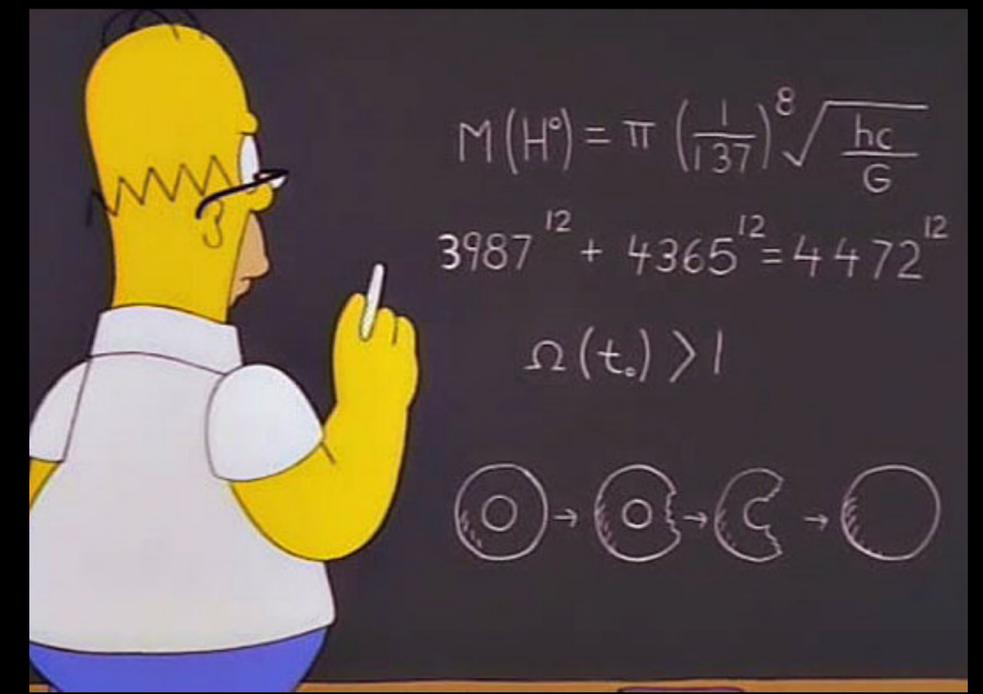
La gente hacen llamadas

+

Las llamadas generan datos

Dónde está la IA

Inteligencia Artificial



Informática

Matemática

¿Qué tal las matemáticas?



No, mates no.

$$X+X+X= 60$$

$$X+Y+Y= 30$$

$$Y-2Z= 3$$

$$Z+X+Y=?$$



¡Un reto!
Me apunto.

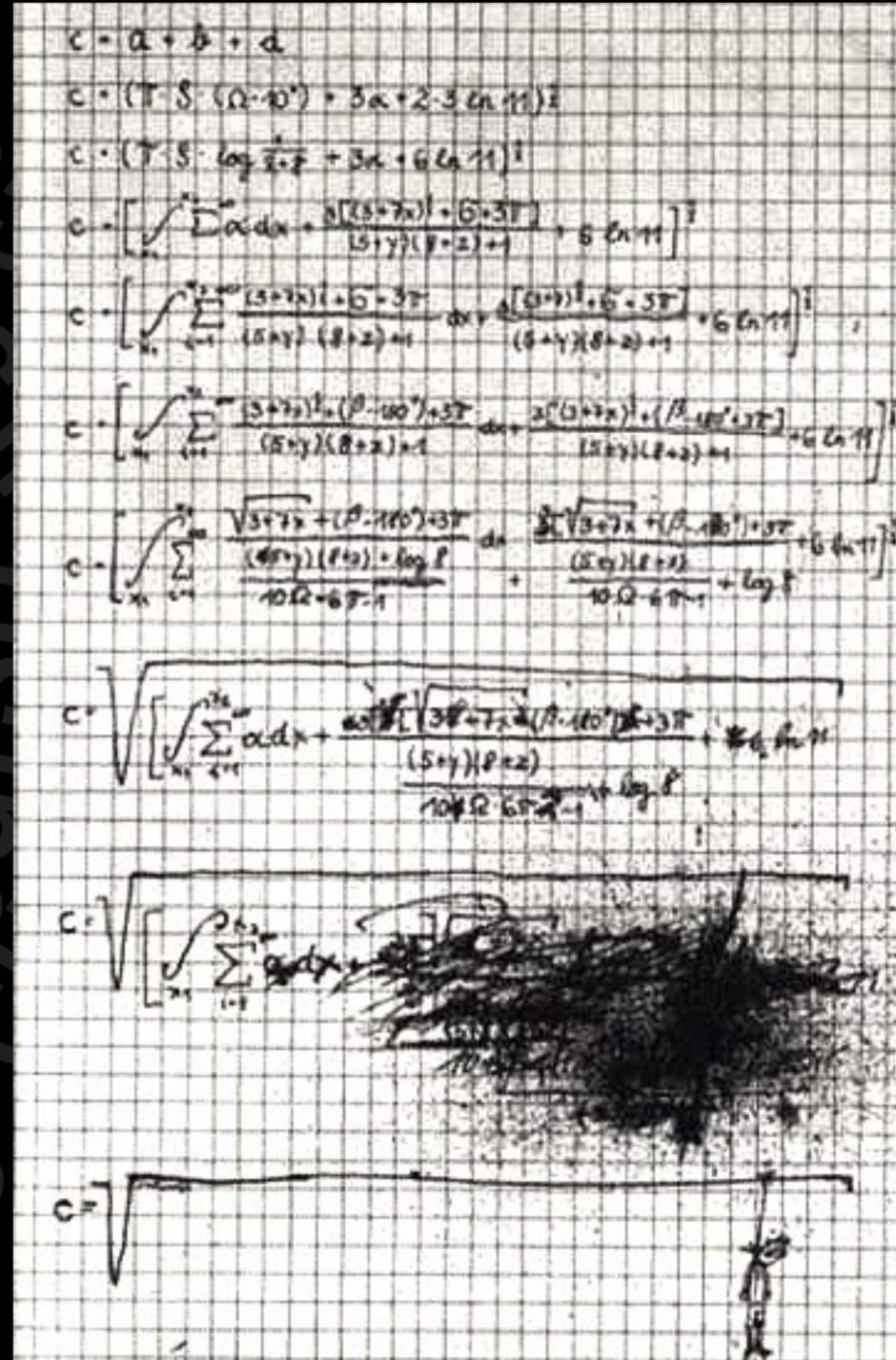
$$\text{🍏} + \text{🍏} + \text{🍏} = 60$$

$$\text{🍏} + \text{🍌} + \text{🍌} = 30$$

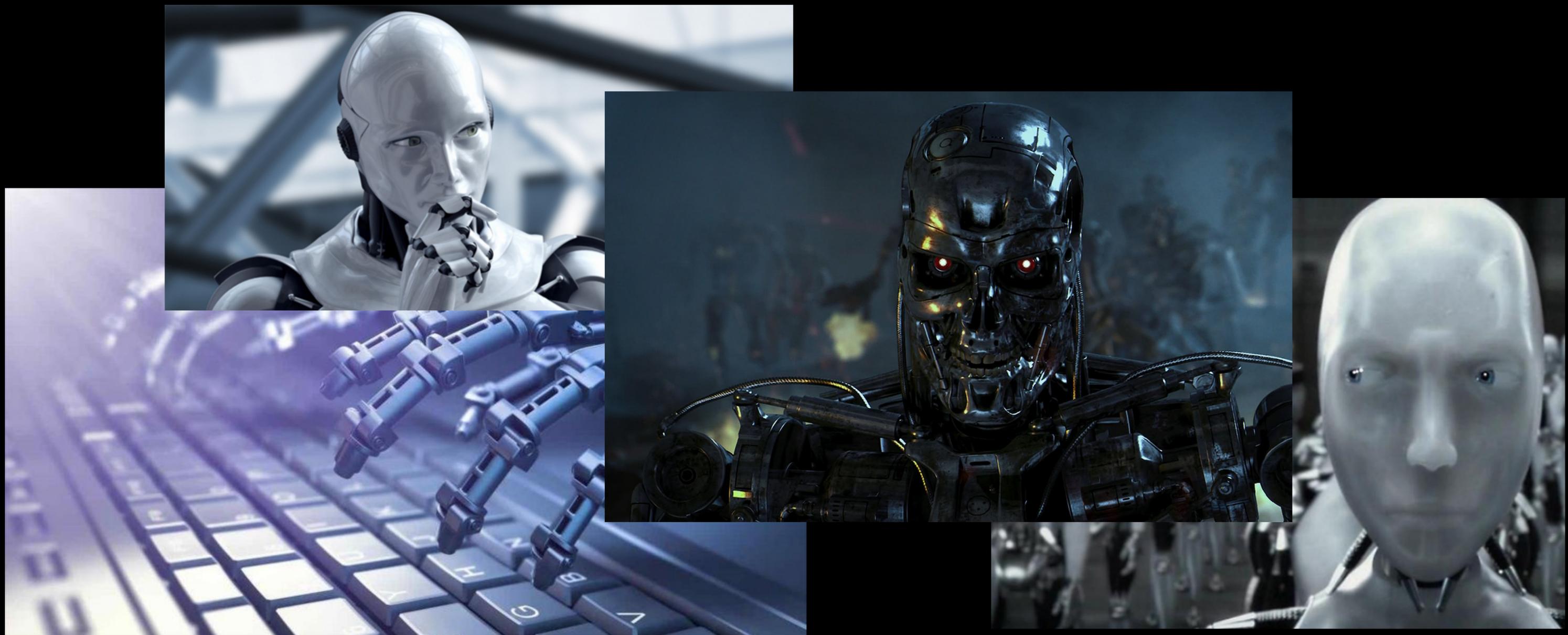
$$\text{🍌} - \text{🥥} = 3$$

$$\text{🥥} + \text{🍏} + \text{🍌} = ?$$

Al comienzo, bastante frustrante



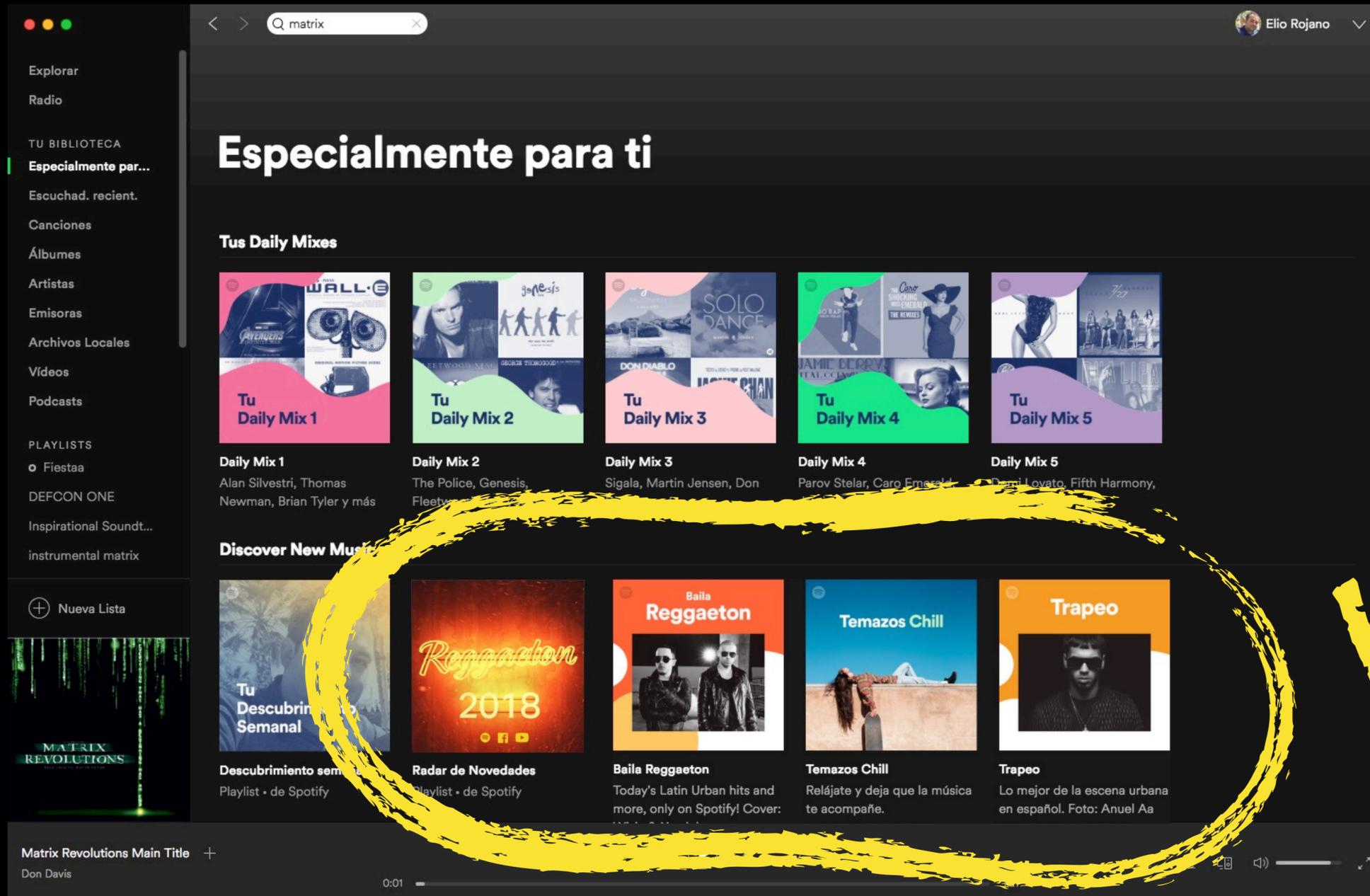
¿Por qué interesa la IA?



¿Por qué interesa la IA?

- Prever posibles datos y acelerar el trabajo más tedioso
- Hacer automáticamente el trabajo más tedioso
- Ser más prácticos en la obtención de objetivos
- Porque es un reto... 🍌 🍌 🍌
- Enseñar al sistema lo que nos gusta y que nos ofrezca lo que podemos desear

...aunque no siempre aciertan...



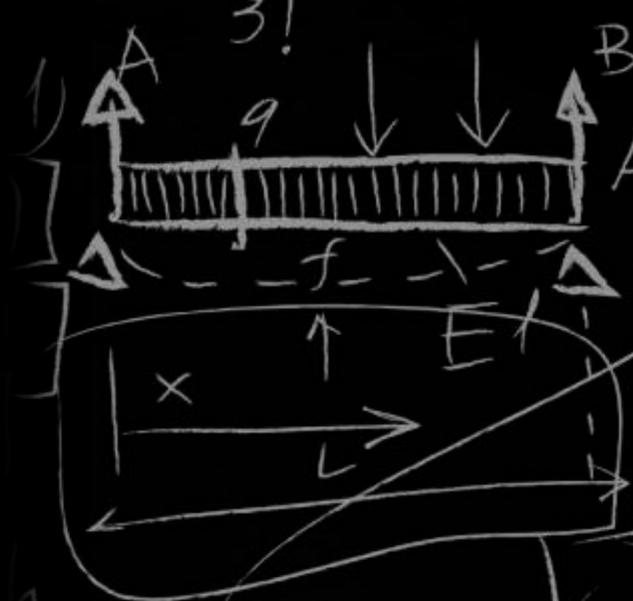
WTF!

Atención

Procesamiento de datos
está muy cerca, es muy parecido, pero no es
Inteligencia Artificial

Inconvenientes

$2 \sin(45^\circ + \alpha)$
 $2 \cos(45^\circ + \alpha)$
 $x^2 + \frac{n(n-1)(n-2)x^3}{3!} + \dots + \frac{x^n}{n!}$
 $\sin \frac{\alpha}{2} = \pm \sqrt{\frac{1 - \cos \alpha}{2}}$



$A = B = \frac{qL}{2}$
 $f = \frac{qL}{384 EI}$
 $T_x = \frac{q(L-2x)}{2}$

$SA = LA + \angle B$
 $\sin(\frac{\pi}{2} \pm \alpha) = \pm \cos \alpha$
 $\cos(\frac{\pi}{2} \pm \alpha) = \mp \sin \alpha$
 $\sin(\frac{\pi}{2} \pm \alpha) = \cos \alpha$
 $\cos(\frac{\pi}{2} \pm \alpha) = \mp \sin \alpha$
 $\sin(\frac{\pi}{2} \pm \alpha) = \cos \alpha$
 $\cos(\frac{\pi}{2} \pm \alpha) = \mp \sin \alpha$
 $\sin(\frac{\pi}{2} \pm \alpha) = \cos \alpha$
 $\cos(\frac{\pi}{2} \pm \alpha) = \mp \sin \alpha$

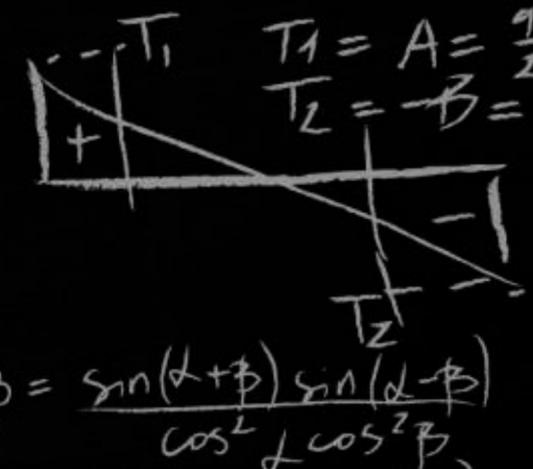
$\sin \alpha = \frac{2 \tan \frac{\alpha}{2}}{1 + \tan^2 \frac{\alpha}{2}}$
 $\cos \alpha = \frac{1 - \tan^2 \frac{\alpha}{2}}{1 + \tan^2 \frac{\alpha}{2}}$
 $\tan \alpha = \frac{2 \tan \frac{\alpha}{2}}{1 - \tan^2 \frac{\alpha}{2}}$

$2w + 2lh + 2wh$

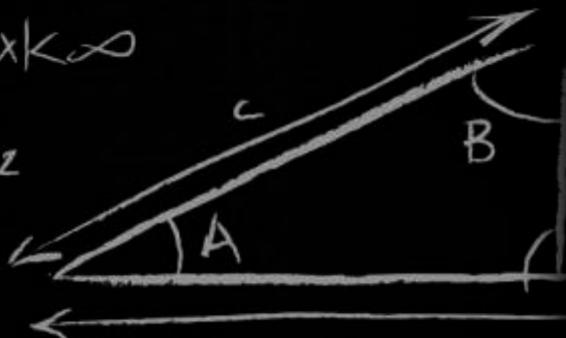
$\sin^2 \frac{\alpha}{2} = \frac{1 - \cos \alpha}{2}$
 $\cos^2 \frac{\alpha}{2} = \frac{1 + \cos \alpha}{2}$
 $\tan^2 \frac{\alpha}{2} = \frac{1 - \cos \alpha}{1 + \cos \alpha}$
 $\cot^2 \frac{\alpha}{2} = \frac{1 + \cos \alpha}{1 - \cos \alpha}$

$1 + x^2 + \dots + x^n + \dots = \sum_{n=0}^{\infty} x^n, |x| < 1$
 $1 - x^2 + \dots + (-x)^n + \dots = \sum_{n=0}^{\infty} (-1)^n x^n, |x| < 1$

$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}, |x| < \infty$
 $T_1 = A = \frac{qL}{2}$
 $T_2 = -B = -\frac{qL}{2}$
 $M = \frac{qL^2}{12}$



