

# Machine Learning

Breve introducción

# Plan de trabajo beca post-doc

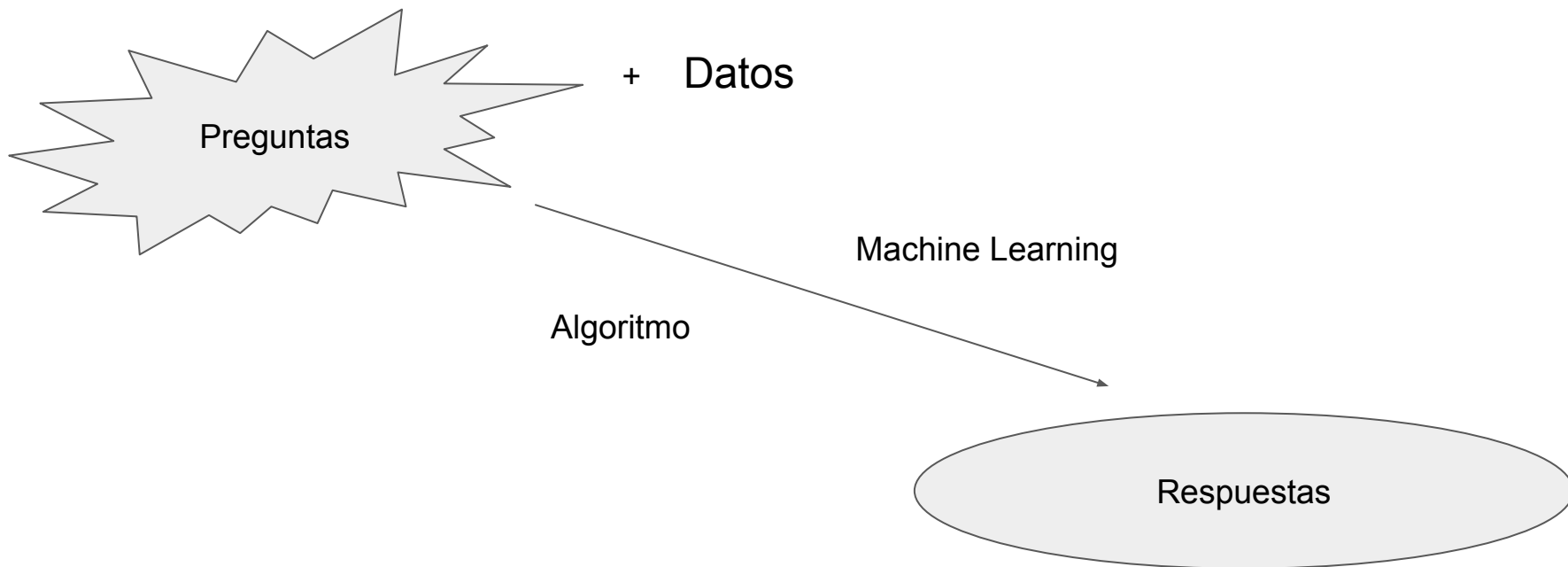
- - Diseñar nuevos inhibidores reversibles de la Cruzipaina (Cz) de mayor eficacia y menores efectos secundarios que los descubiertos hasta el momento.

## *Particulares*

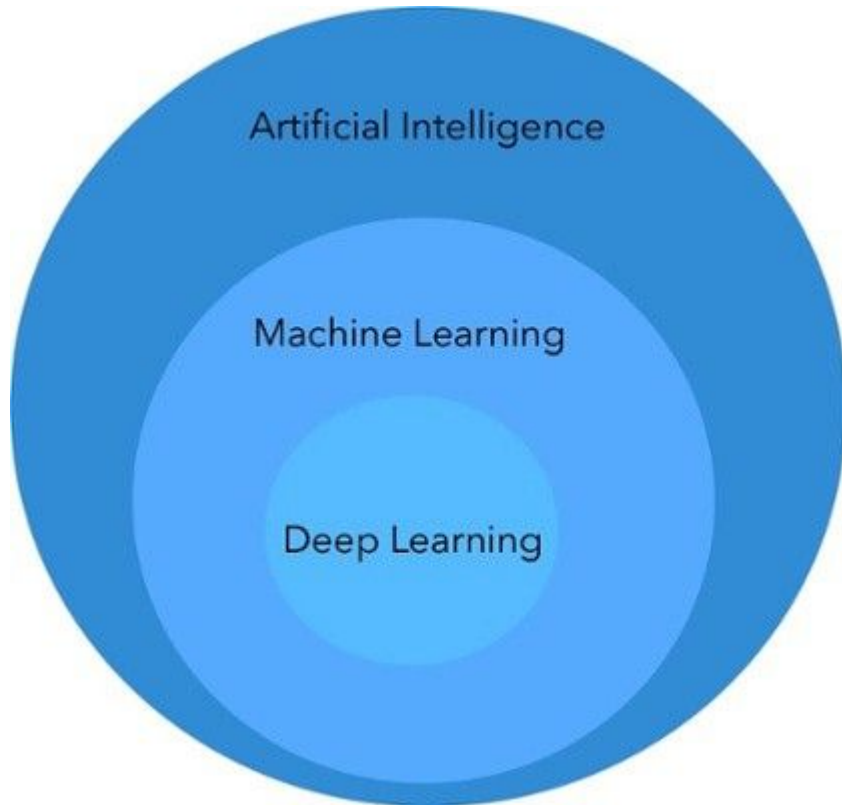
- - Entrenar modelos de aprendizaje automático para predecir actividades de nuevos candidatos a inhibidores de Cruzipaina.
- - Identificar las interacciones intermoleculares claves que permitan entender el mecanismo de inhibición de la enzima.