$A = B = \frac{qL}{2}$
 $T_1 = -T_2 = A$
 $f = \frac{qL^2}{384 EI}$

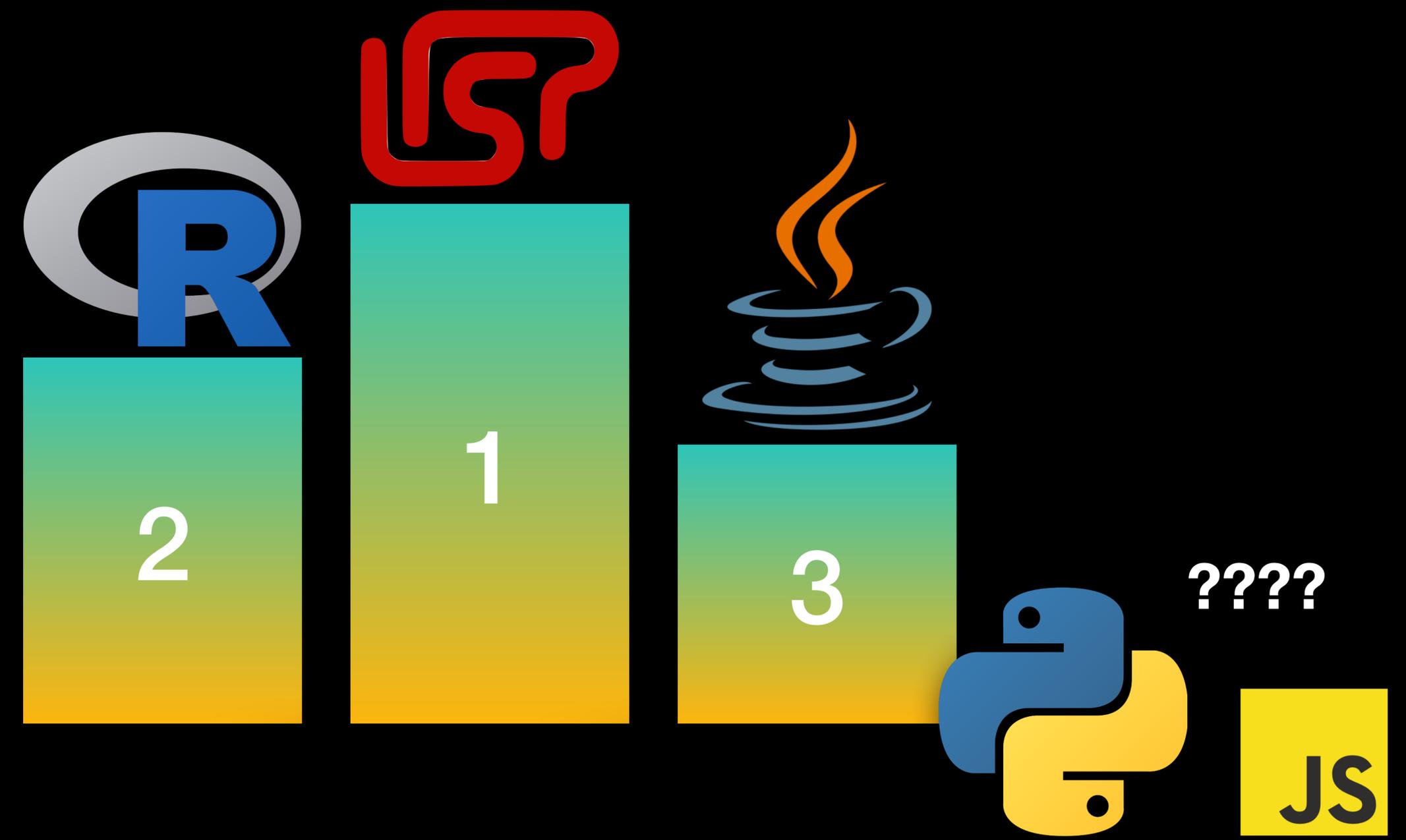


$\sin A = \frac{a}{c}$
 $\cos B = \frac{a}{c}$
 $\cot A = \frac{b}{a}$
 $\sec B = \frac{c}{a}$
 $\sin B = \frac{b}{c}$
 $\tan A = \frac{a}{b}$
 $\cot B = \frac{a}{b}$
 $\csc A = \frac{c}{b}$
 $\cos A = \frac{b}{c}$
 $\tan B = \frac{c}{a}$
 $\sec A = \frac{c}{b}$
 $\csc B = \frac{c}{a}$

Inconvenientes



Inconvenientes



Inconvenientes

1°. Lisp

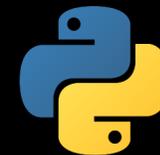
2°. R

3°. Java

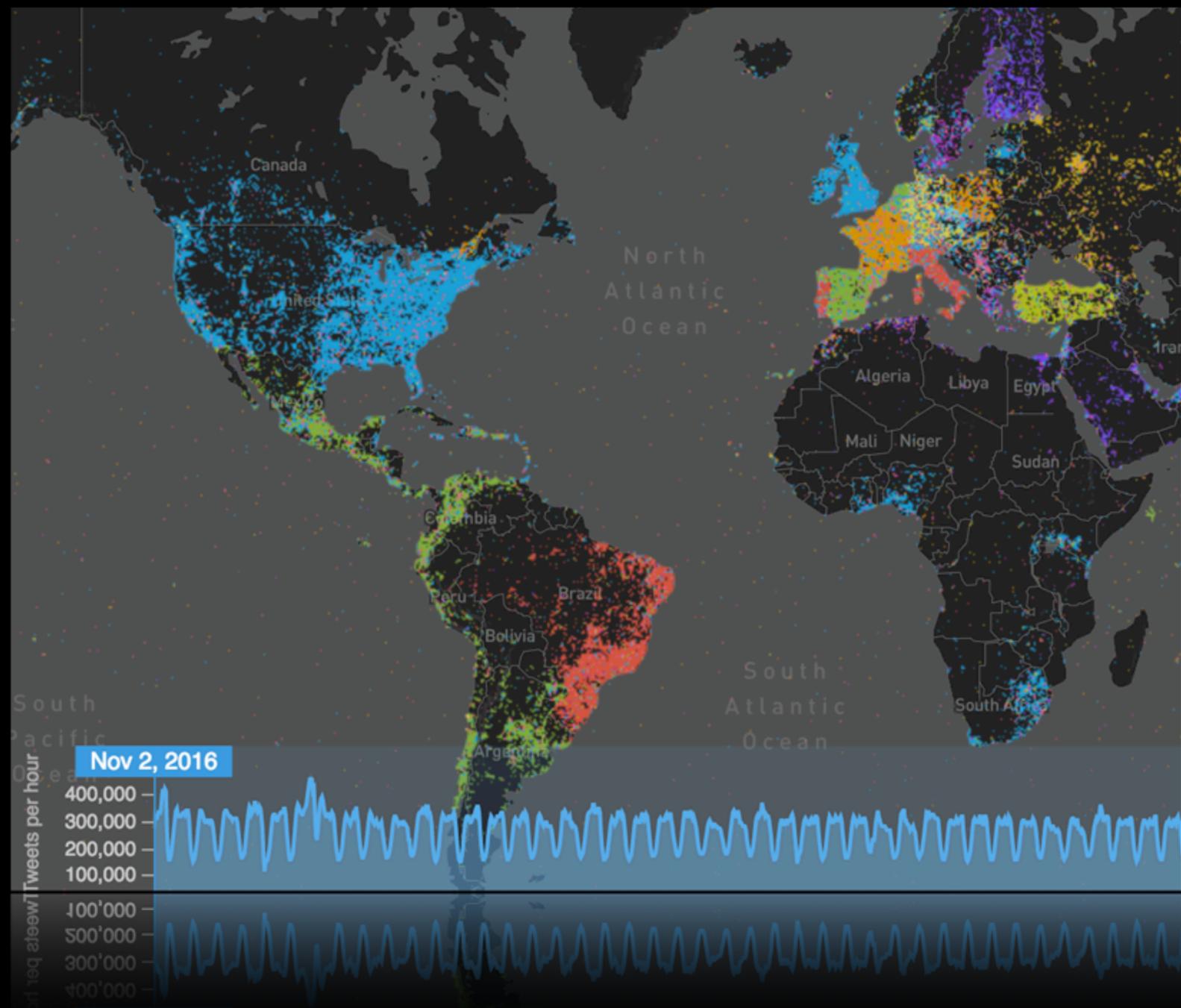
4°. Python

5°. Javascript

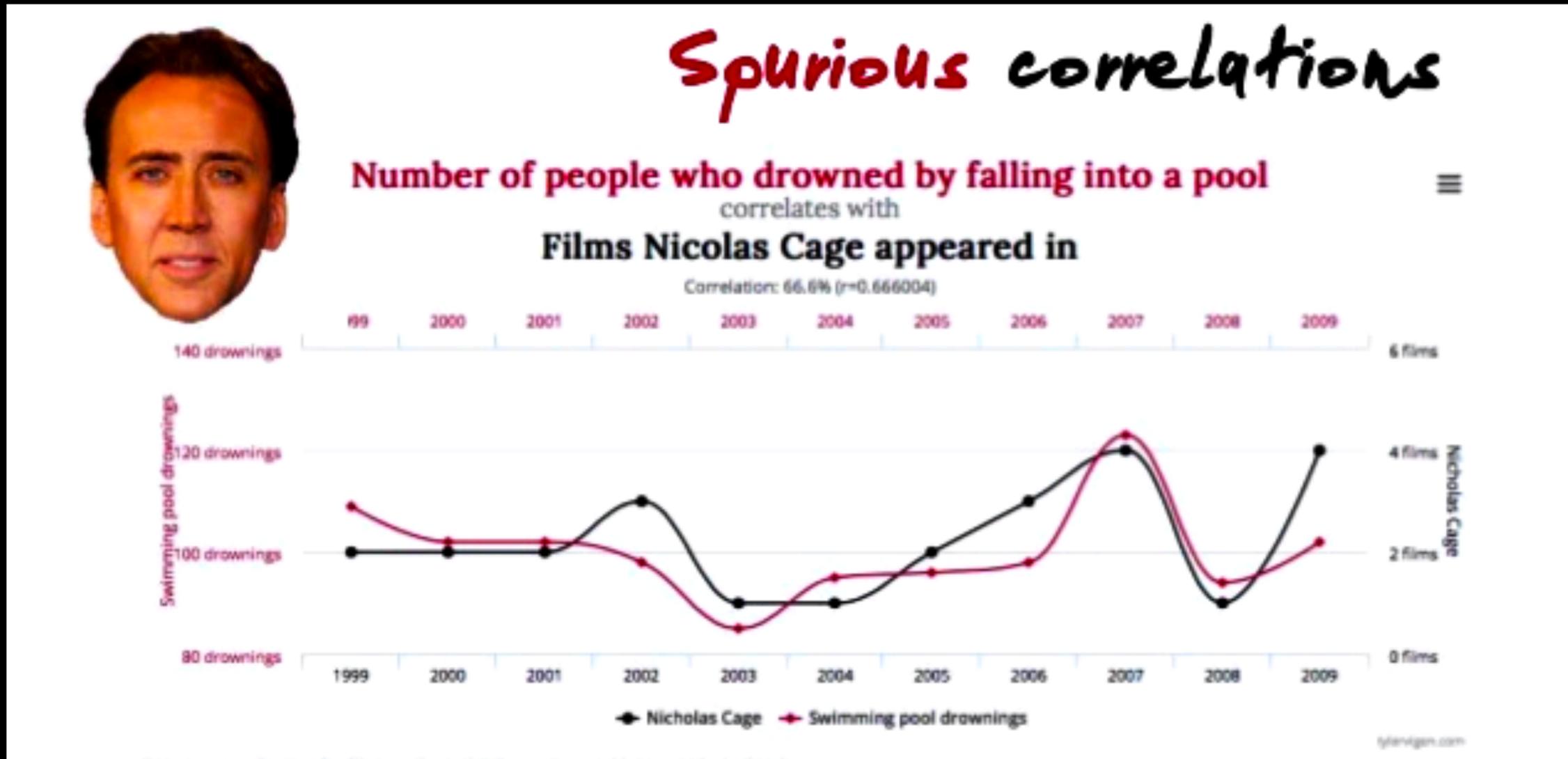
6°. PHP



Inconvenientes



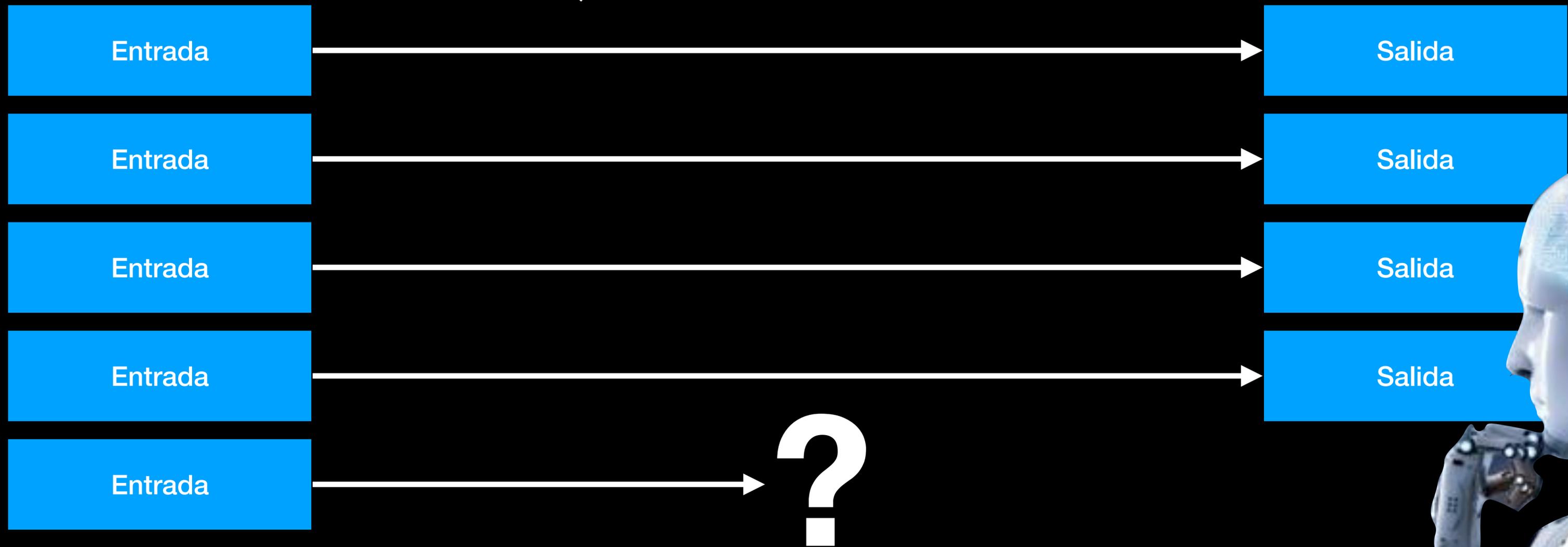
Correlación no implica causalidad



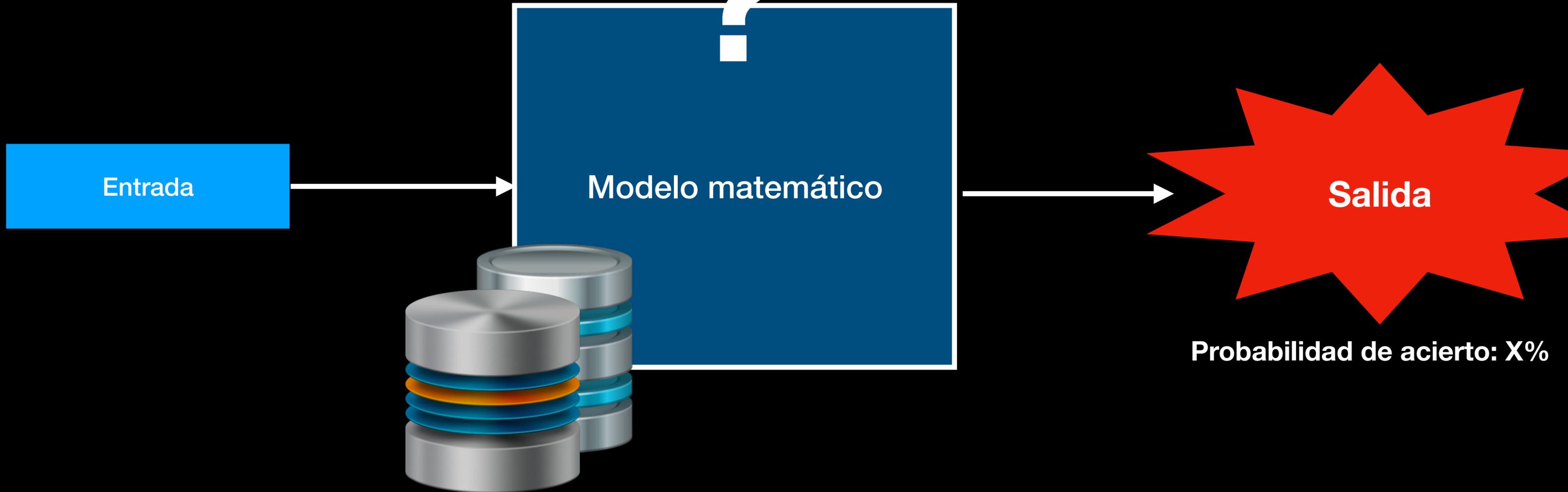
Número de ahogados por caídas en piscina en relación con películas de Nicolas Cage

[Empecemos]

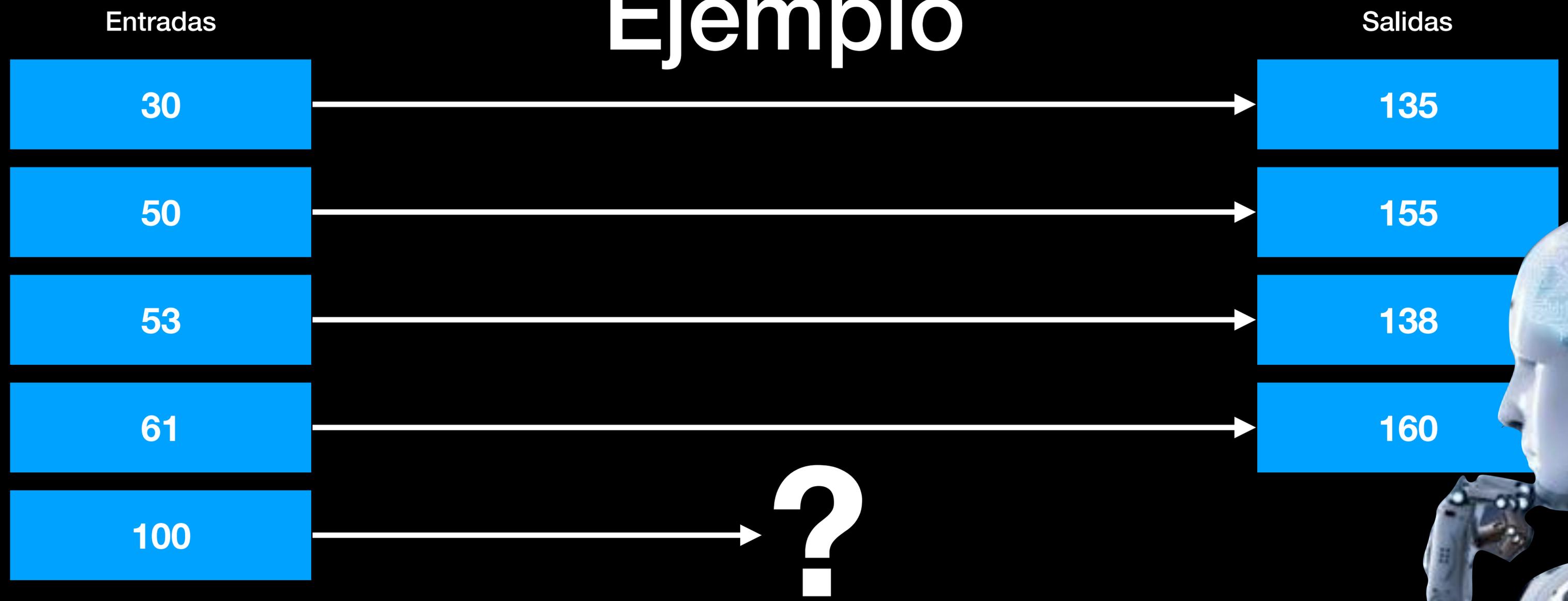
Qué es la IA



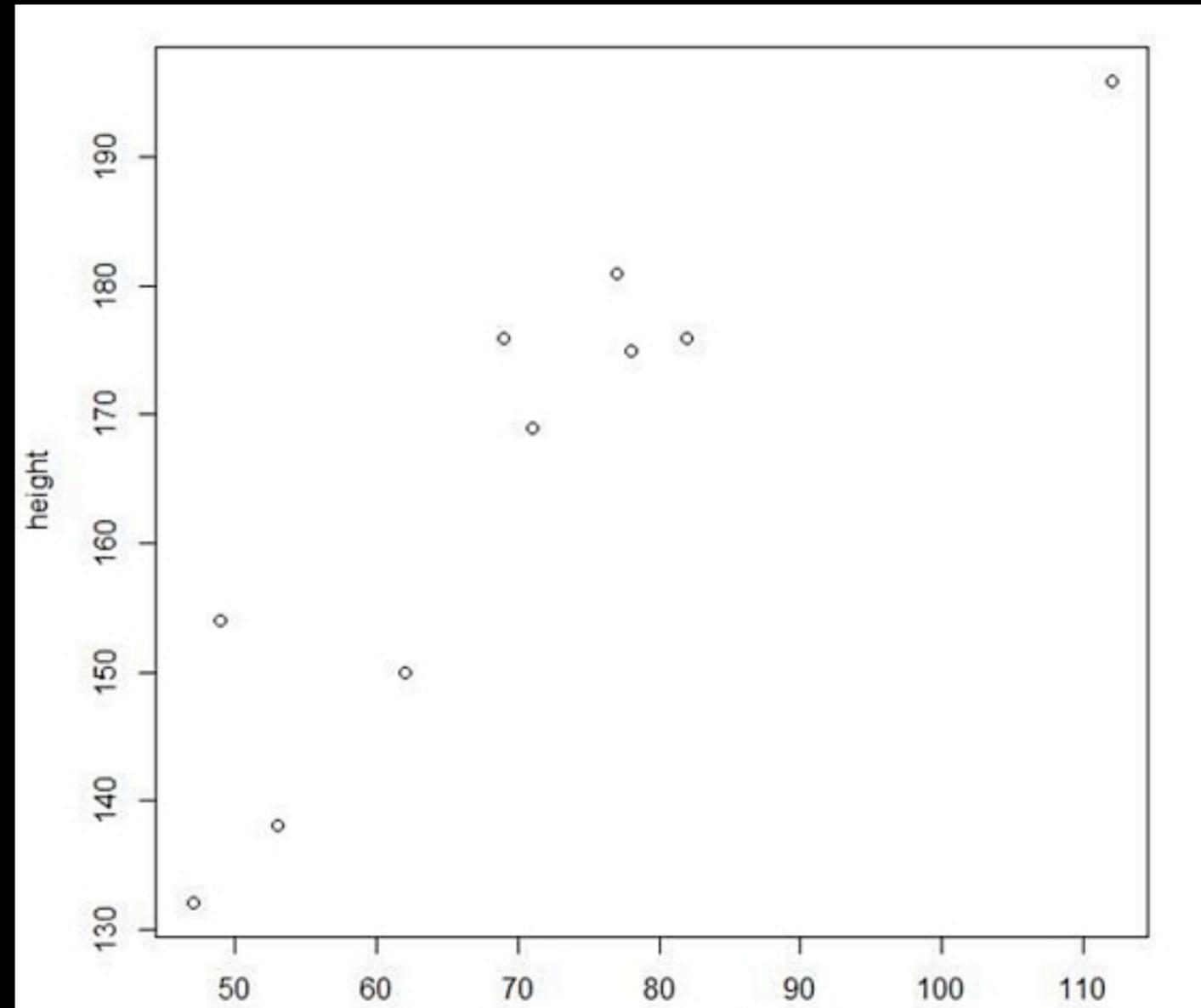
Qué es la IA



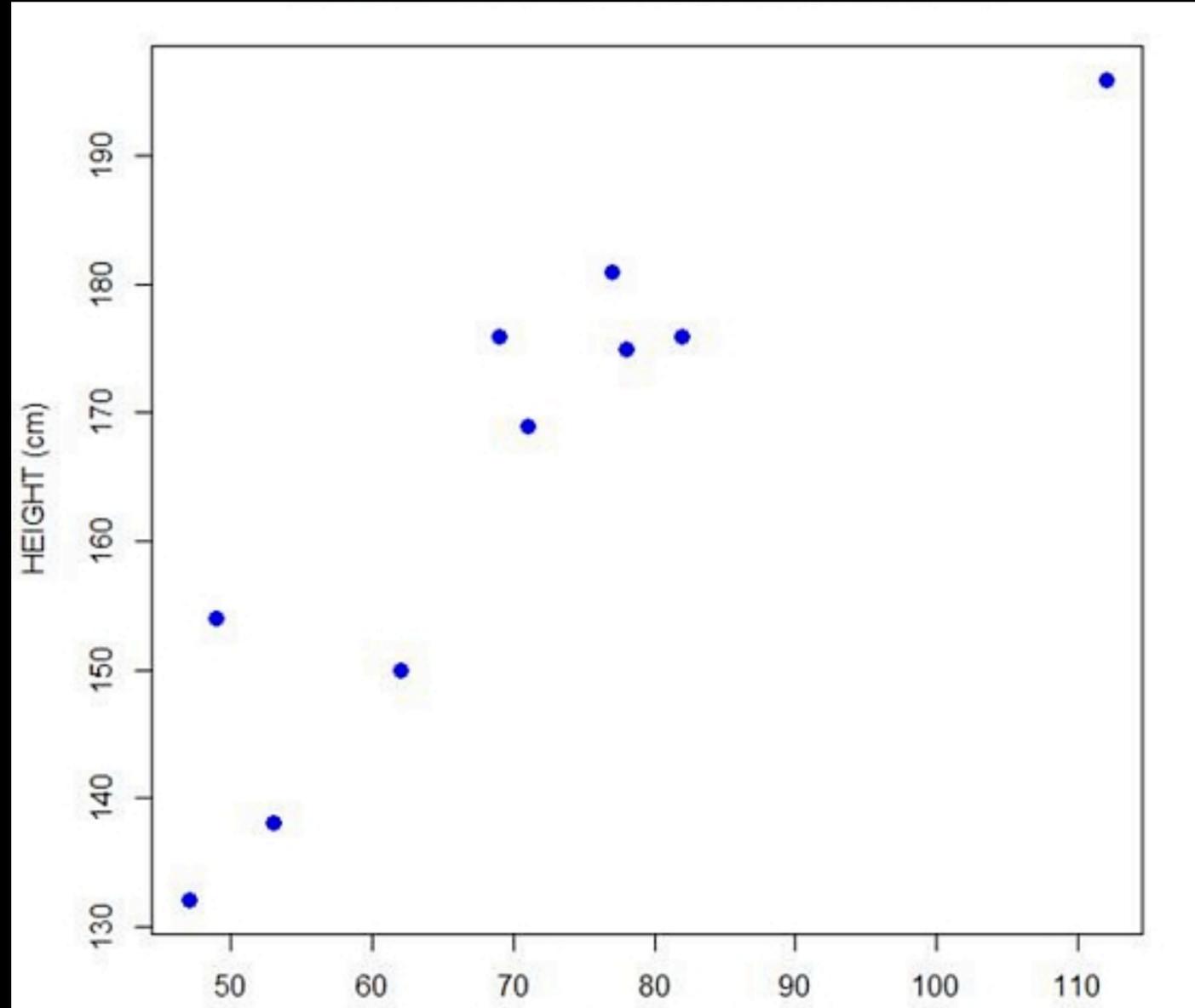
Ejemplo



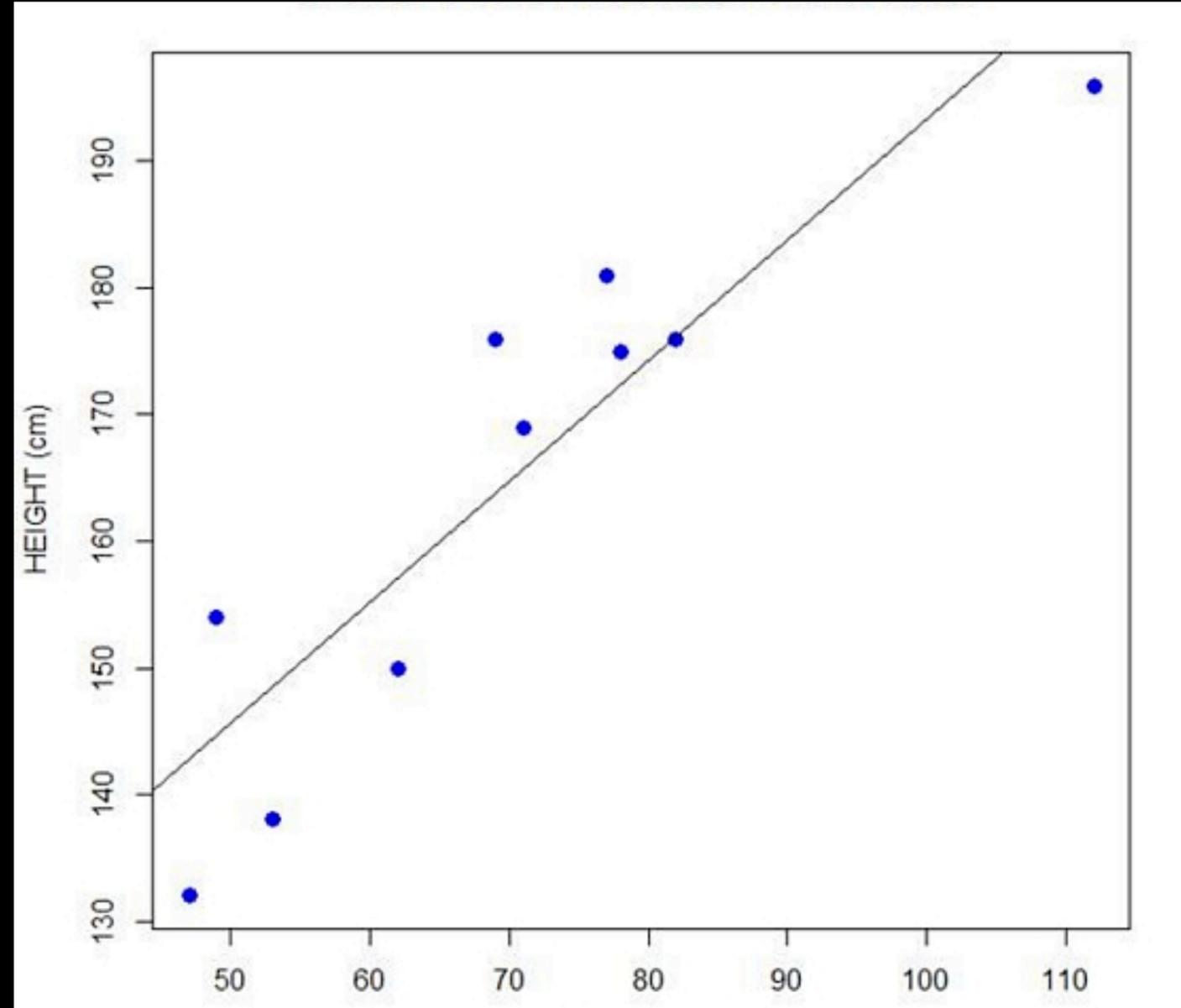
Ejemplo



Ejemplo



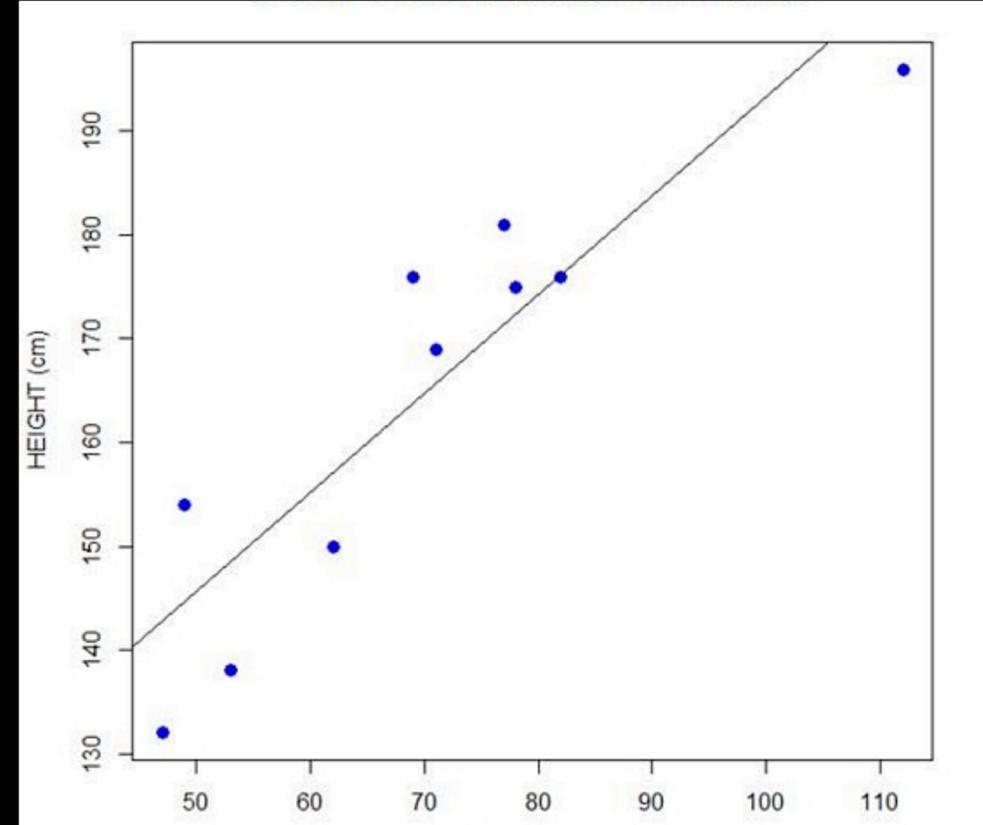
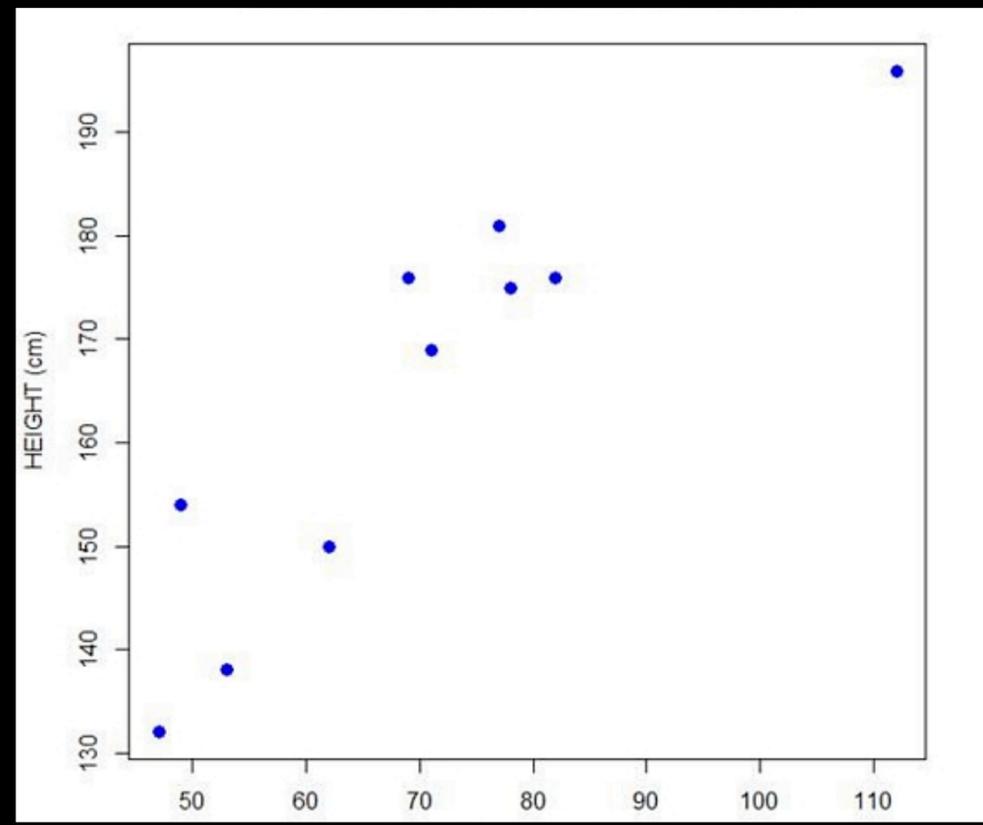
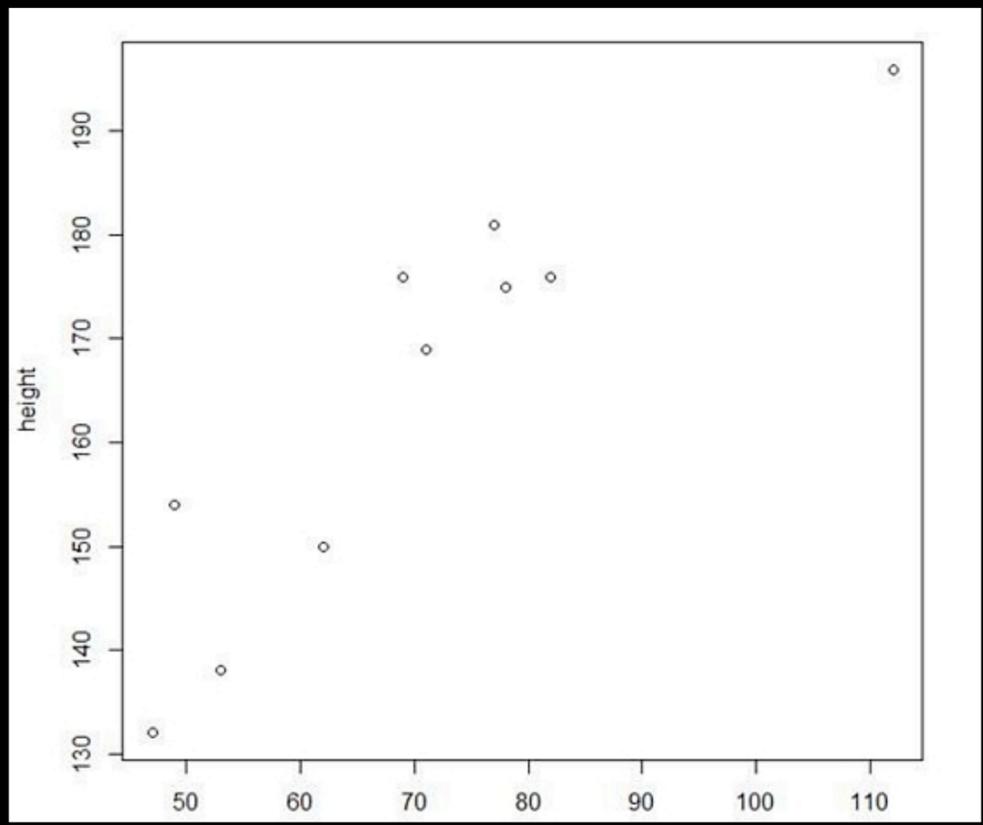
Ejemplo



Los datos suficientemente claros

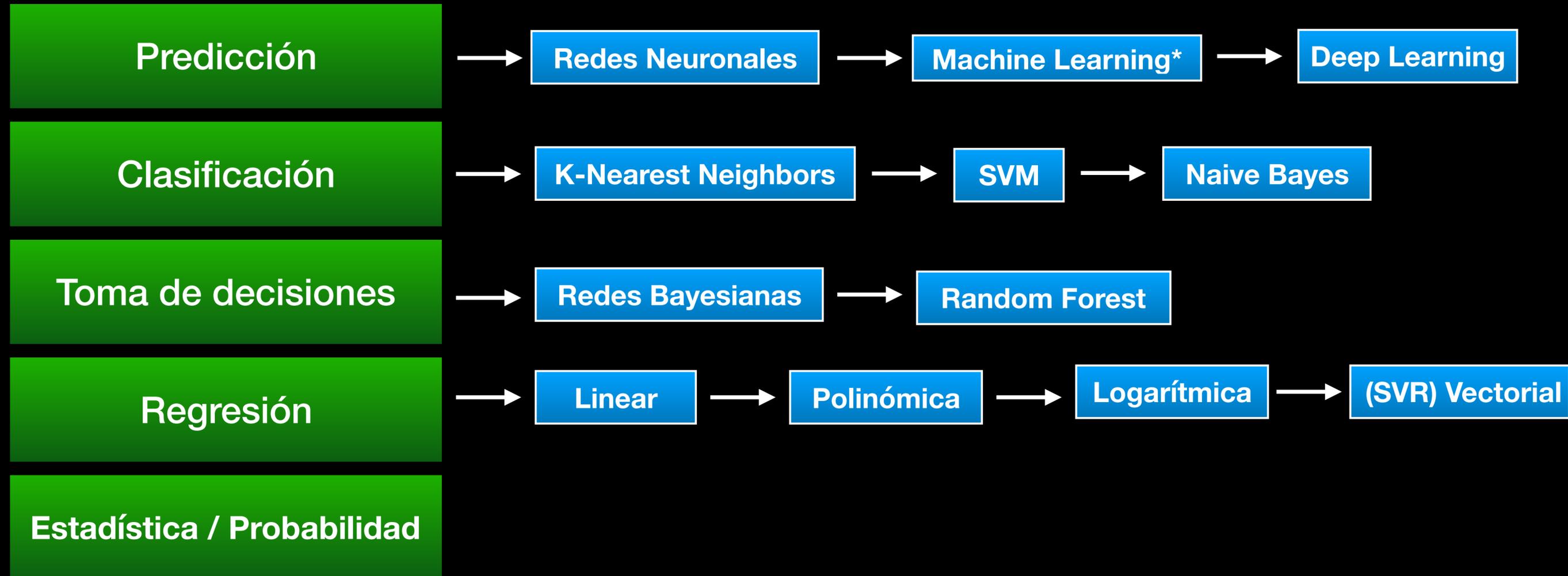
El objetivo es desarrollar un **modelo** que nos permita prever los valores que nos interesa obtener de forma automática.

Nos basamos en datos antiguos para obtener información automática.