# ¿Qué es Machine Learning?

“Machine learning is about extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science.”



# Machine Learning / IA



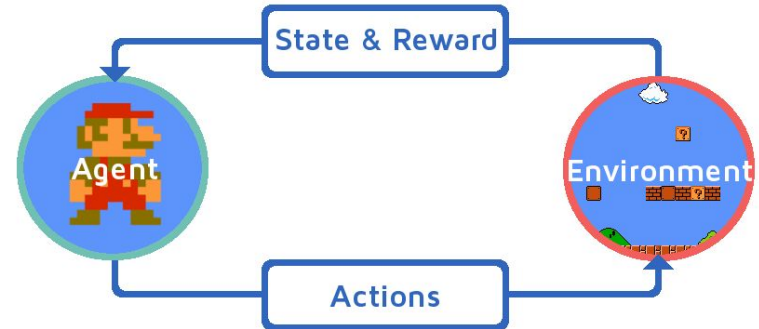
## Aprendizaje Supervisado

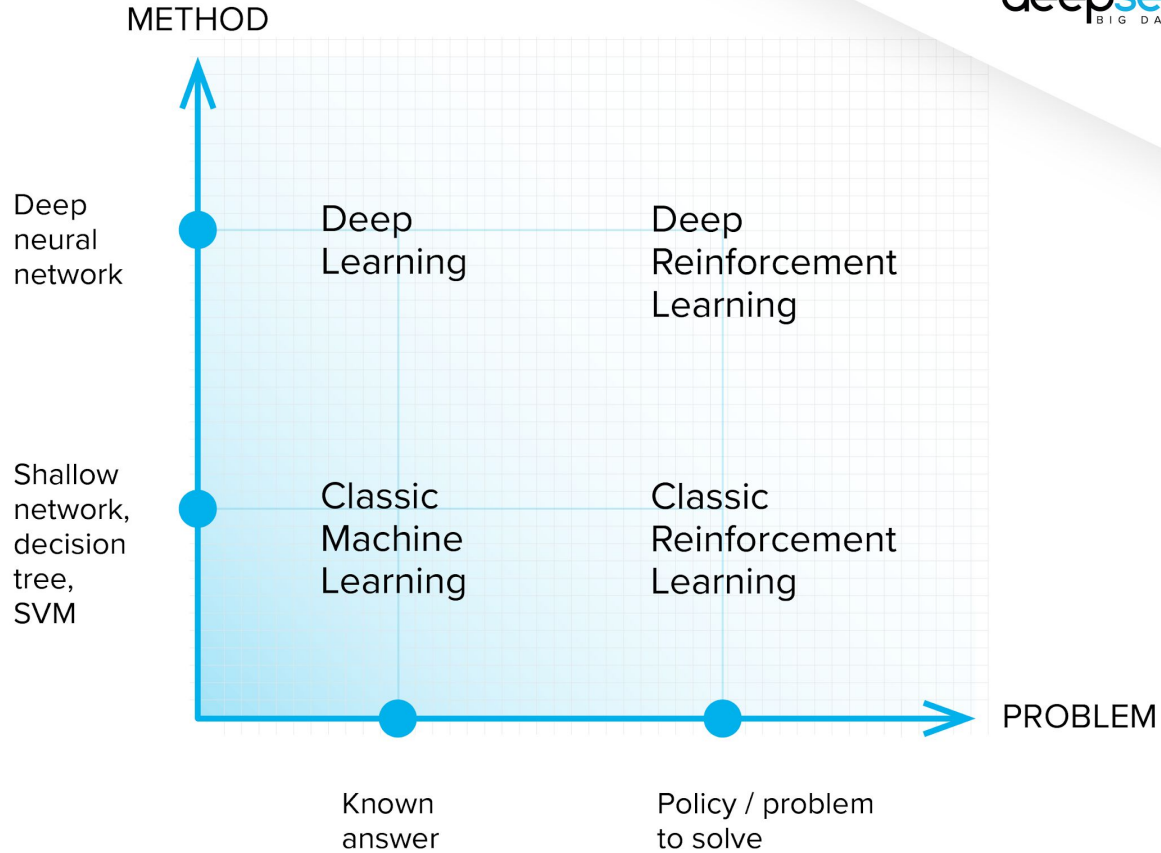
Regresión  
Clasificación

## Aprendizaje No Supervisado

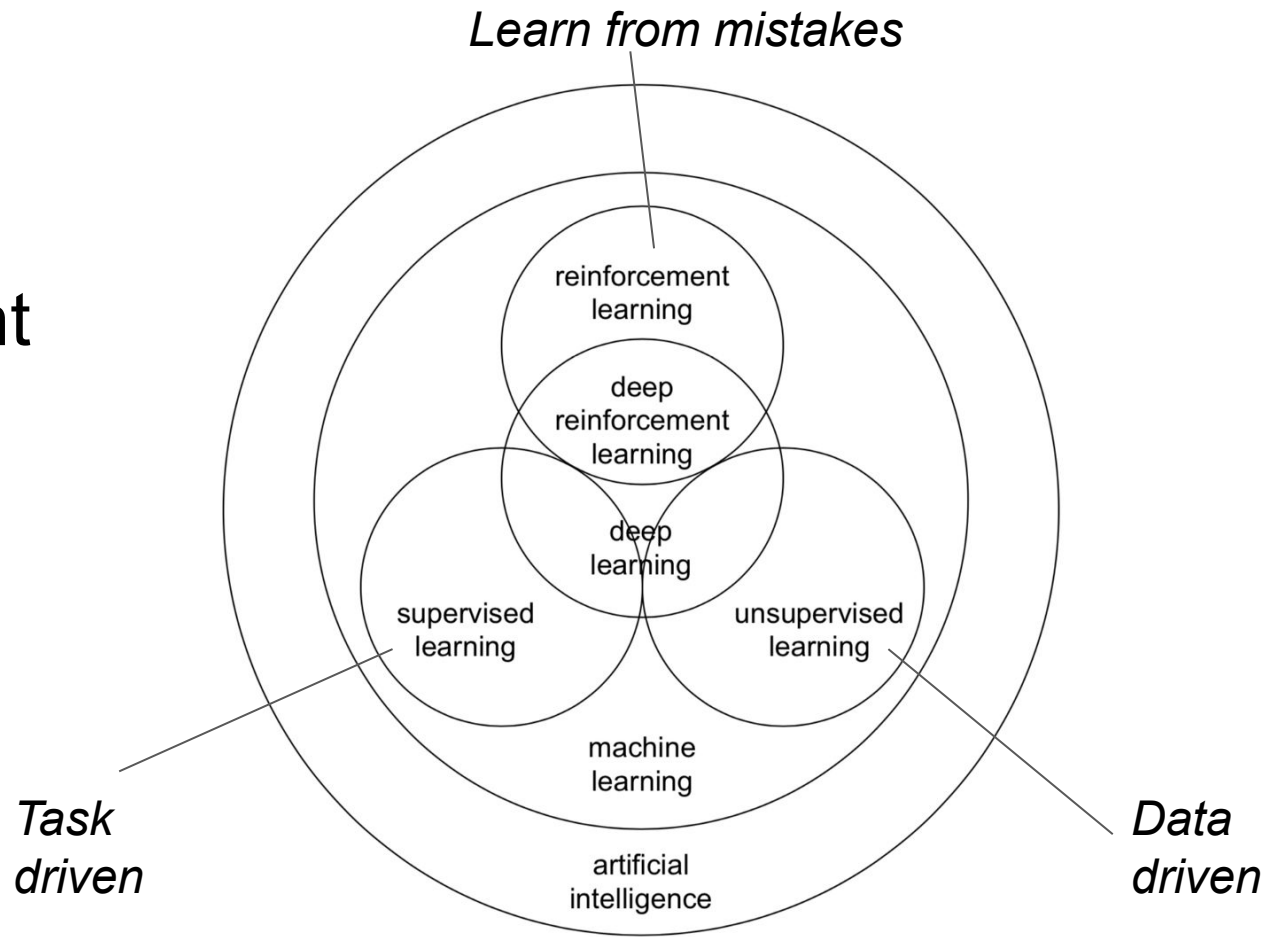
Clusterizado  
Reducción de dimensiones

## Aprendizaje Reforzado





# Supervised Unsupervised Reinforcement Learning



# ¿Qué tipo de problemas puede ser abordados mediante ML?

- Determinar si un tumor es maligno o benigno en una imagen médica.
  - Detectar actividad fraudulenta en transacciones de tarjeta de crédito.
  - Identificar el código zip de letras manuscritas en un sobre.
- 
- Detectar patrones de acceso anormales en un website
  - Segmentación de usuarios según preferencias de usuario.
  - Identificar tópicos (temas) en posteos de blogs.