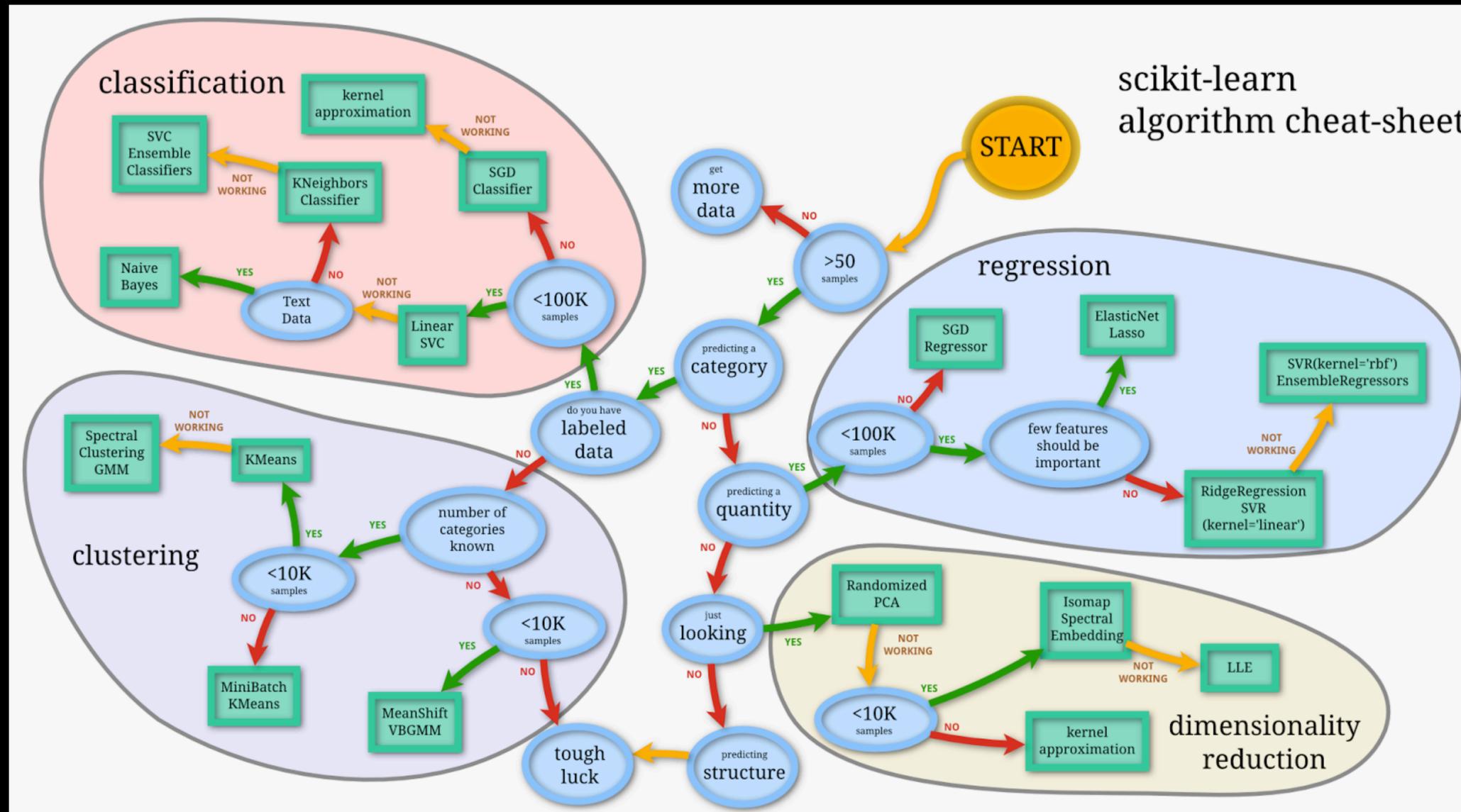


[Tipos de IA]

Tipos de IA



Existen otras clasificaciones...

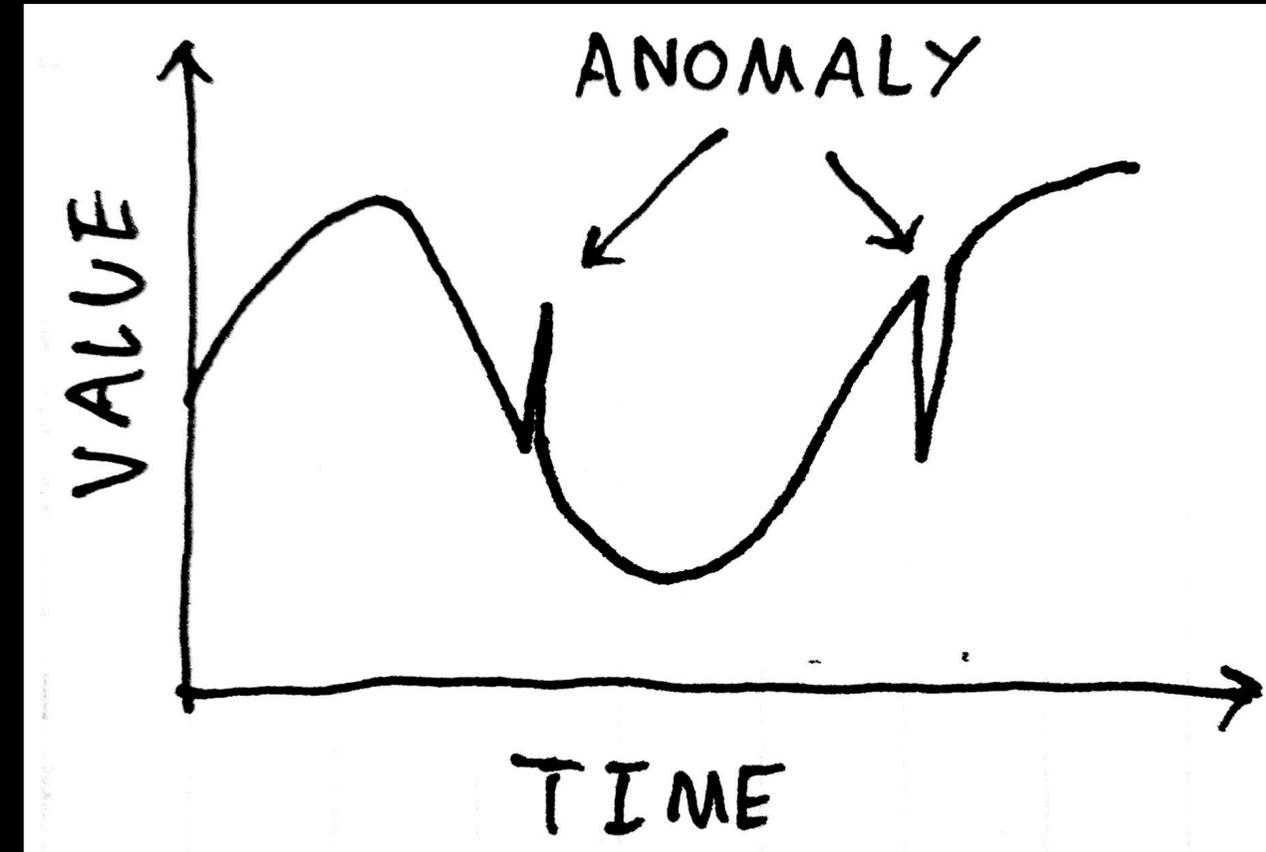
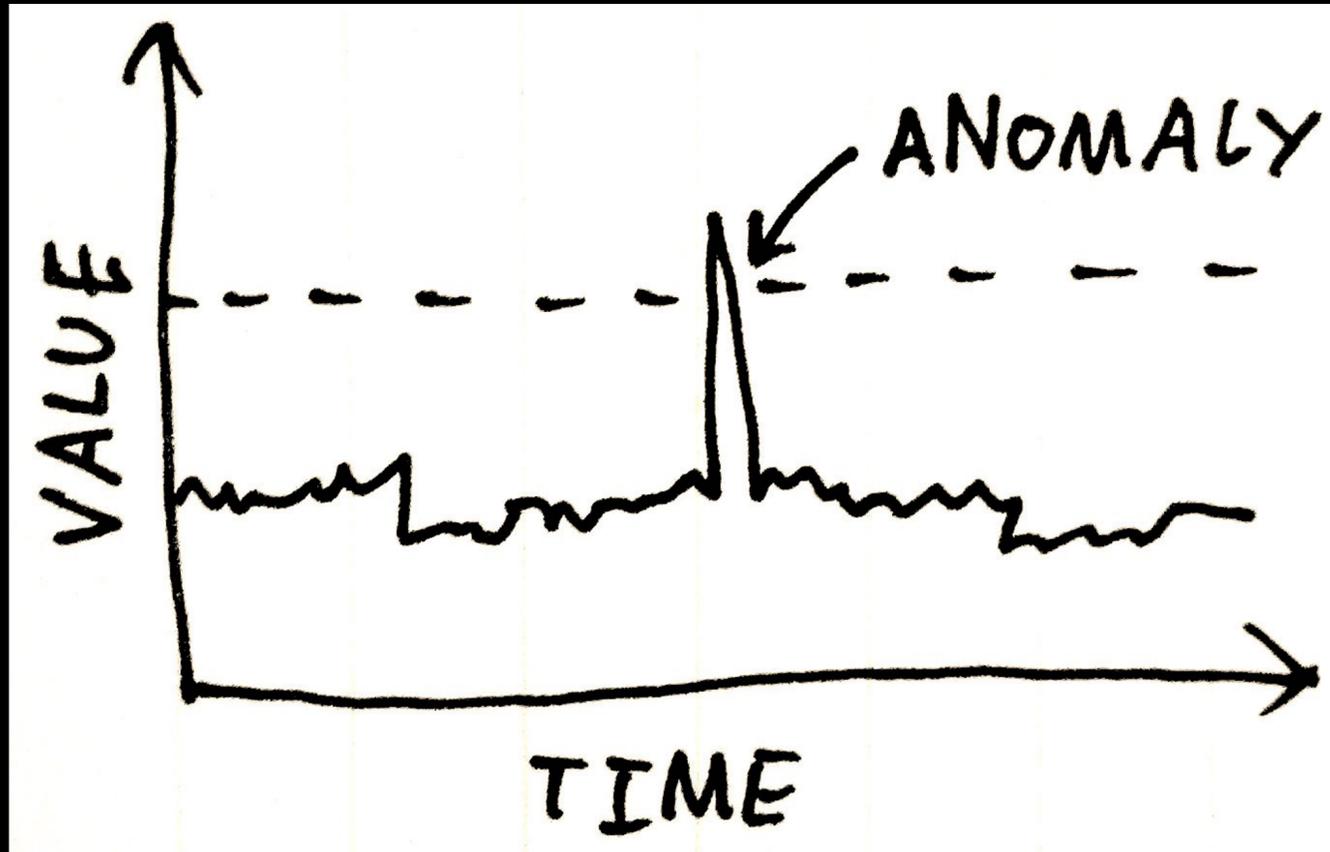


<http://peekaboo-vision.blogspot.com/2013/01/machine-learning-cheat-sheet-for-scikit.html>

[Probabilidad y Estadística]

Probabilidad y Estadística

Detección de anomalías



Demo

Anaconda es un distribución **libre y abierta** de los lenguajes **Python y R**, utilizada en ciencia de datos, y aprendizaje automático (*machine learning*). Esto incluye procesamiento de grandes volúmenes de información, análisis predictivo y cómputos científicos. Está orientado a simplificar el despliegue y administración de los paquetes de software.

Las diferentes versiones de los paquetes se administran mediante el sistema de administración del paquete *conda*, el cual lo hace bastante sencillo de instalar, correr, y actualizar software de ciencia de datos y aprendizaje automático como ser Scikit-team, TensorFlow y SciPy.³

La distribución Anaconda es utilizada por **6 millones de usuarios** e incluye más de 250 paquetes de ciencia de datos válidos para Windows, Linux y MacOS.



Más información: <https://www.anaconda.com>

Demo

jupyter Untitled2 Last Checkpoint: 09/11/2018 (autosaved) Python 3

File Edit View Insert Cell Kernel Widgets Help

Run Code

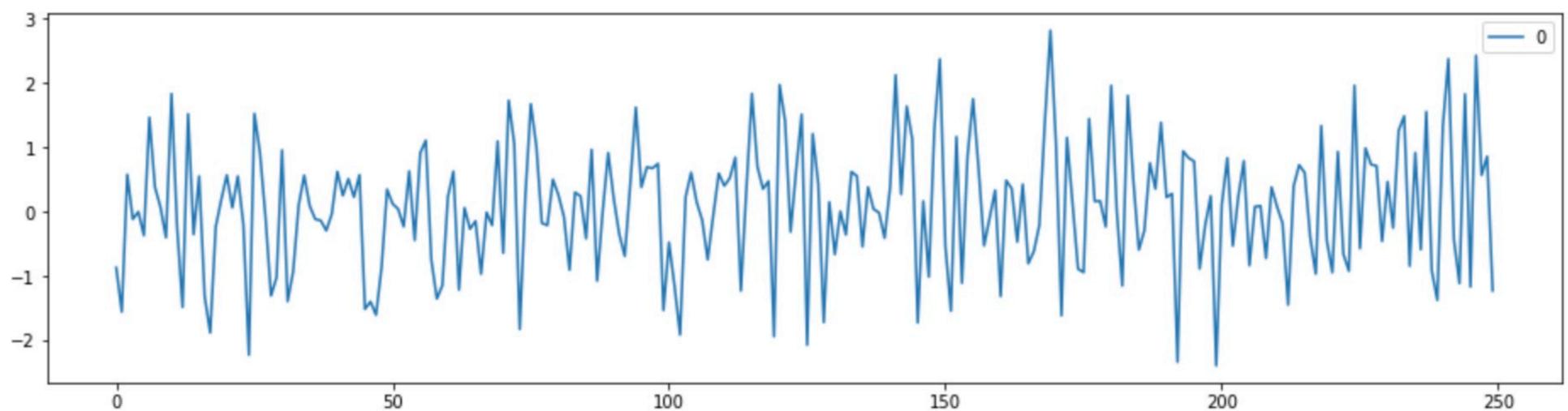
```
In [1]: import pandas as pd
import numpy as np
```

```
In [7]: %matplotlib inline

import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (16.0, 4.0)
```

```
In [12]: data = pd.DataFrame(np.random.randn(250))
data.plot()
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b9bfae80>

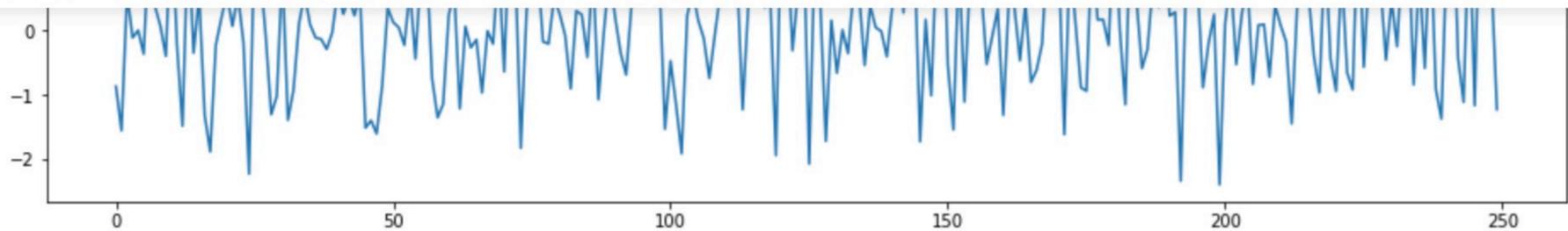


Demo

jupyter Untitled2 Last Checkpoint: 09/11/2018 (autosaved) Python 3 Logout

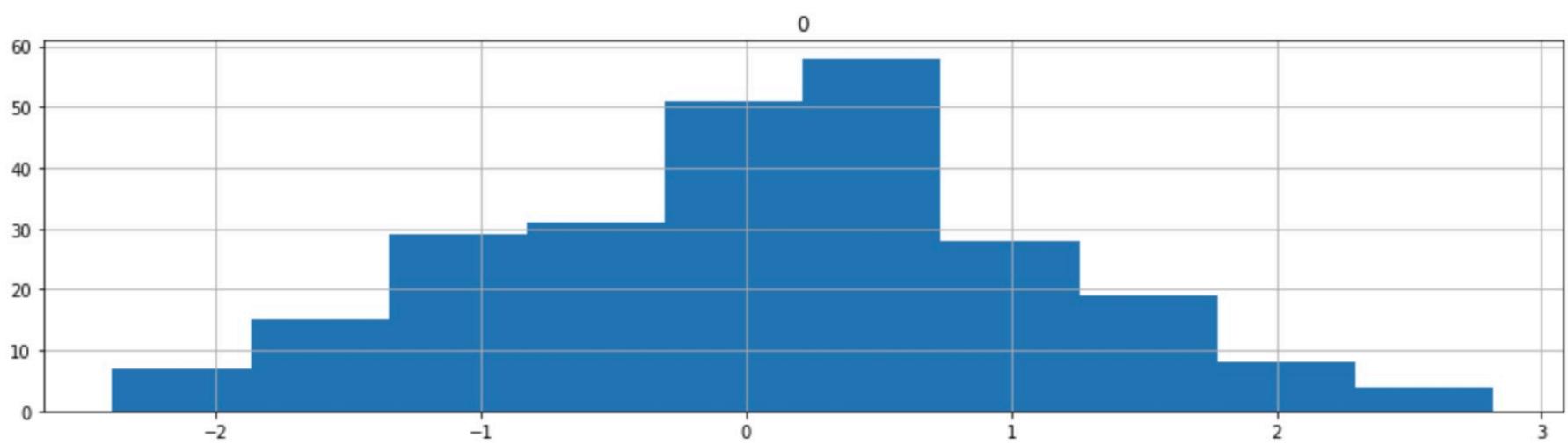
File Edit View Insert Cell Kernel Widgets Help Trusted

Code



In [13]: `data.hist()`

Out[13]: `array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7fe7b83826a0>]], dtype=object)`



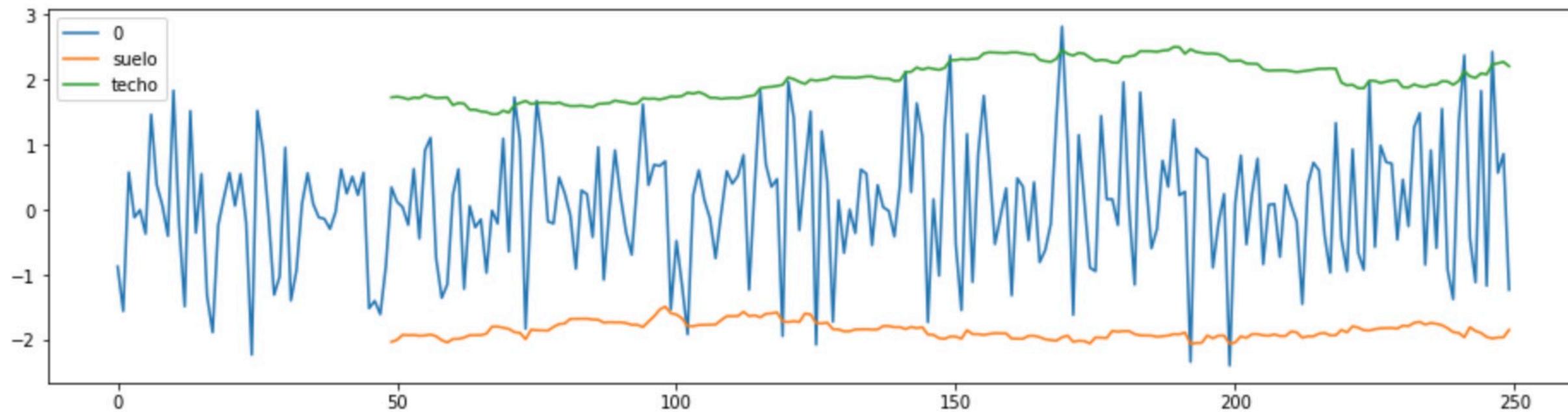
In [14]: `data.std()`

Out[14]: `0 0.989745`
`dtype: float64`

Demo

```
In [21]: wind = 50
sigma = 2
data["suelo"] = data[0].rolling(window=wind).mean() - (sigma * data[0].rolling(window=wind).std())
data["techo"] = data[0].rolling(window=wind).mean() + (sigma * data[0].rolling(window=wind).std())
data.plot()
```

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b2eba208>

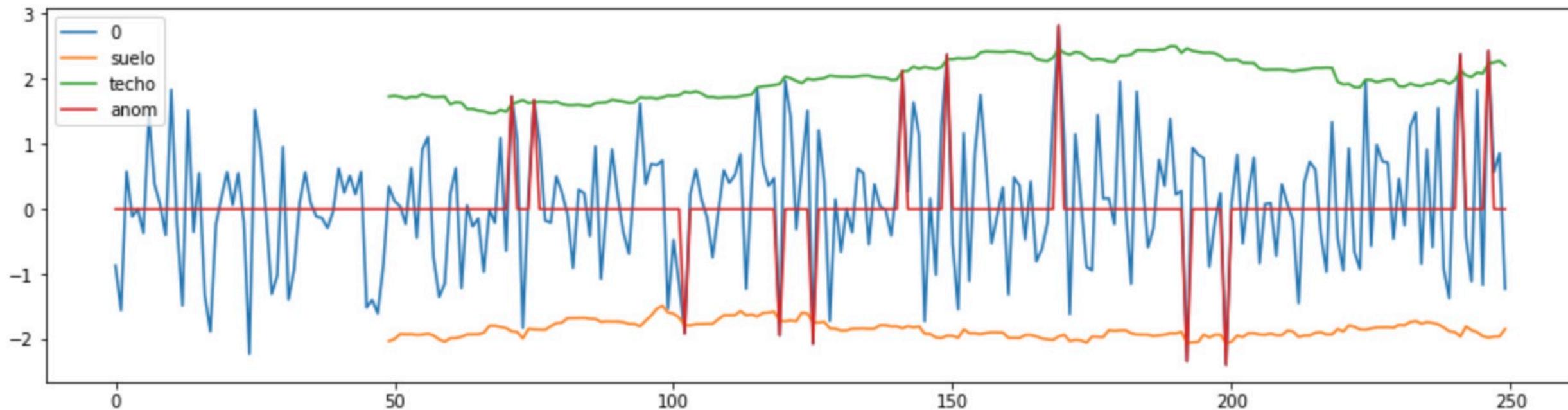


Demo

```
In [22]: data["anom"] = data.apply(  
        lambda row: row[0] if (row[0]<=row["suelo"] or row[0]>=row["techo"]) else 0, axis=1)
```