# Conocer los datos y la tarea (task) a realizar

- What question(s) am I trying to answer? Do I think the data collected can answer that question?
- What is the best way to phrase my question(s) as a machine learning problem?
- Have I collected enough data to represent the problem I want to solve?
- What features of the data did I extract, and will these enable the right predictions?
- How will I measure success in my application?
- How will the machine learning solution interact with other parts of my research or business product?



# Modelización

Machine Learning

Clasificaciones

Regresiones

Deep Learning

Clasificaciones

Regresiones

**con una red neuronal  
(cambio en la arquitectura según el  
objetivo)**

Reinforcement learning

Modelar la respuesta (comportamiento) de un agente  
(puede ser una red neuronal) en un entorno de  
simulación.

c  
o  
m  
p  
l  
e  
j  
i  
d  
a  
d



# Modelización

## ■ "Pure" Reinforcement Learning (cherry)

- ▶ The machine predicts a scalar reward given once in a while.

- ▶ **A few bits for some samples**

## ■ Supervised Learning (icing)

- ▶ The machine predicts a category or a few numbers for each input

- ▶ Predicting human-supplied data

- ▶ **10→10,000 bits per sample**

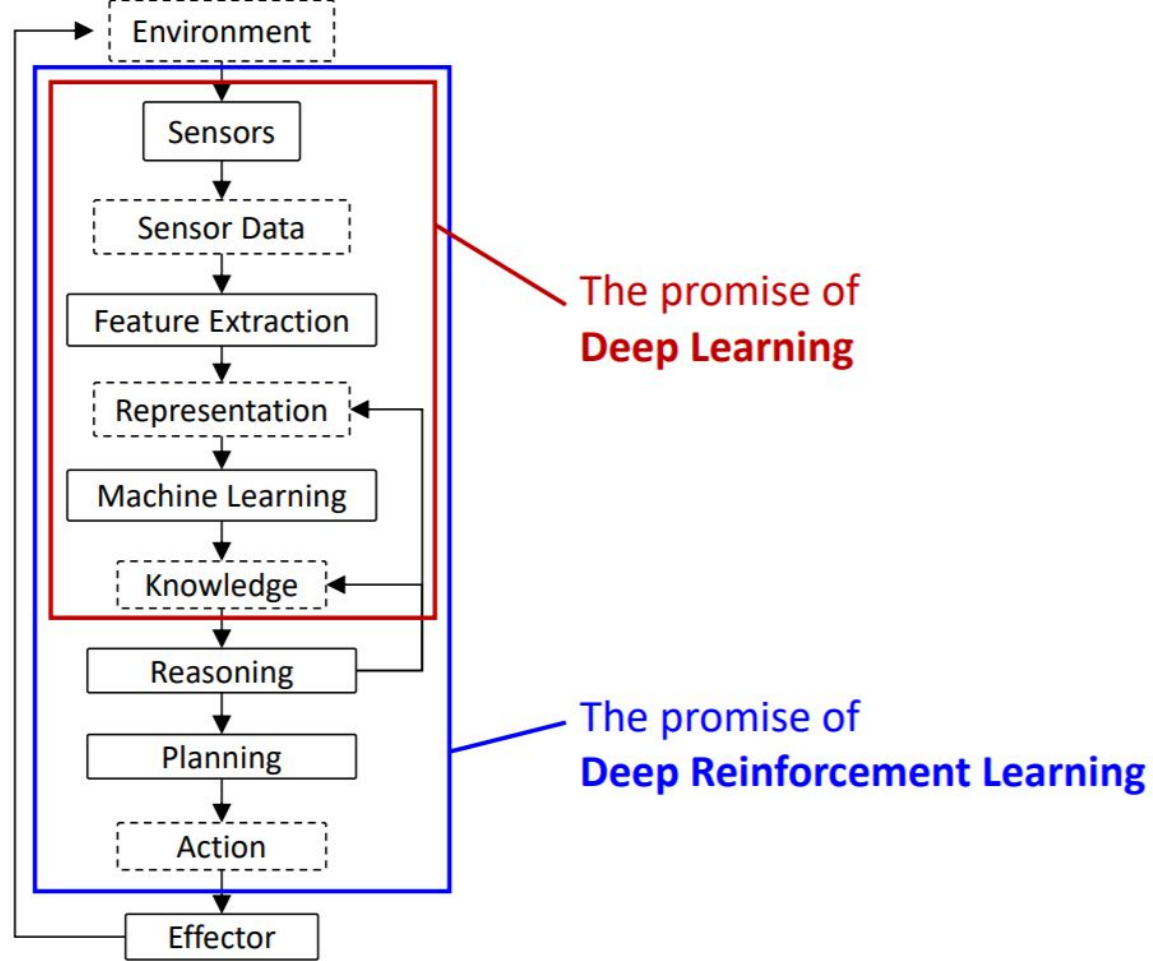
## ■ Unsupervised/Predictive Learning (cake)