```
In [23]: data.plot()
```

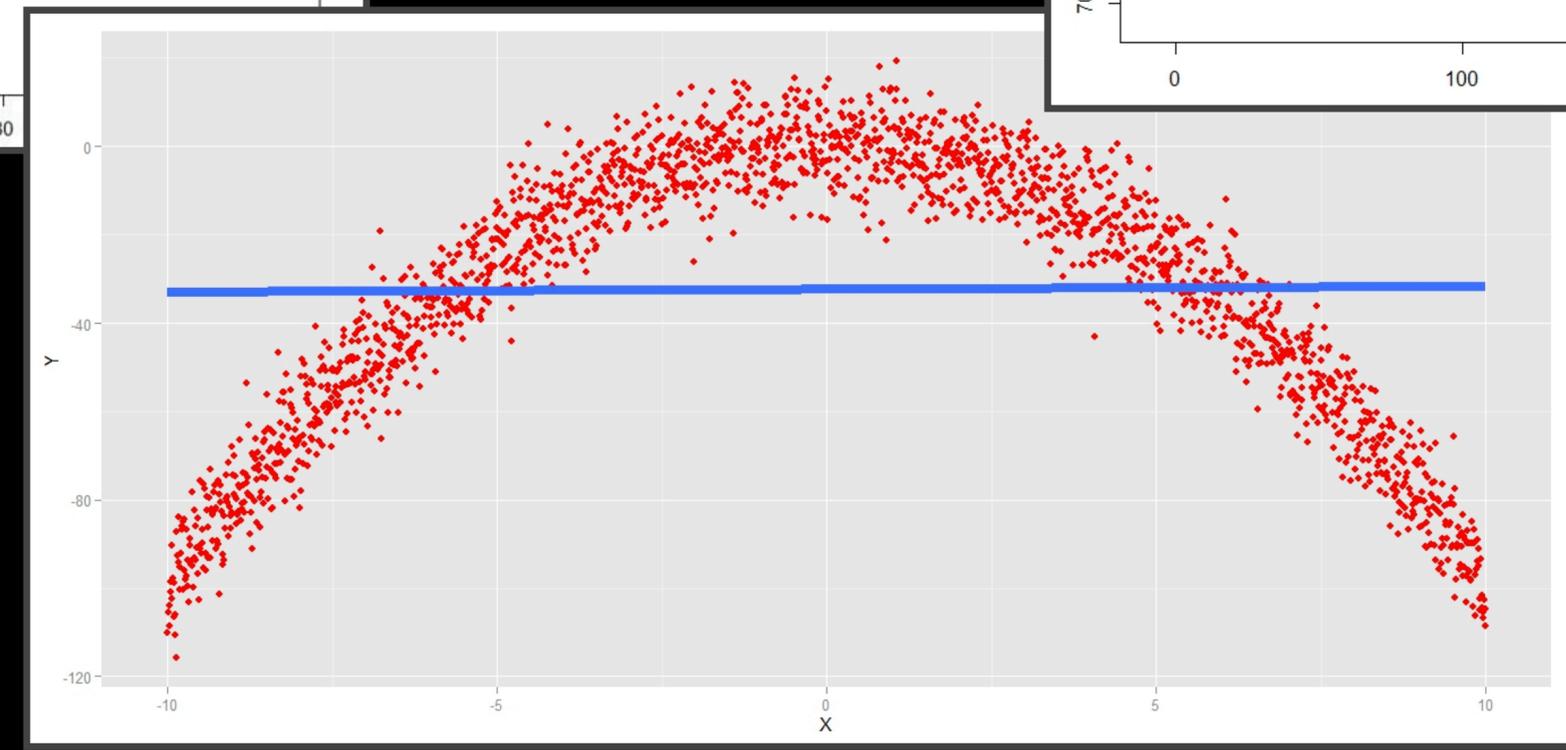
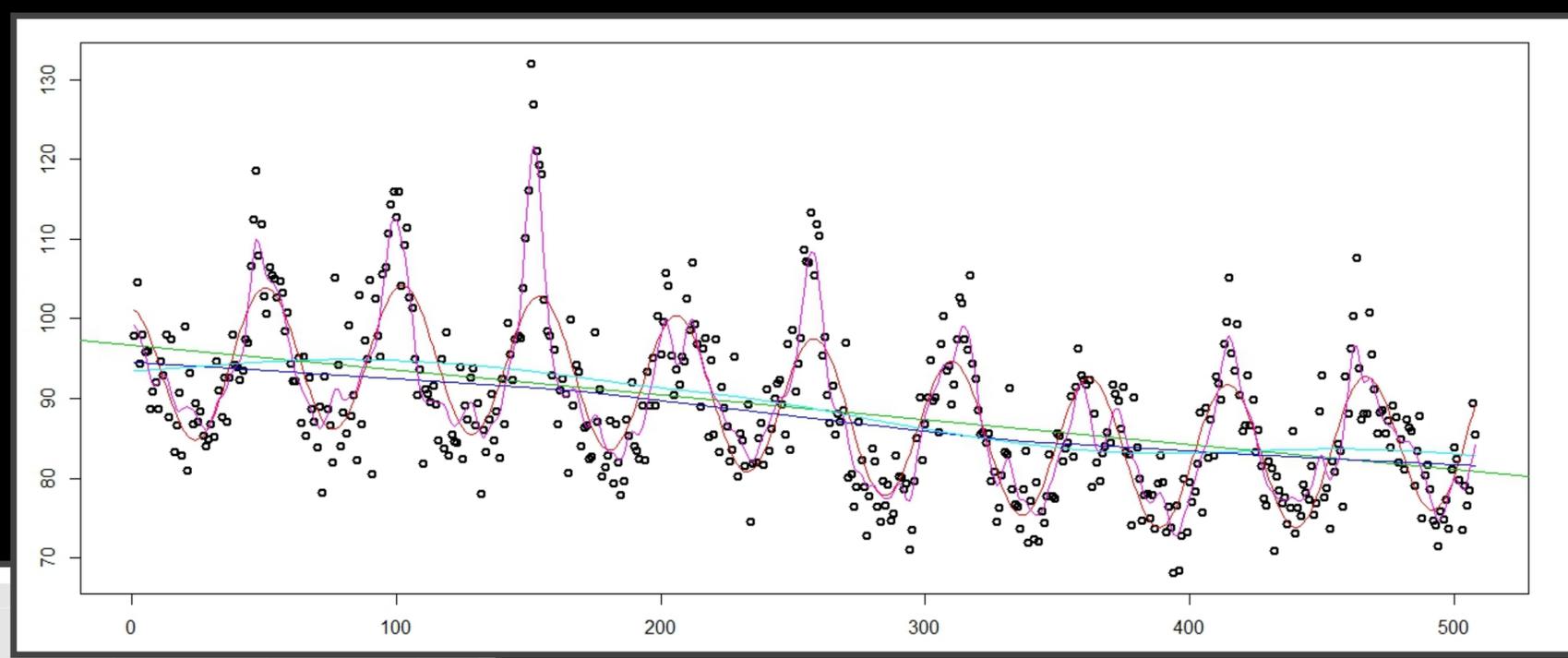
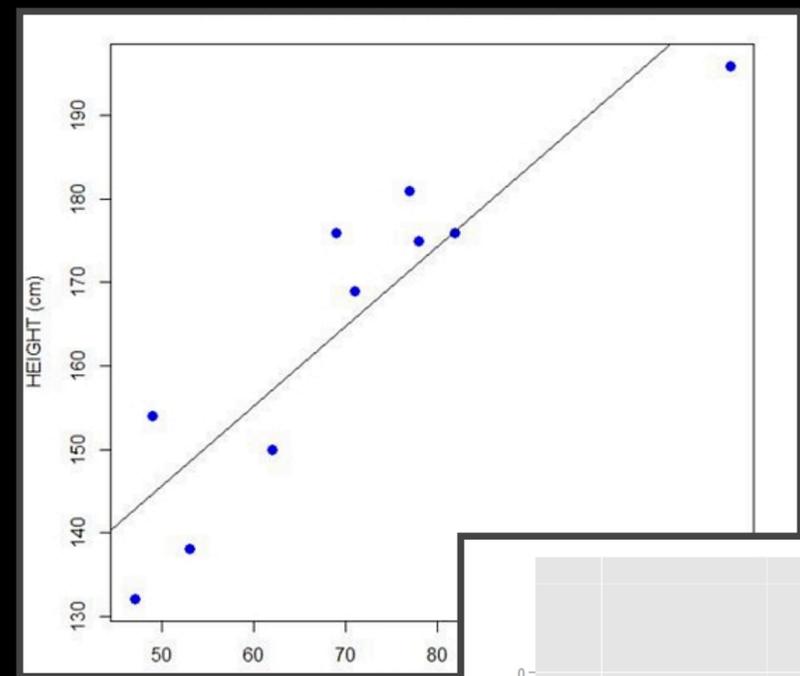
```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b2e0aac8>
```



<http://sukiweb.net/archivos/2018/03/04/deteccion-de-anomalias-con-python-pandas/>

Regresiones

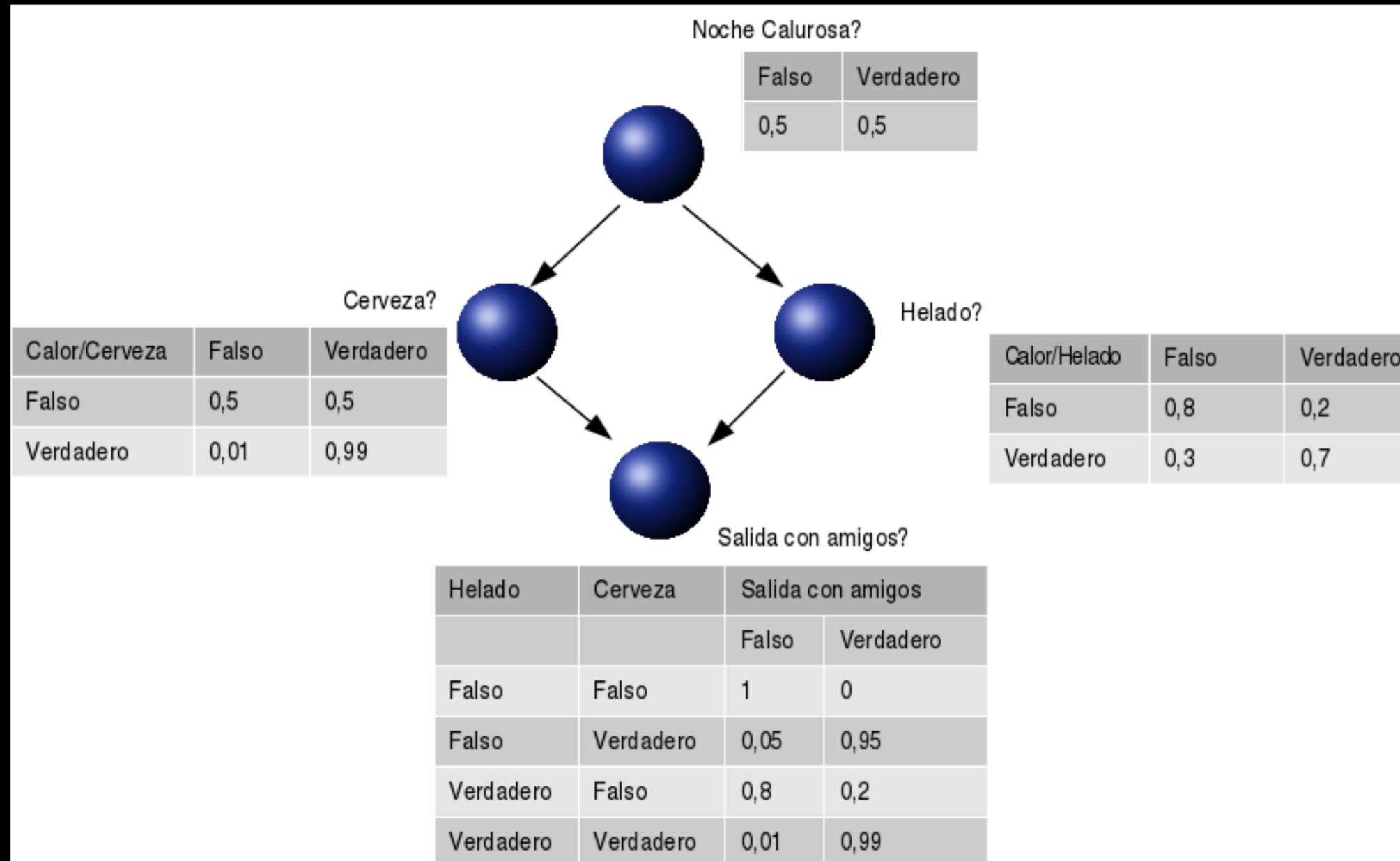
Regresiones



Toma de decisiones

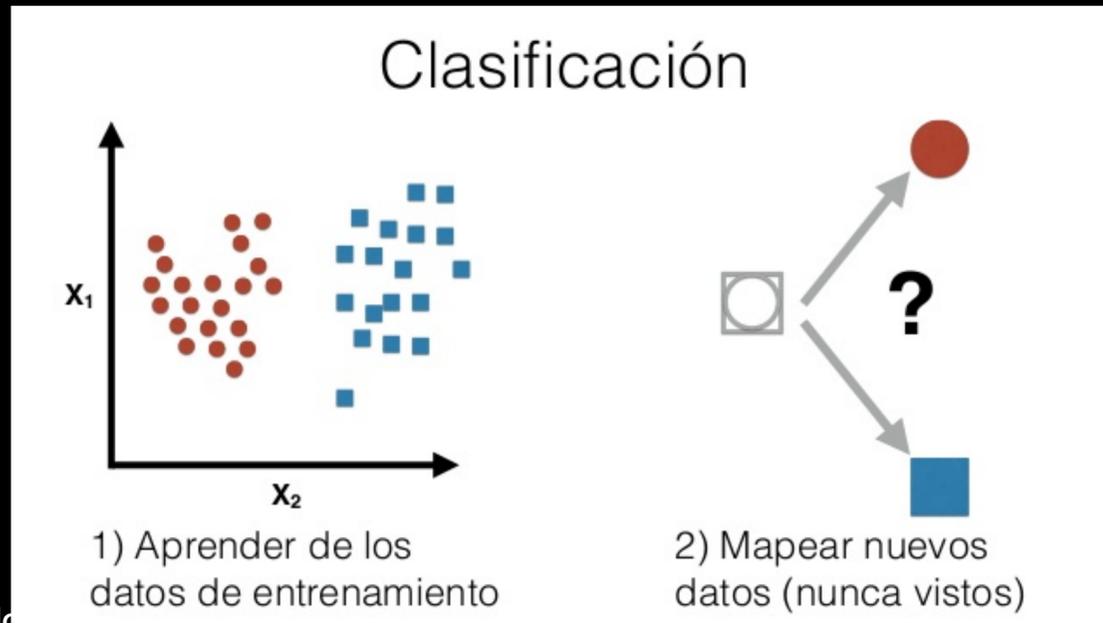
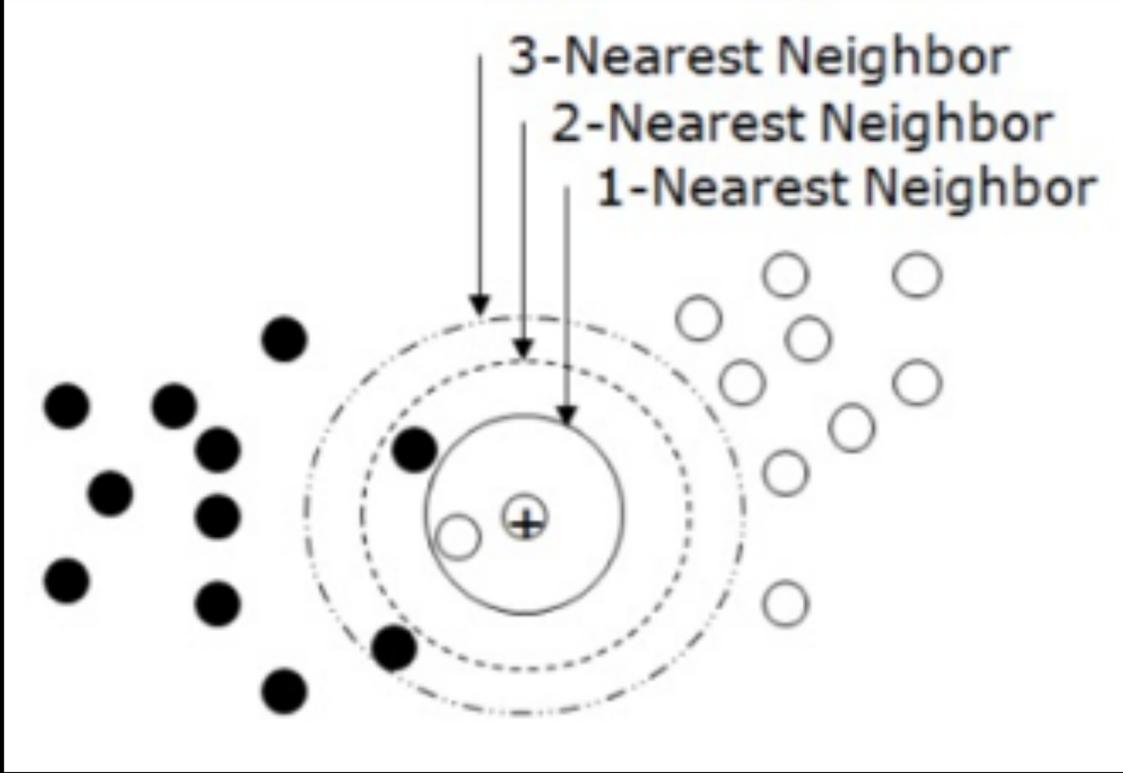
Toma de decisiones

Redes Bayesianas



Clasificación

Clasificación



Credit scoring

1 Necesitamos una base de datos con los clientes

Ciente	Ingresos Mes	Nivel Estudios	Score
Ana	1.100 €	1	Malo
Luis	983 €	0	Malo
Pedro	1.203 €	1	Malo
Diego	995 €	3	Malo
Lucas	725 €	1	Malo
José	1.750 €	2	Malo
Manuel	2.500 €	5	Bueno
Tere	2.650 €	4	Bueno
Sofía	1.700 €	3	Bueno
Manu	1.321 €	4	Bueno
Johr	2.311 €	3	Bueno
Peter	1.512 €	5	Bueno
Xiao	2.432 €	2	Bueno
Art	2.550 €	1	Bueno
Cesar	1.567 €	4	Bueno
Marga	1.250 €	5	Bueno
Pere	1.003 €	2	Bueno
Tania	2.005 €	5	Bueno
Silvia	1.580 €	2	Bueno
Flor	1.350 €	3	Bueno
Diana	1.301 €	2	Bueno

2 Implementaremos el algoritmo K-NN

Vecino más cercano

3 ¿Concedemos crédito a Juan?

Juan

- Ingresos mes: 1.800 euros
- Nivel de estudios, de 0 a 5
- 0=Sin estudios
- 1=Básico
- 2=Secundaria
- 3=Bachiller
- 4=Graduado Universidad
- 5=Máster

Clasificación

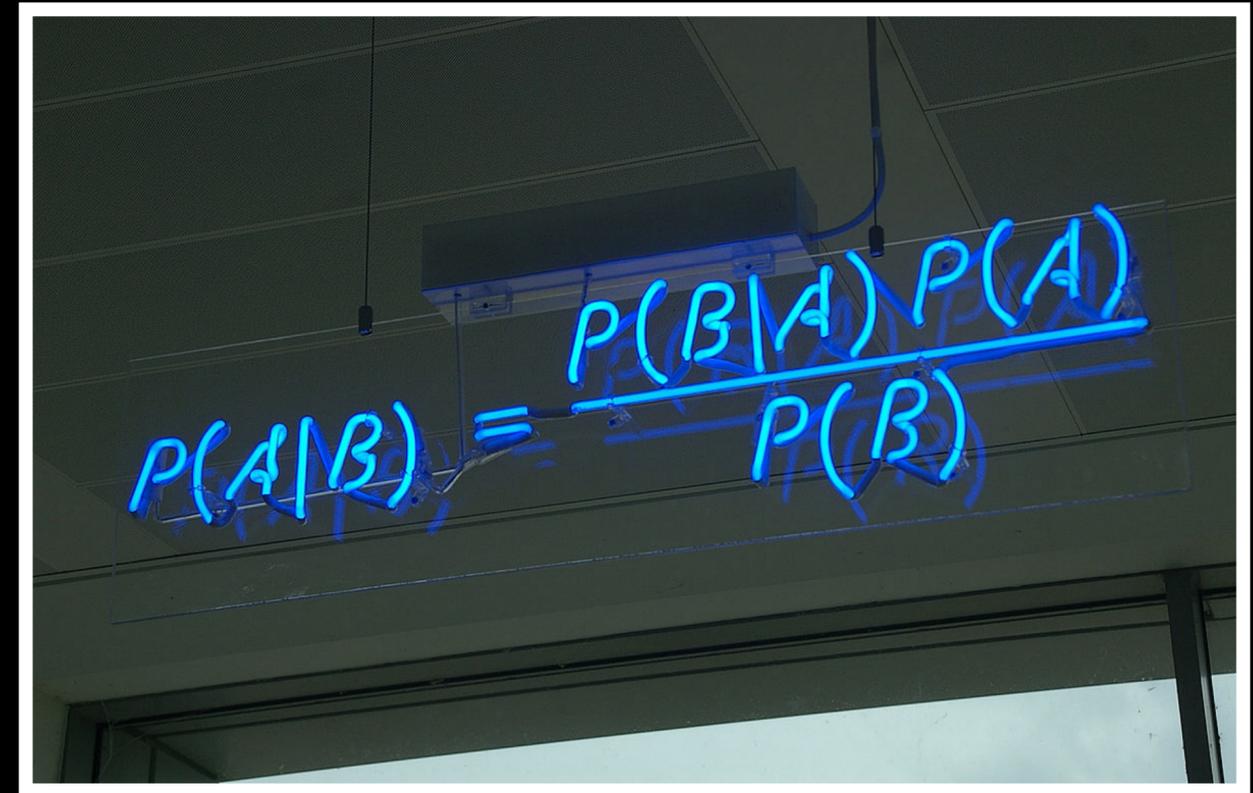
Naive Bayes

¿Te suena SpamAssassin?

Cálculo de la probabilidad de que un mensaje que contiene una palabra dada es spam

Combinaciones de probabilidades individuales

Heurística


$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Predicción

Predicción

Machine Learning



Predicción

Literatura ?

MOTHERBOARD Moveable Hacking Environment Space Gaming Health Tech Science Influence

A Neural Network Wrote the Next 'Game of Thrones' Book Because George R.R. Martin Hasn't

"The Winds of Winter" is already here ... sorta.

SHARE TWEET



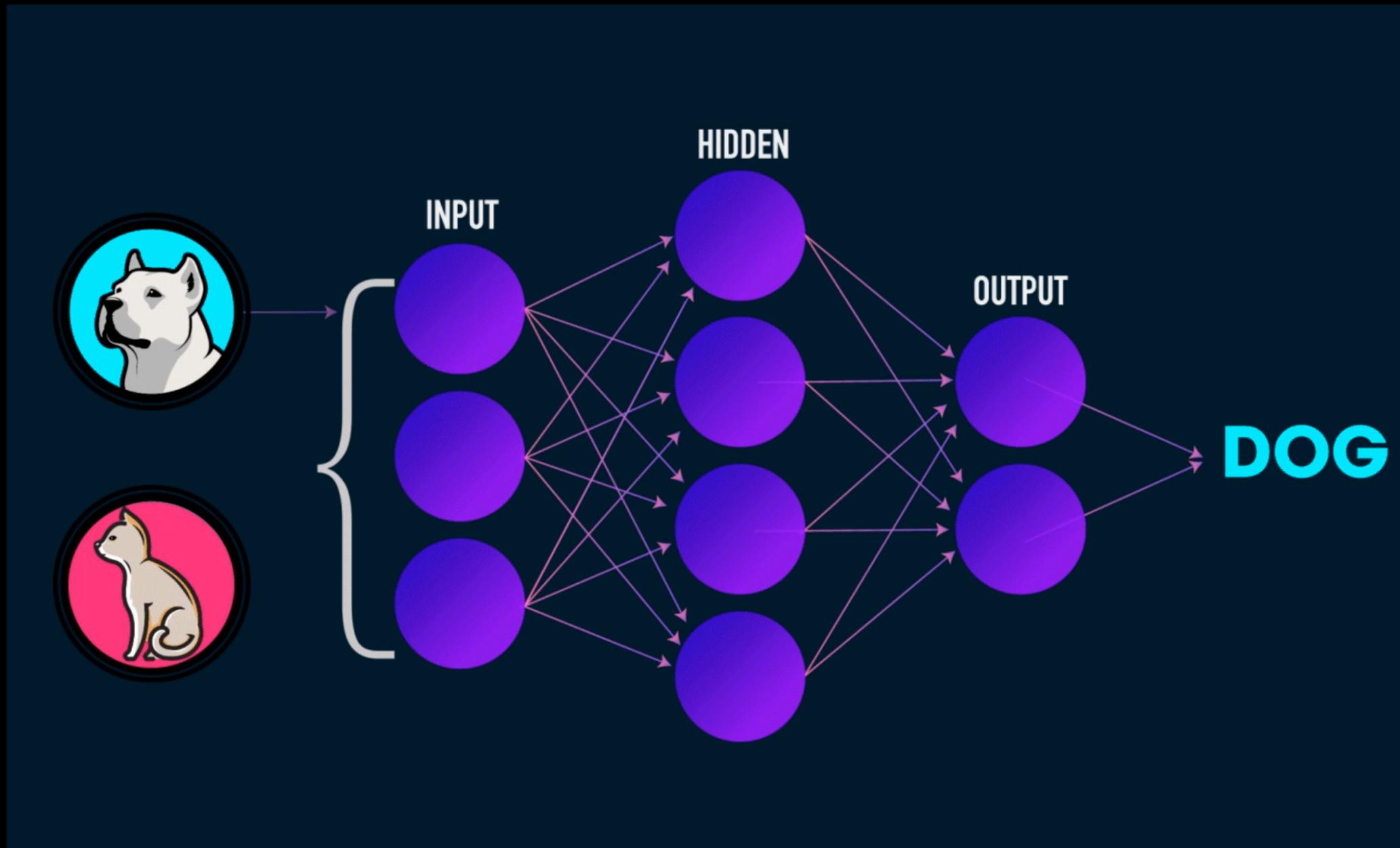
Immagine: HBO/Shutterstock. Composizione: Jason Koebler



Arte ?



Predicción



[BigData]

Machine Learning necesita datos



¿Cuántos datos?

Cuantos más ejemplos le enseñemos al modelo, más acertada será la predicción.

¿Cuántos datos?

MySQL no es muy práctico aquí:

the data cache stored in RAM can grow very large and be subjected to thousands or even millions of requests per second.

Hacen falta motores muy orientado a la entrega rápida de datos con muy bajo consumo y posibilidad de redundancia, escalabilidad y distribución

Volumen

Variiedad

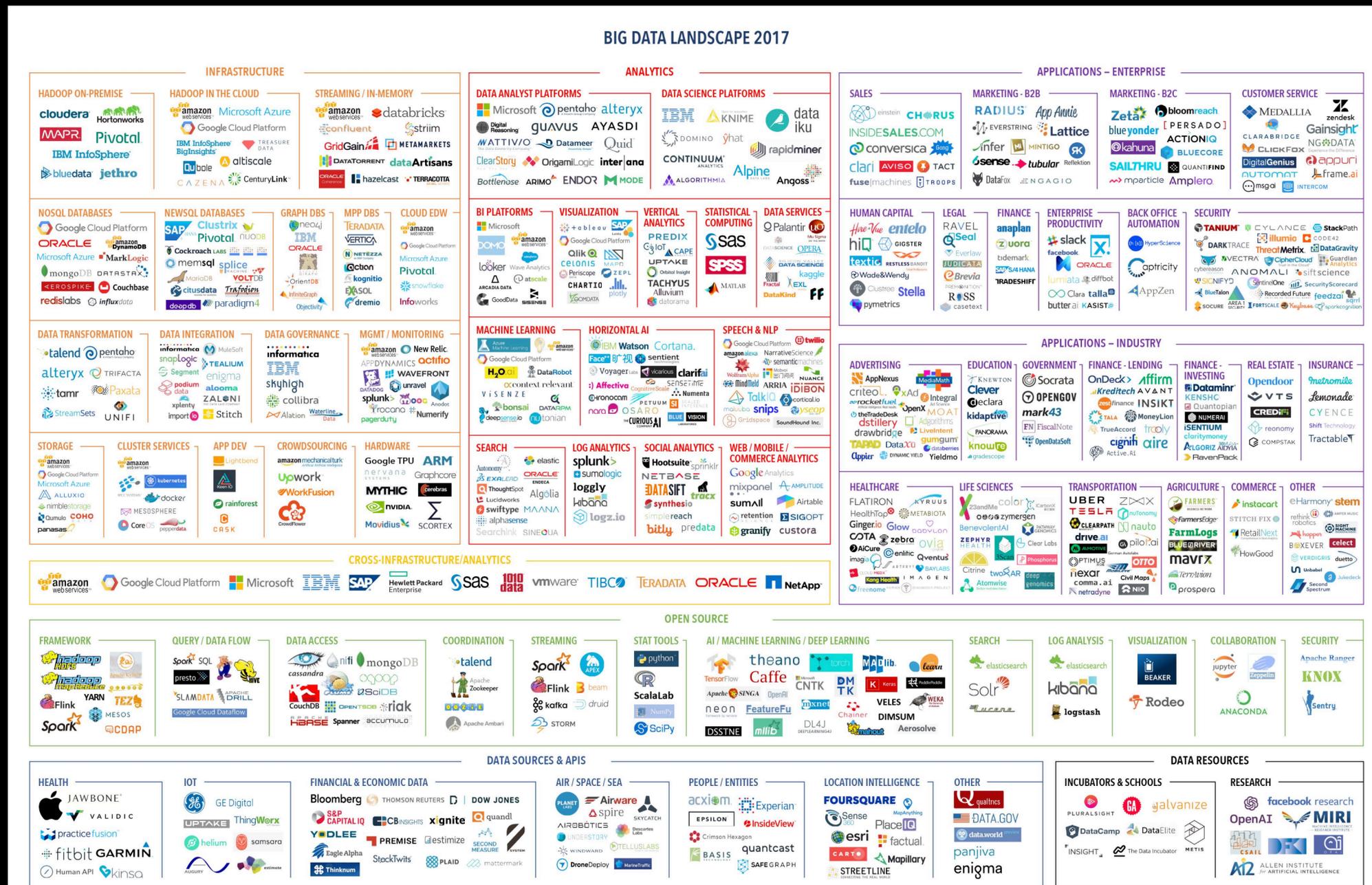
Velocidad

Veracidad

Variabilidad

Valor

Bases de datos BigData hay muchas



¿Mi Asterisk™ es compatible con eso?

Todos los motores disponen de sistemas para insertar datos de una forma fácil y cómoda.

Asterisk no tiene un módulo para enviar datos a la base de datos, pero podemos implementarlo muy fácilmente.

Ejemplo

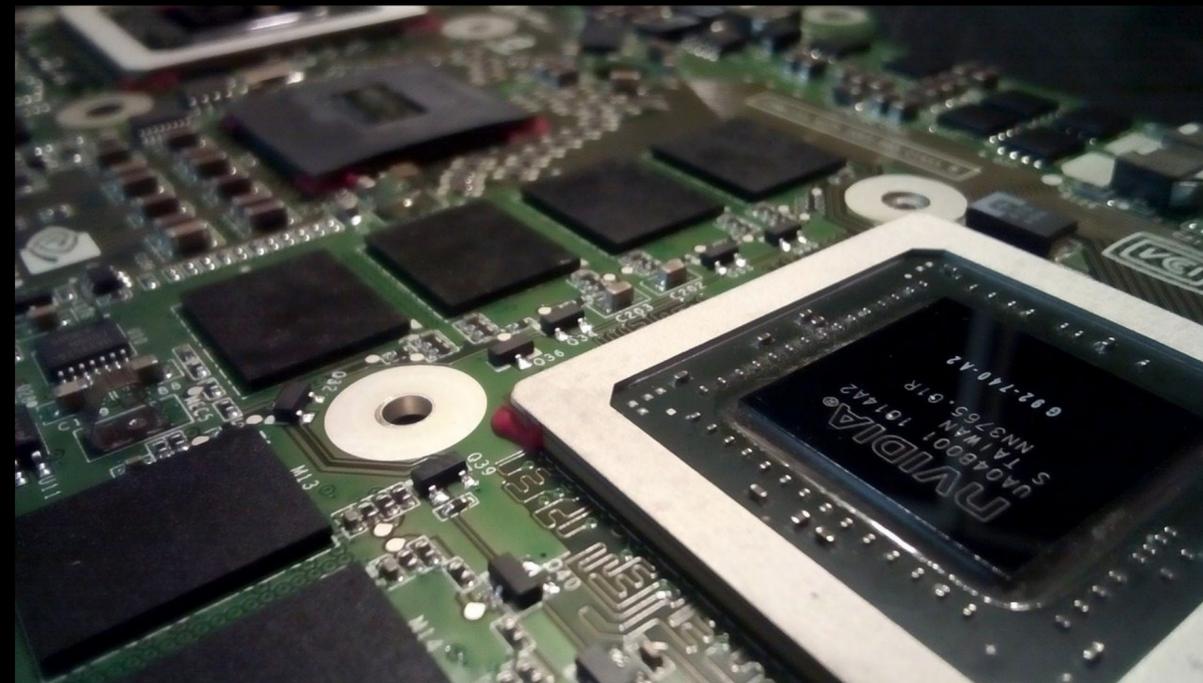
```
#!/usr/bin/env node
var MongoClient = require('mongodb').MongoClient;
var ami = require('asterisk-manager')(5038, 'localhost', 'amiuser', 'amip4ssw0rd', true);
var url = "mongodb://localhost:27017/";
ami.on('cdr', function(evt) {
  MongoClient.connect(url, function(err, db) {
    if (err) throw err;
    var dbo = db.db("cdr");
    myobj = {
      date: evt.starttime,
      src: evt.source,
      dst: evt.destination,
      res: evt.disposition,
      time: parseInt(evt.billableseconds),
    };
    dbo.collection('cdr').insertOne(myobj, function(err, res) {
      if (err) throw err;
      db.close();
    });
  });
});
});
```

[Revolucion de la IA]

Machine Learning
Deep Learning

¿Por qué ahora está en auge?

Nuevos procesadores hacen millones de pruebas por segundo.



¿Por qué ahora está en auge?

Nuevas librerías y herramientas facilitan su accesibilidad.



IP[y]: IPython
Interactive Computing



PYTORCH

theano

gensim

Caffe



SM

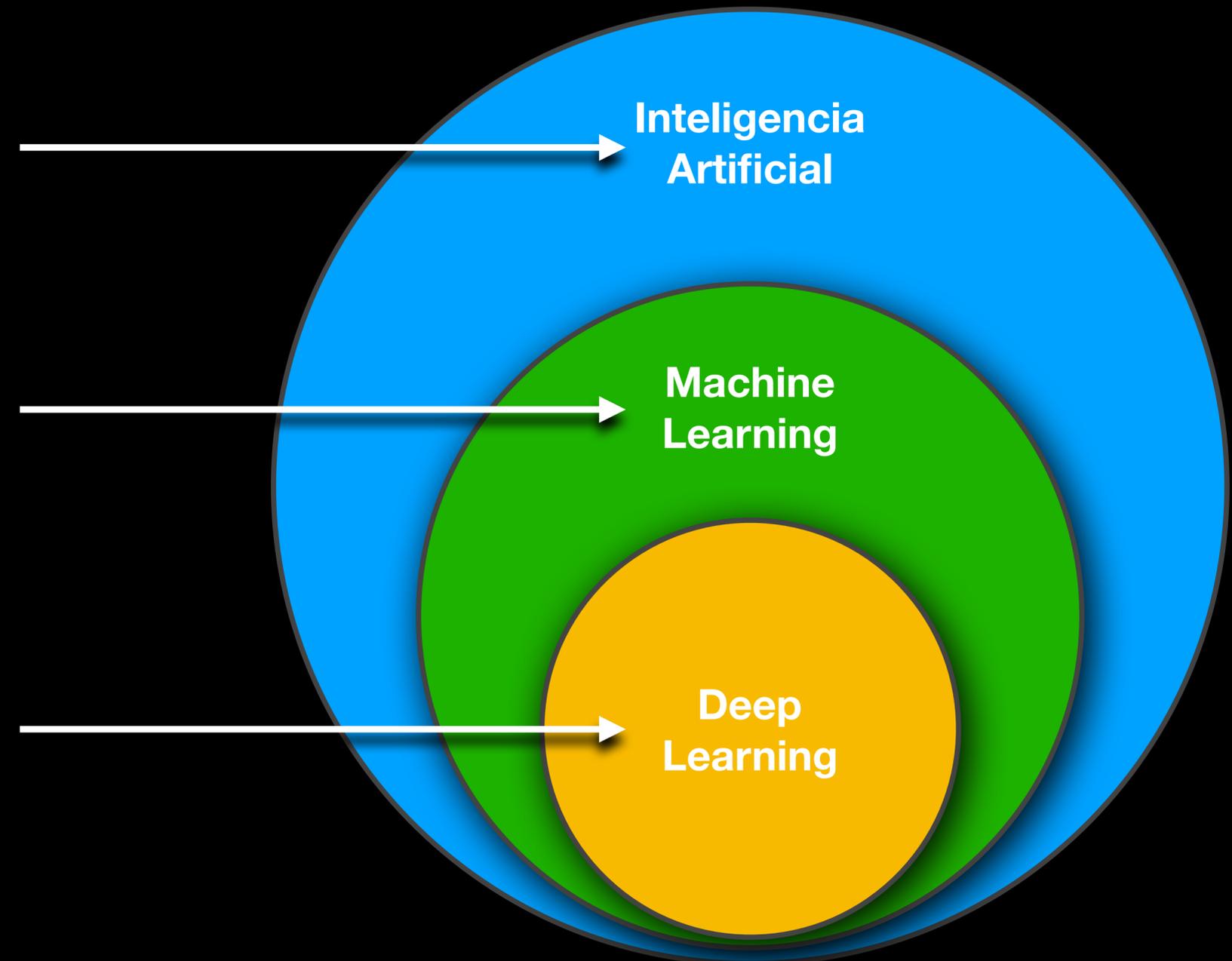
將軍
sho gun

Machine Learning vs Deep Learning

Técnica que permite a un sistema
IMITAR
el comportamiento humano

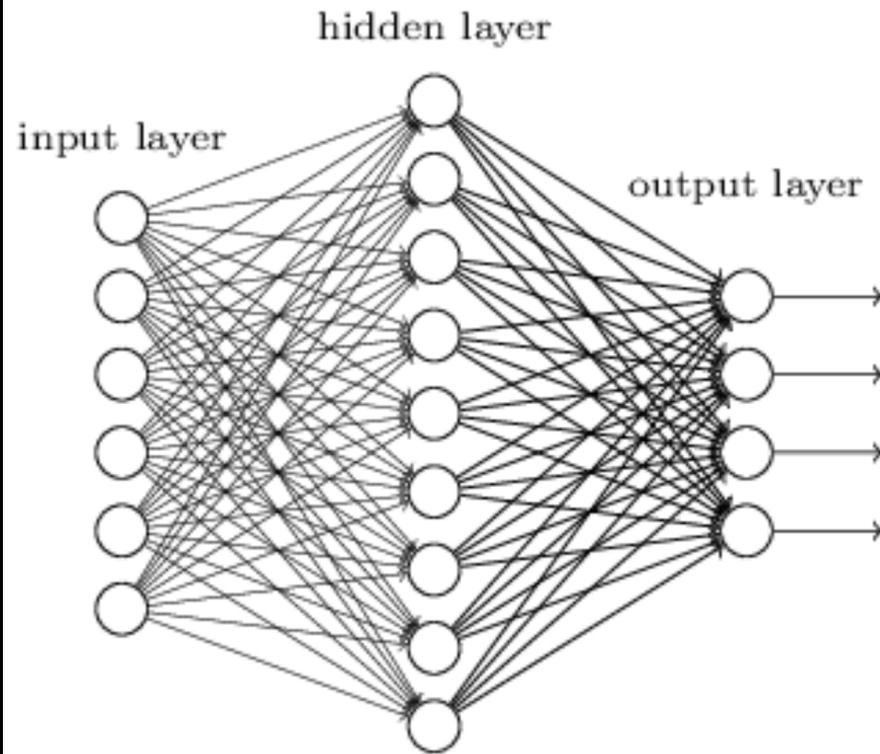
Técnica que permite a un sistema
MEJORAR
en base a la experiencia

Técnica que permite a un sistema
PERFECCIONAR
en base a su propia experiencia

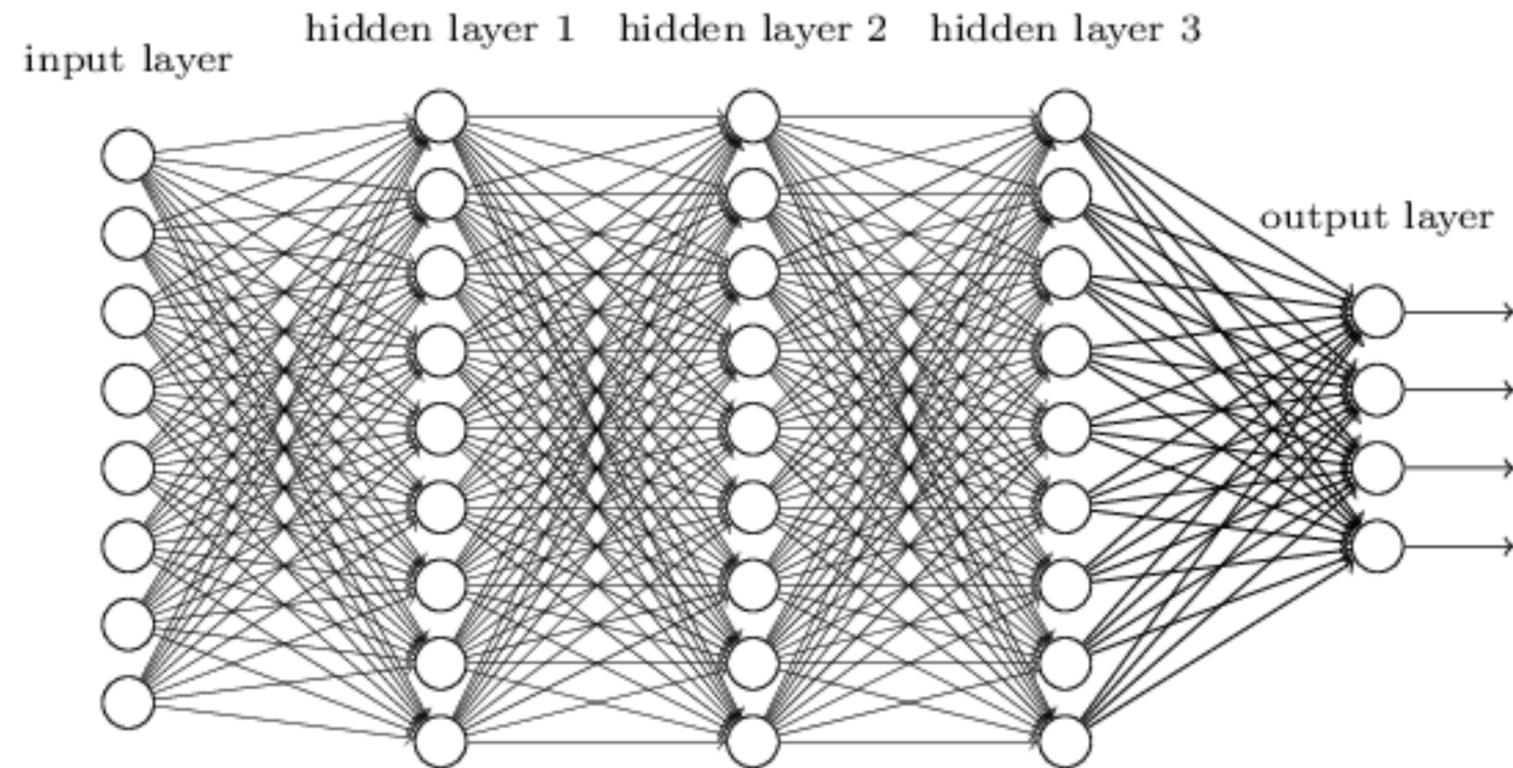


Machine Learning vs Deep Learning

"Non-deep" feedforward neural network



Deep neural network



DEMO

TensorFlow es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos.

Todos los productos de Google utilizan alguna característica de Inteligencia Artificial.

TensorFlow está originalmente desarrollado en Python, aunque hay una versión para Javascript.

<https://www.tensorflow.org>



DEMO