- ▶ The machine predicts any part of its input for any observed part.

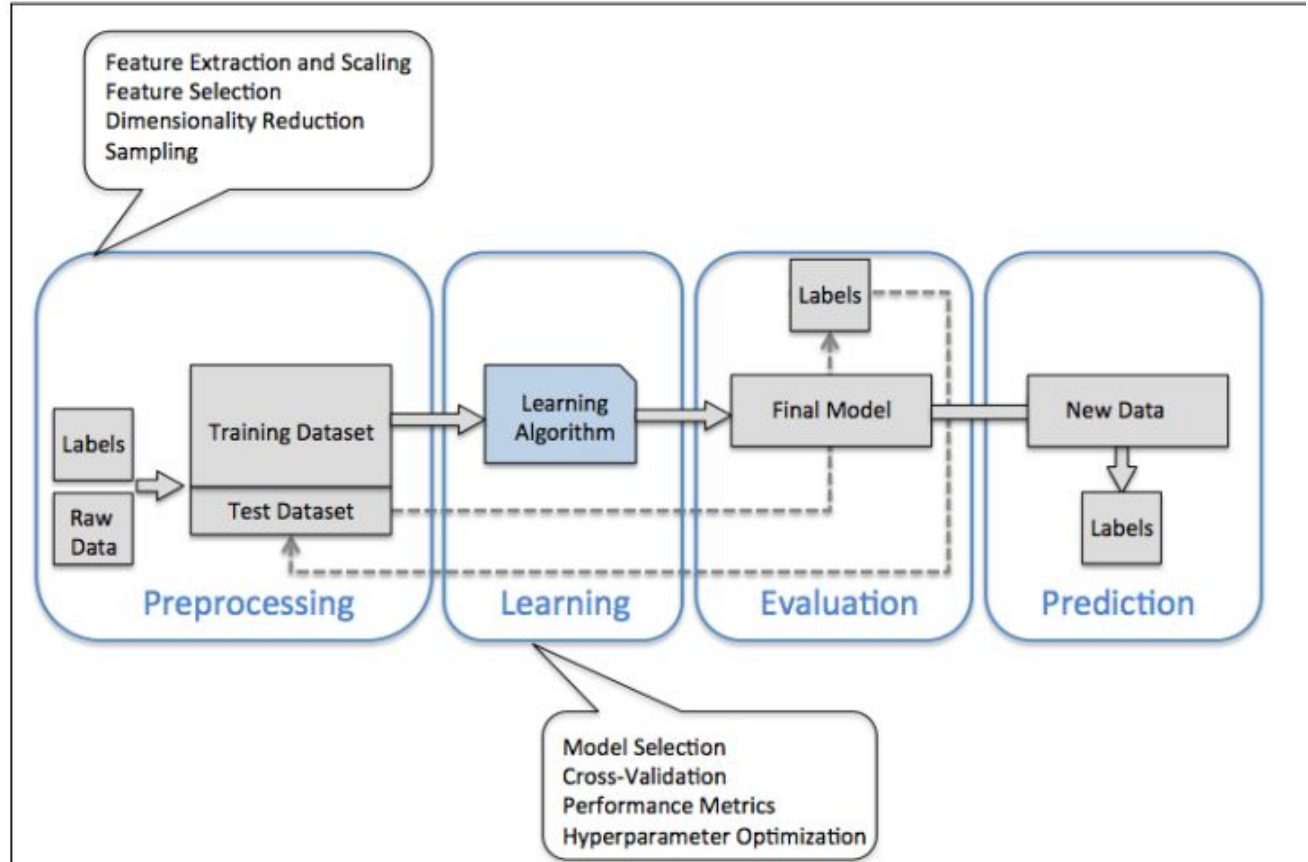
- ▶ Predicts future frames in videos

- ▶ **Millions of bits per sample**





# ML pipeline



# Datos ordenados (tidy data)

Data cleaning/ data wrangling/ data munging.