```
<html>
<head>
<!-- Load TensorFlow.js -->
<!-- Get latest version at https://github.com/tensorflow/tfjs -->
<script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs@0.11.2">
</script>
<style>body{ font-size: 50px; }</style>
</head>
<body>
  <div id="output_field"></div>
</body>
<script>
async function learnLinear(){
  const model = tf.sequential();
  model.add(tf.layers.dense({units: 1, inputShape: [1]}));
  model.compile({
    loss: 'meanSquaredError',
    optimizer: 'sgd'
  });

  const xs = tf.tensor2d([-1, 0, 1, 2, 3, 4], [6, 1]);
  const ys = tf.tensor2d([-3, -1, 1, 3, 5, 7 ], [6, 1]);

  await model.fit(xs, ys, {epochs: 500});

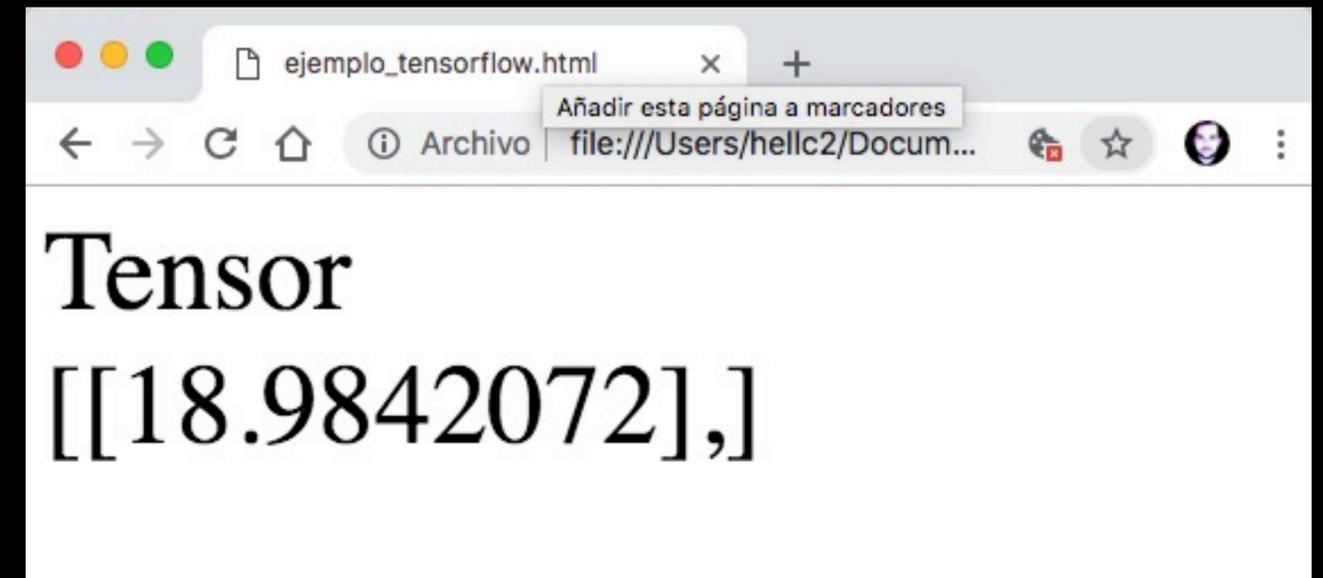
  document.getElementById('output_field').innerText =
  model.predict(tf.tensor2d([10], [1, 1]));
}
learnLinear();
</script>
</html>
```

Entrenamiento...

```
const xs = tf.tensor2d([-1, 0, 1, 2, 3, 4], [6, 1]);
const ys = tf.tensor2d([-3, -1, 1, 3, 5, 7 ], [6, 1]);
```

Solicitamos una predicción...