country	year	cases	population
Afghanistan	1999	7725	19987071
Afghanistan	2000	8666	2059360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	214258	1272915272
China	2000	216766	128042583

variables

country	year	cases	population
Afghanistan	1999	7725	19987071
Afghanistan	2000	8666	2059360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	214258	1272915272
China	2000	216766	128042583

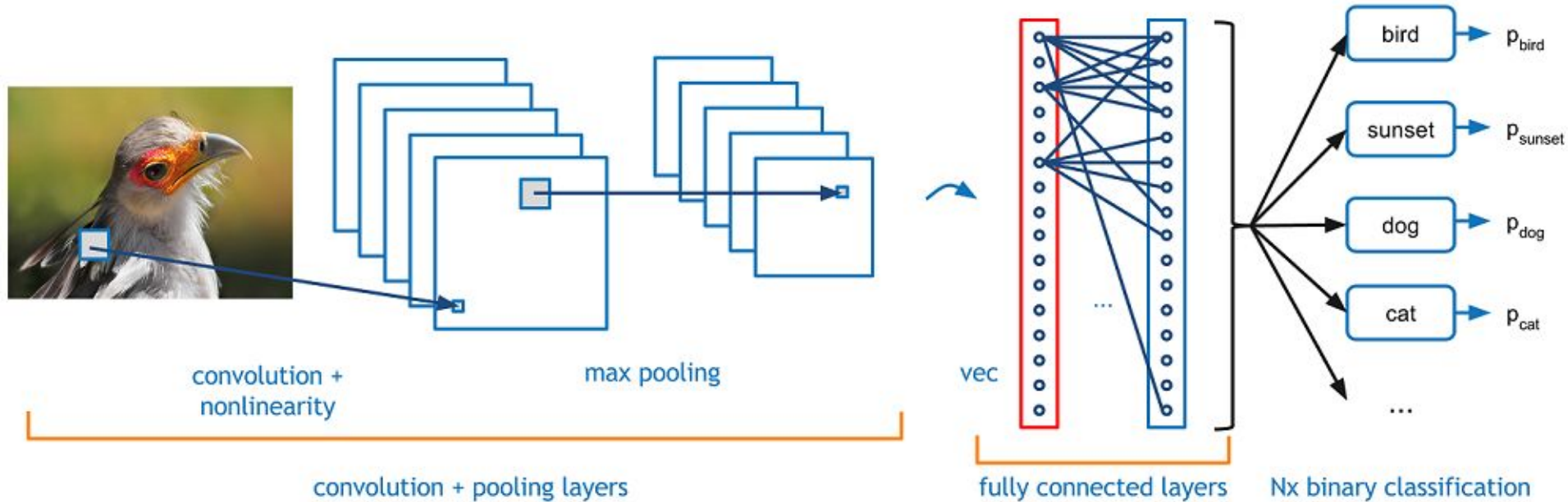
observations

country	year	cases	population
Afghanistan	1999	7725	19987071
Afghanistan	2000	8666	2059360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	214258	1272915272
China	2000	216766	128042583