```
document.getElementById('output_field').innerText =
model.predict(tf.tensor2d([10], [1, 1]));
```

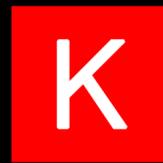


Machine Learning vs Deep Learning



[tensorflow.org](https://www.tensorflow.org)

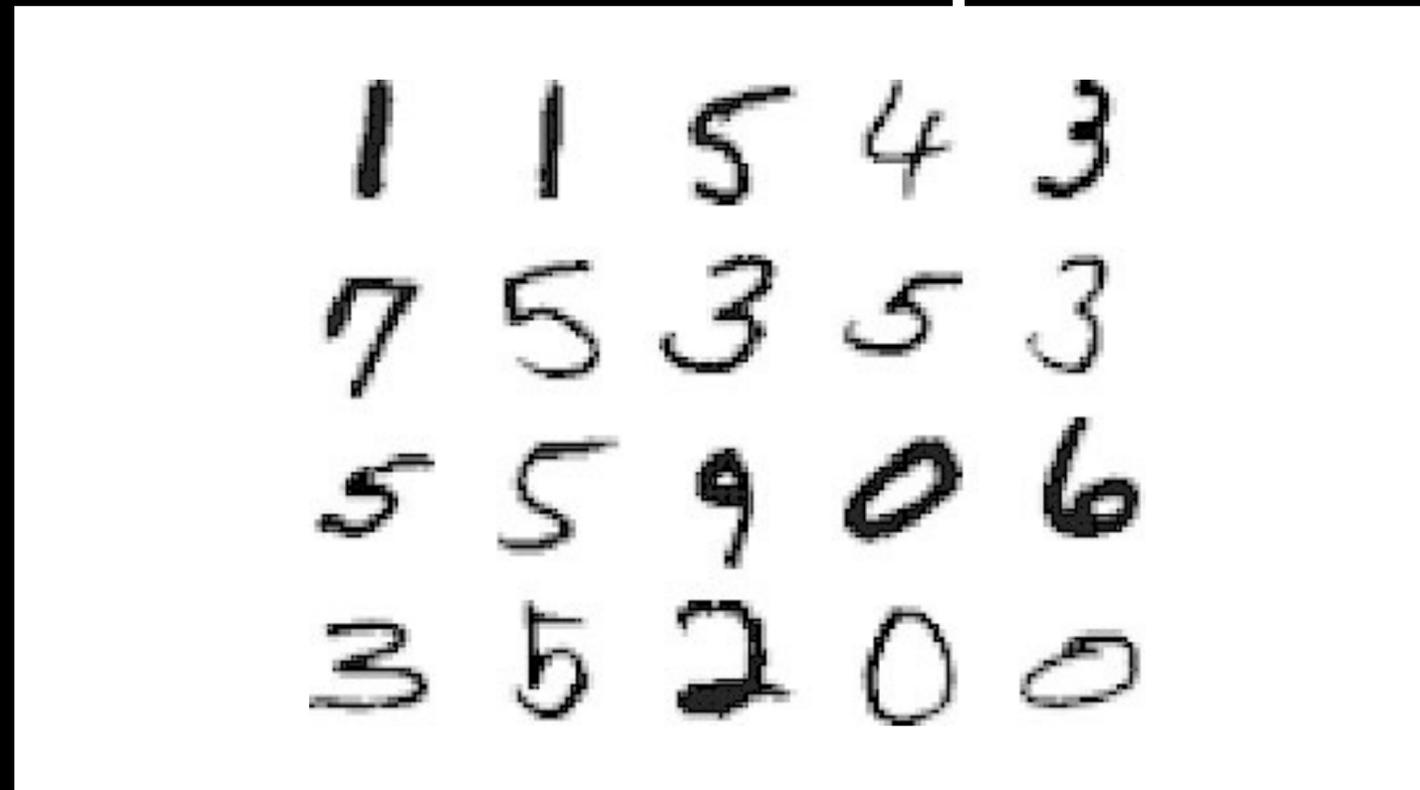
+



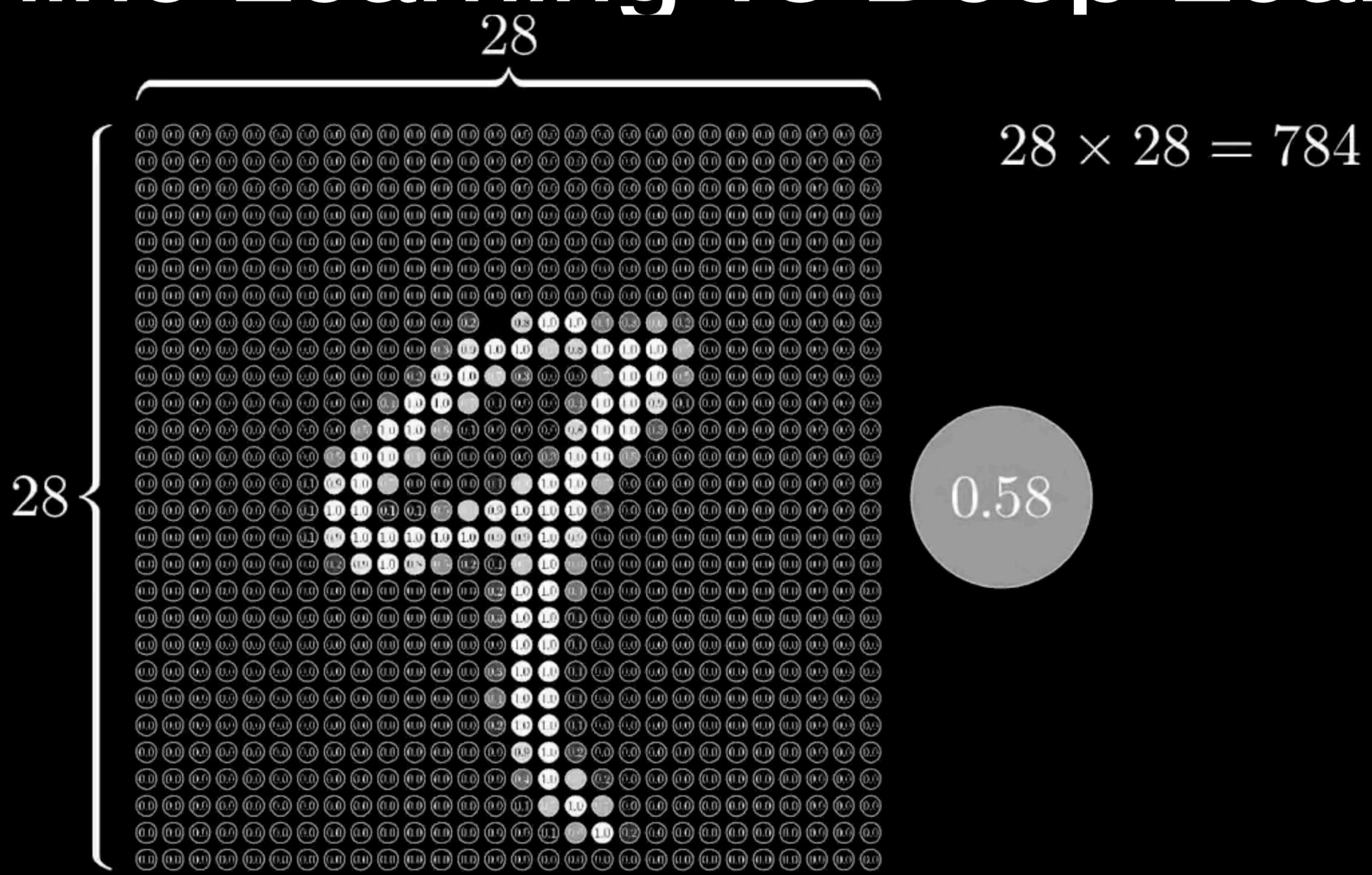
Keras

keras.io

Reconocimiento de patrones



Machine Learning vs Deep Learning

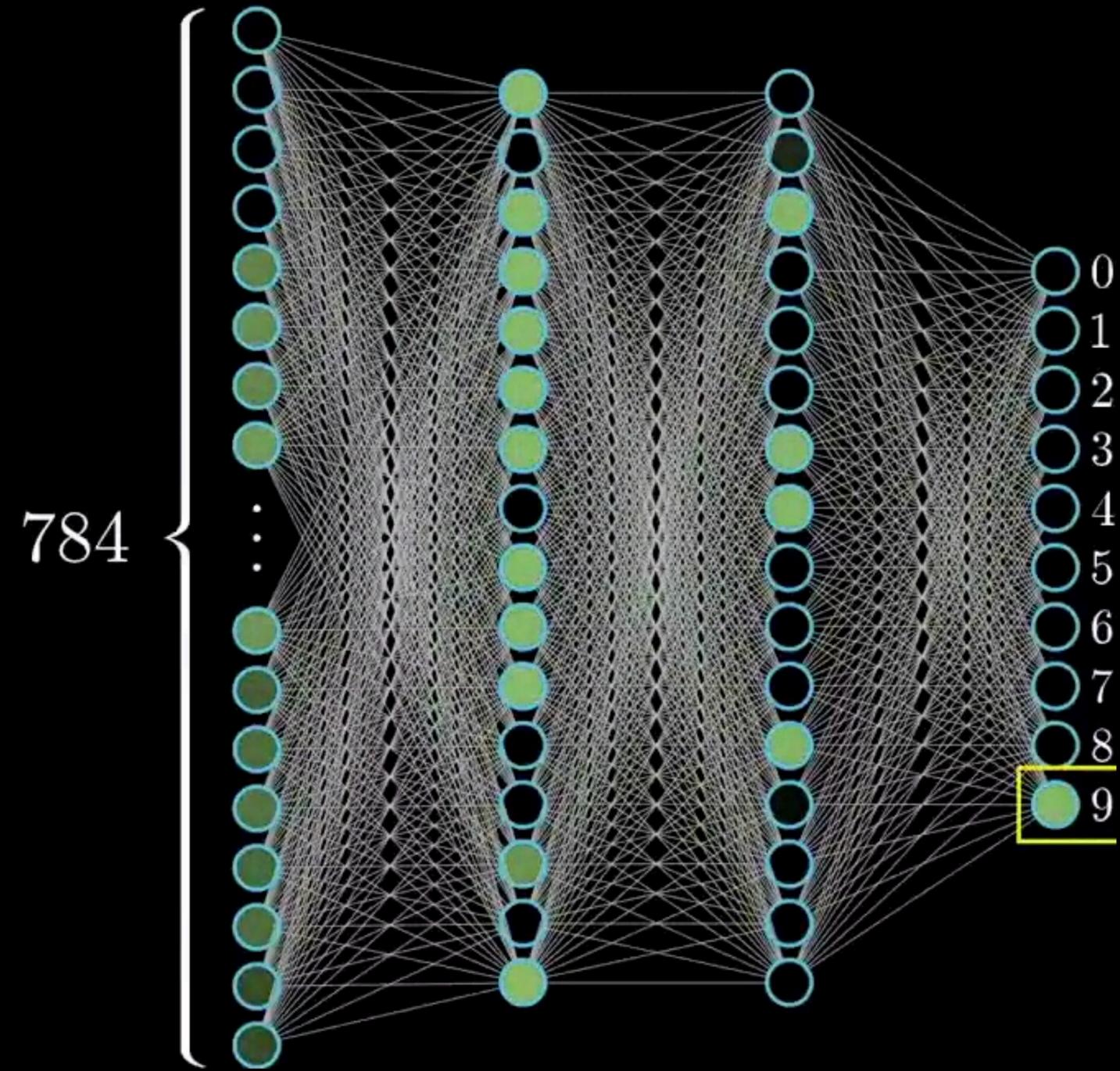


Machine Learning vs Deep Learning

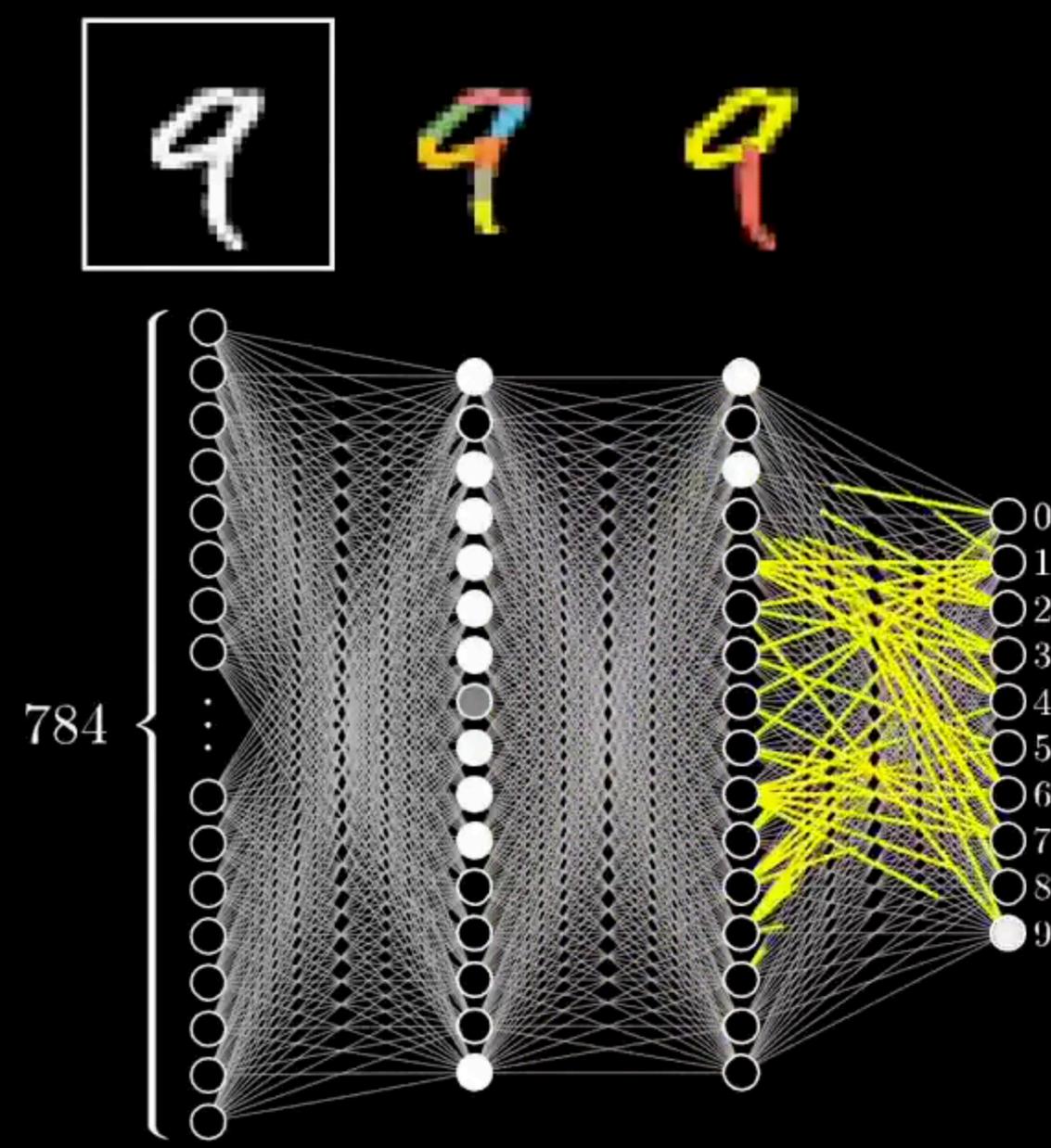
784



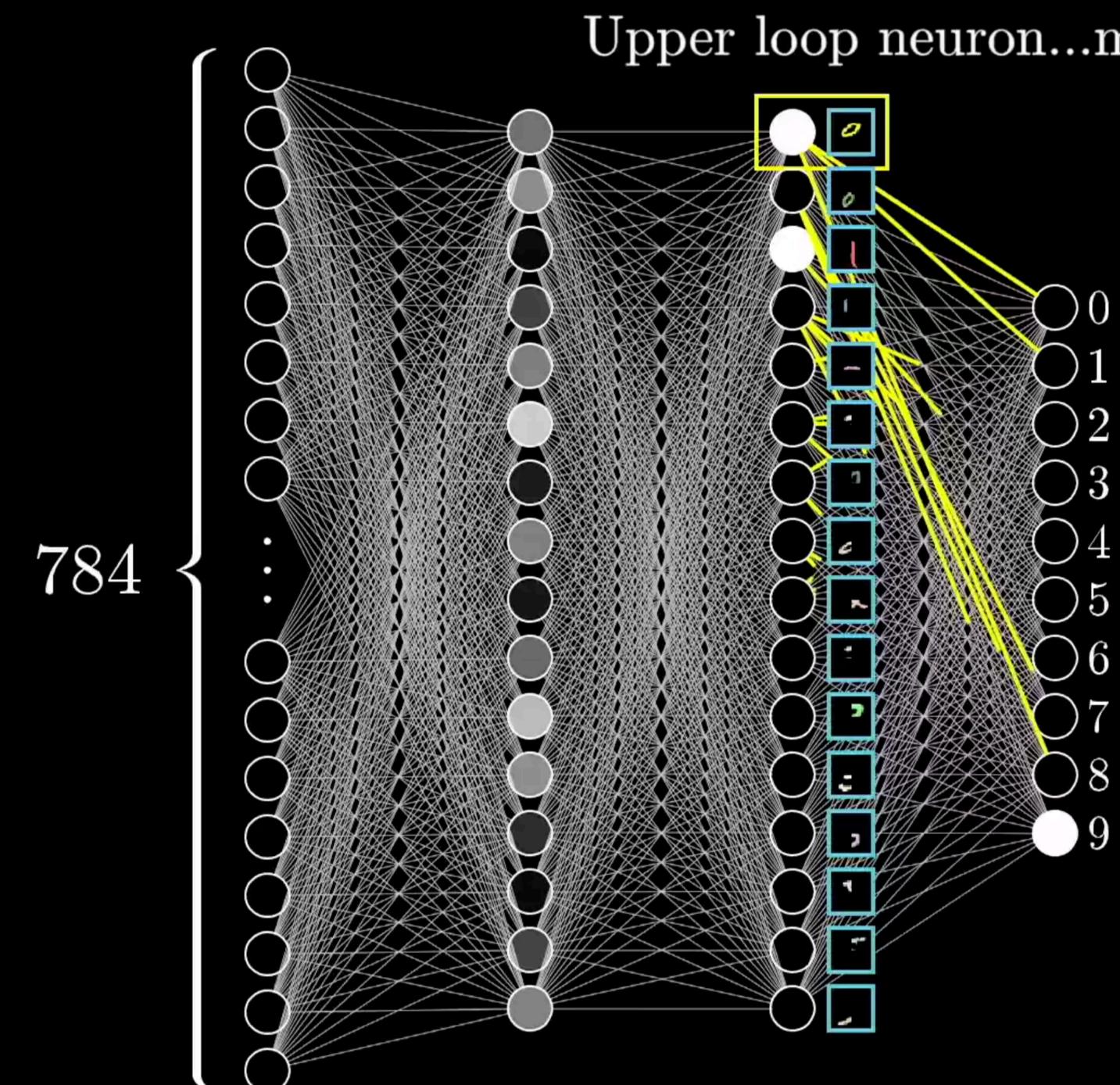
Machine Learning vs Deep Learning



Machine Learning vs Deep Learning

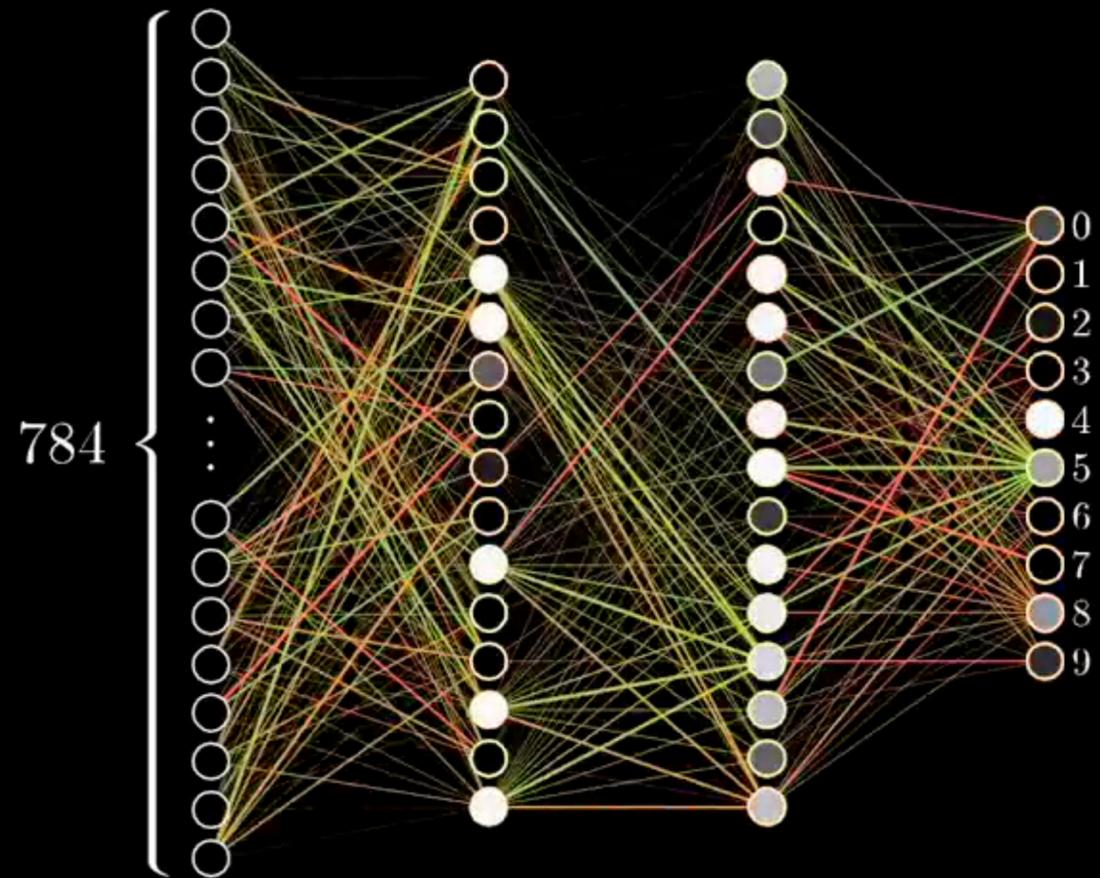
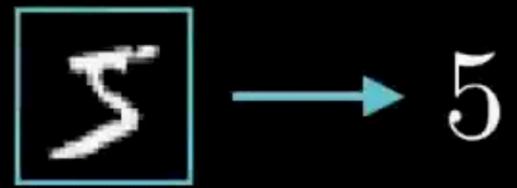


Machine Learning vs Deep Learning



Machine Learning vs Deep Learning

Training in progress...



[WebRTC e IA]

WebRTC

Gracias al WebRTC y la IA podemos hacer cosas como:

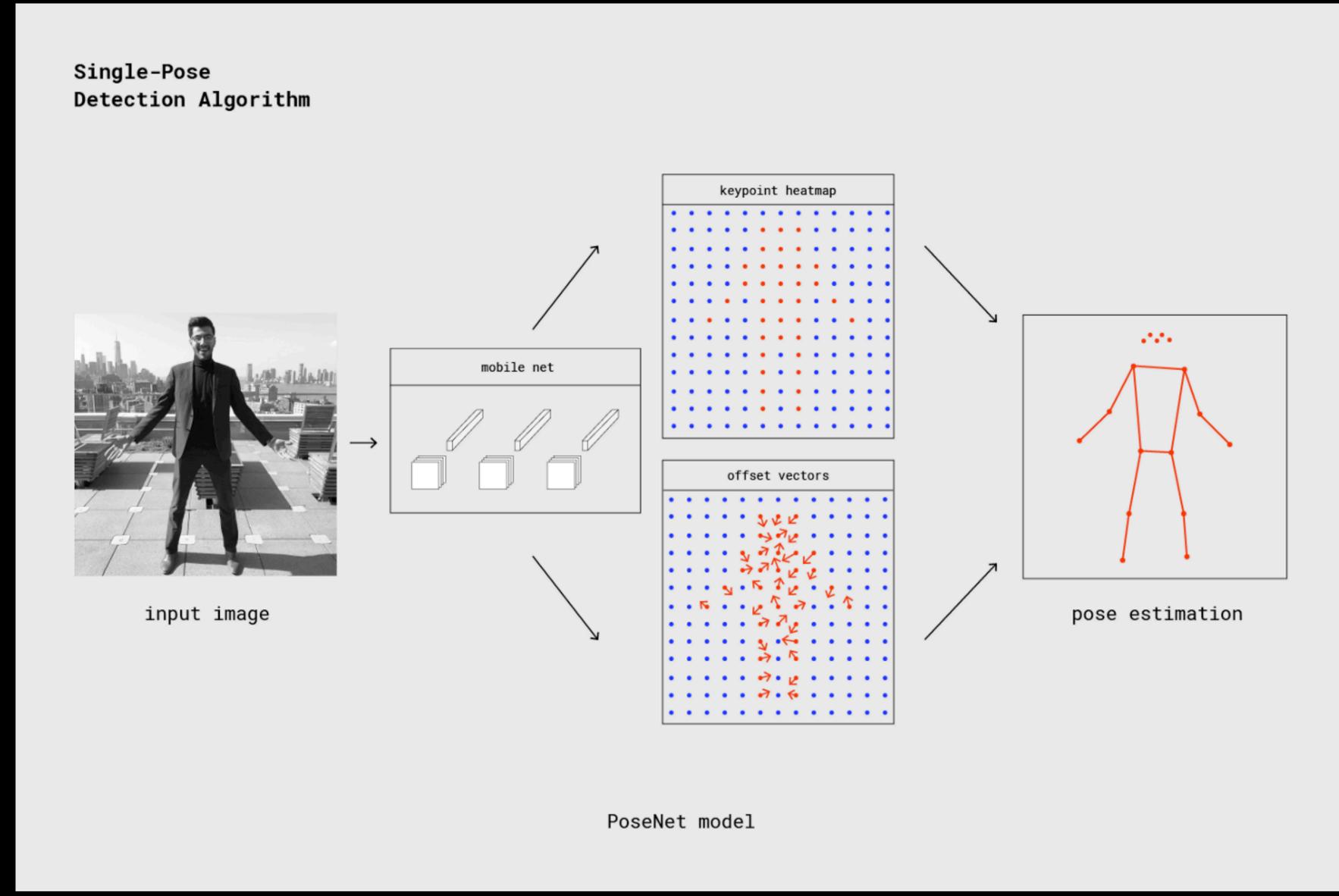
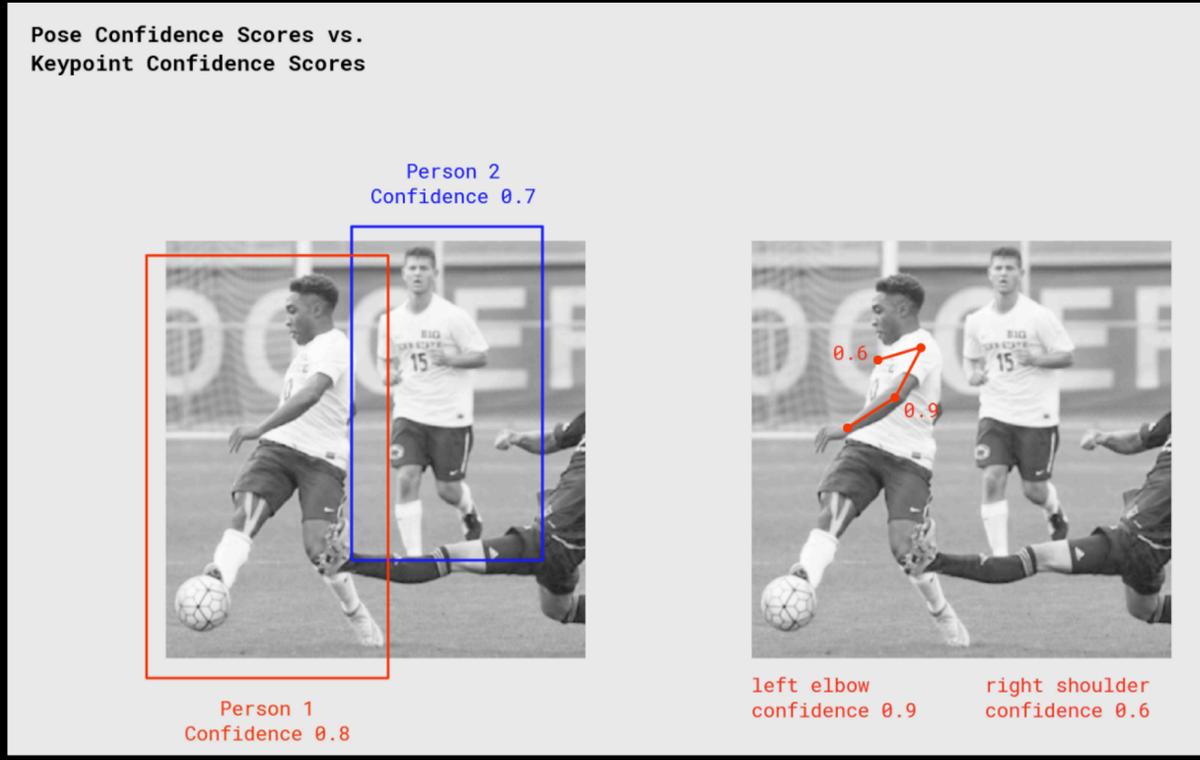
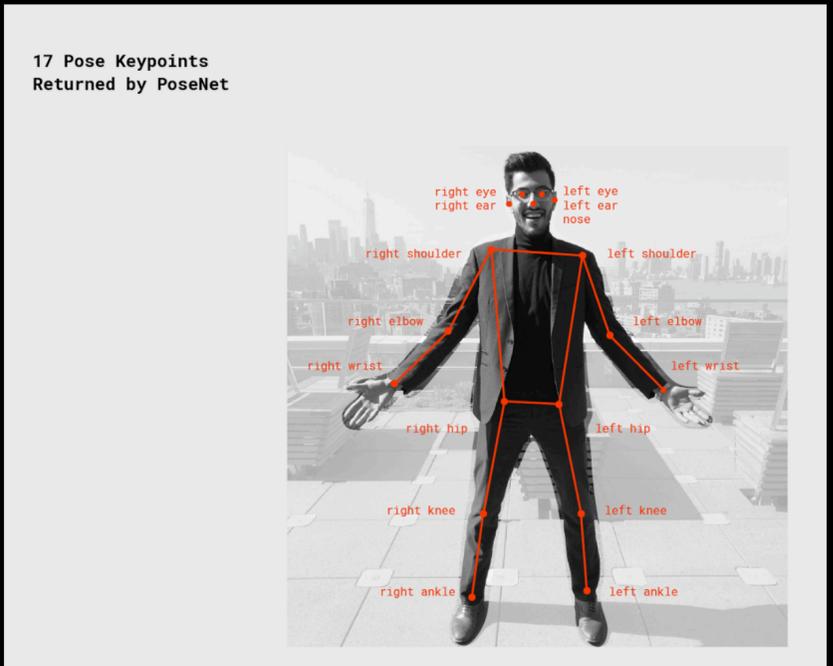
- Reconocimiento de imágenes (MobileNet)
- Reconocimiento del estado de ánimo
- Reconocimiento del número de personas
- Reconocimiento de qué están haciendo esas personas
- Filtrar el audio para obtener una mejora en la calidad de la voz
- ...

Demo



WebRTC

Demo



[Otros usos]

Otros usos

- **Detección de anomalías** en parámetros de sistemas, (procesador, memoria, hilos, procesos, etc.)
- Predicción de **incidencias del sistema** en base a sus parámetros.
- Creación de **bots conversacionales** de ayuda.
- **Identificación** de la persona que habla mediante “Voice ID Fingerprint”.
- **Aceleración en la gestión de incidencias** de equipos de Help Desk.
- Ajuste de los parámetros de **búsqueda de posibles leads** en base a las ventas logradas.
- Mejora sustancial de los **sistemas de reconocimiento de voz**.
- ...

Un futuro muy presente...



Google Duplex



HSBC Voice ID

¿Preguntas?



Agradecimientos

- a **Rosa y Pablo** por su paciencia y apoyo.
- a **Guillermo Ruiz** por su apoyo y recomendaciones.
- a **David Martín (Suki)** por sus ejemplos y tutoriales.
- a todos los que publican libremente su trabajo para enseñarnos cómo se hacen las cosas.
- a la **organización del VoIP2DAY** por invitarme a este evento.
- a **VOZ.com** por apoyarme con estos y otros proyectos.
- a los lectores y patrocinadores de **Sinologic.net**.
- y especialmente a...

... a Daniel Maestre



Gracias!