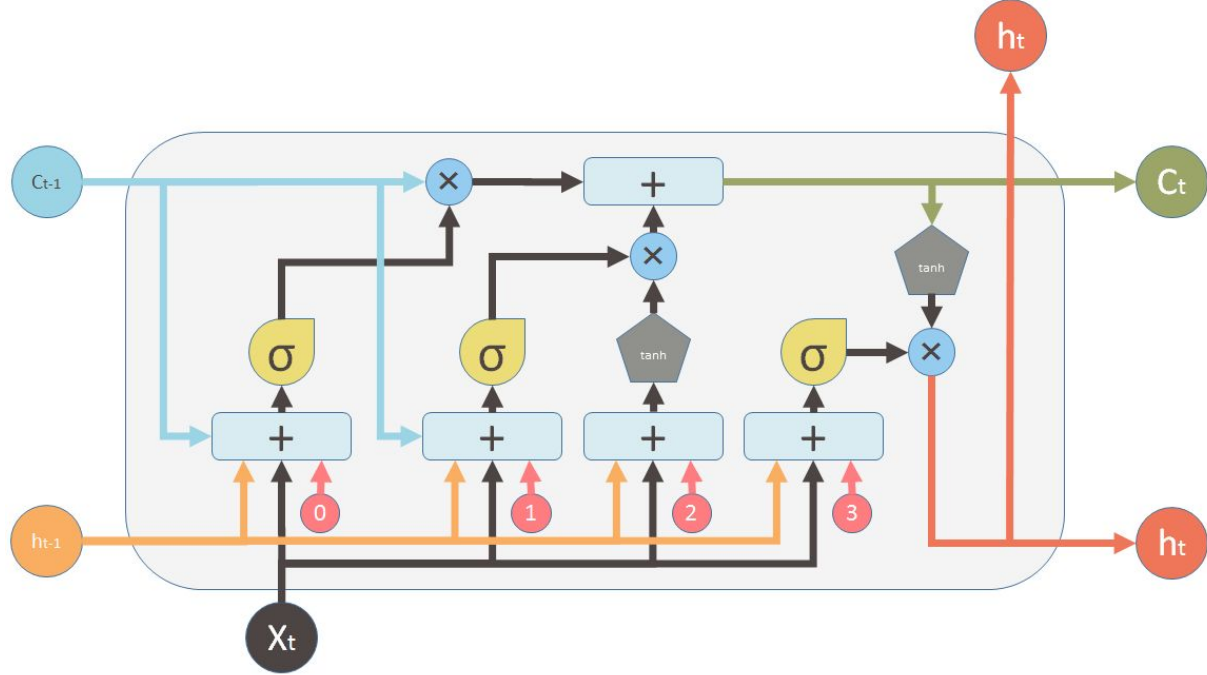
values

Algunas  
arquitecturas  
importantes

# Convolutional Neural Networks (CNNs)



# Recurrent Neural Networks (LSTM)



Inputs:



Input vector



Memory from previous block



Output of previous block

outputs:



Memory from current block



Output of current block

Nonlinearities:



Sigmoid



Hyperbolic tangent

Bias:



Vector operations:



Element-wise multiplication



Element-wise Summation / Concatenation



Qué se realizó

- Paper

COMBINING CHARGE DENSITY ANALYSIS  
WITH MACHINE LEARNING TOOLS TO  
INVESTIGATE CRUZAIN INHIBITION  
MECHANISM



- Paper

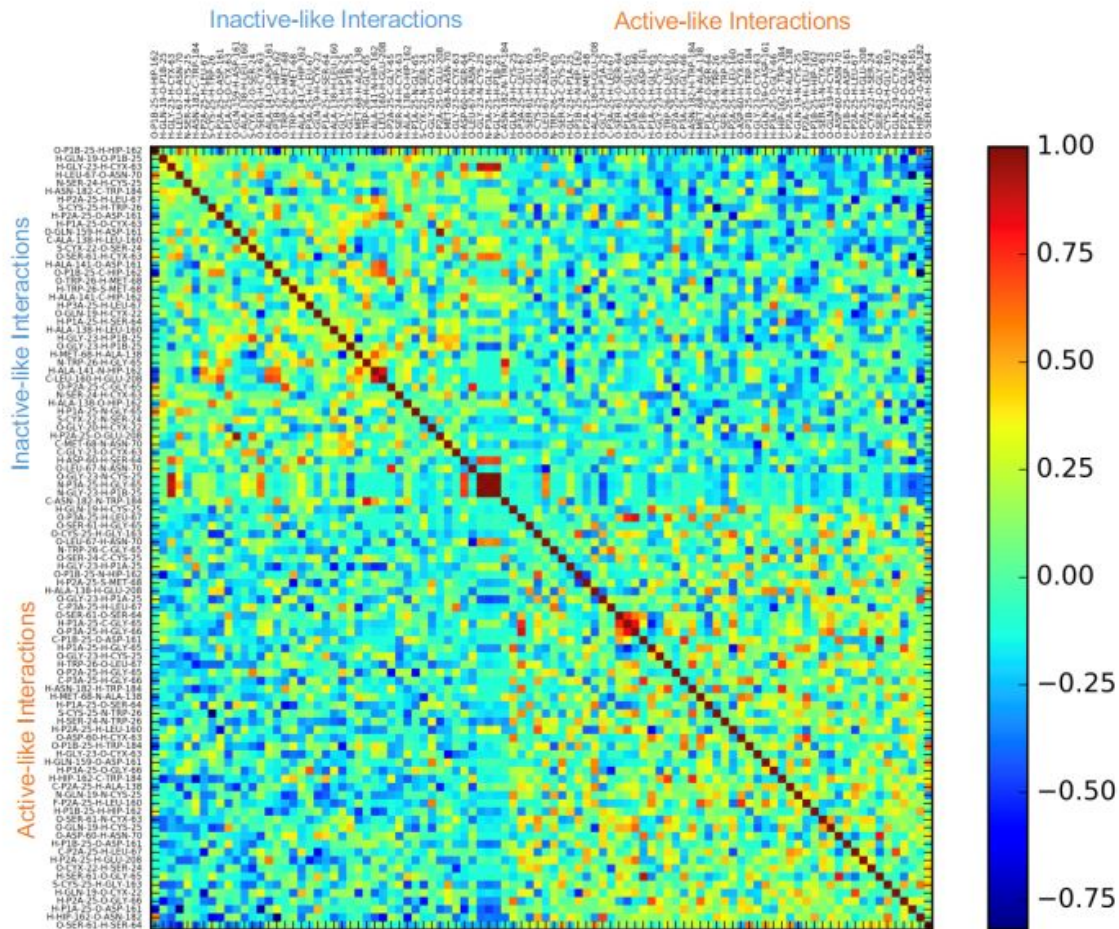


Figure 3. Correlation Matrix based on charge density data from interactions in Cz-Inh complexes.

# ● Halógenos

- Optimización de random forest para seleccionar variables. Las muestras corresponden a compuestos con y sin halógeno.



Siguiente

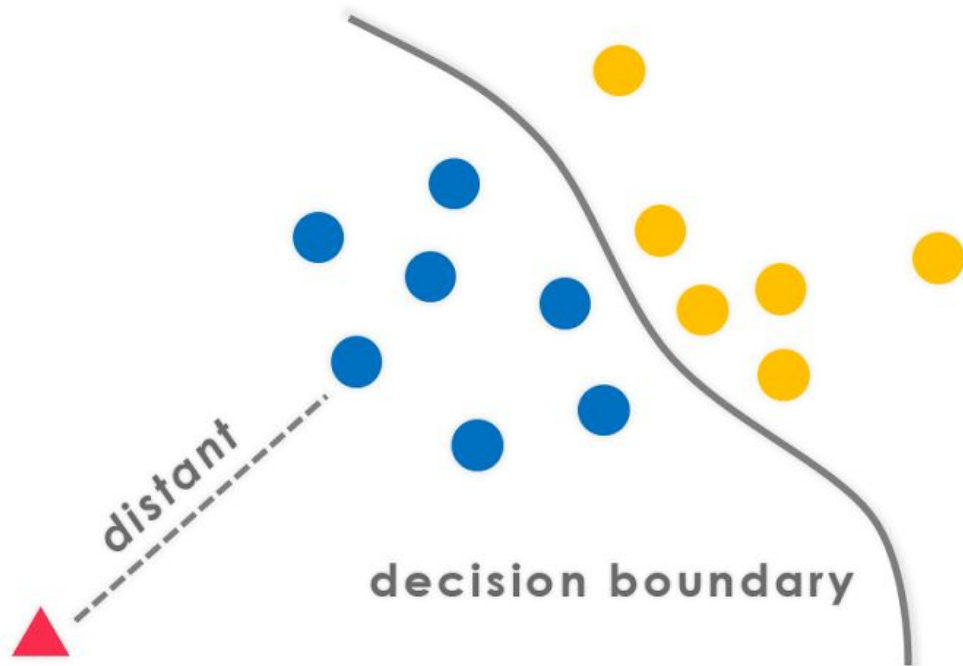
- Optimización de un algoritmo para selección de variables pero para interacciones intramoleculares de proteína.

# Perspectivas a futuro

- + Notebooks con modelos generativos
- + Uso de herramientas de ML para clasificar entre compuestos activos aislados y los que tienen efecto sobre el parásito. Calcular la probabilidad de que el compuesto activo llegue al blanco y haya efecto.

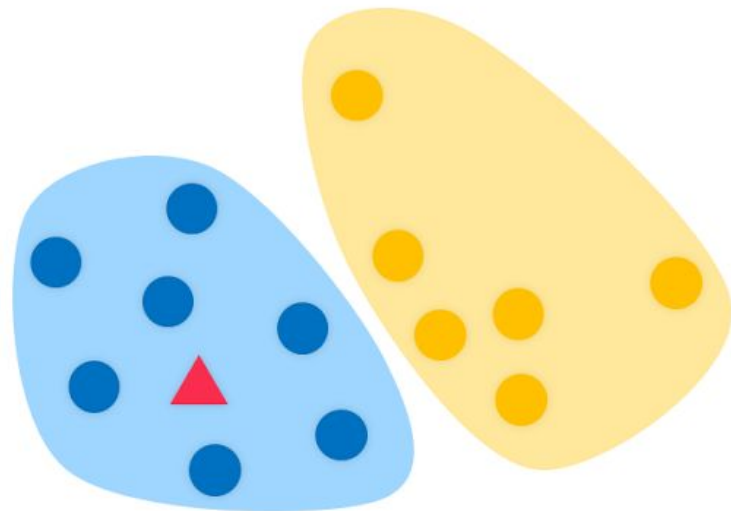
# Modelos Discriminativos / Generativos

Discriminative



**Asume la existencia de  $P(Y|X)$  y lo calcula a partir de datos de entrenamiento**

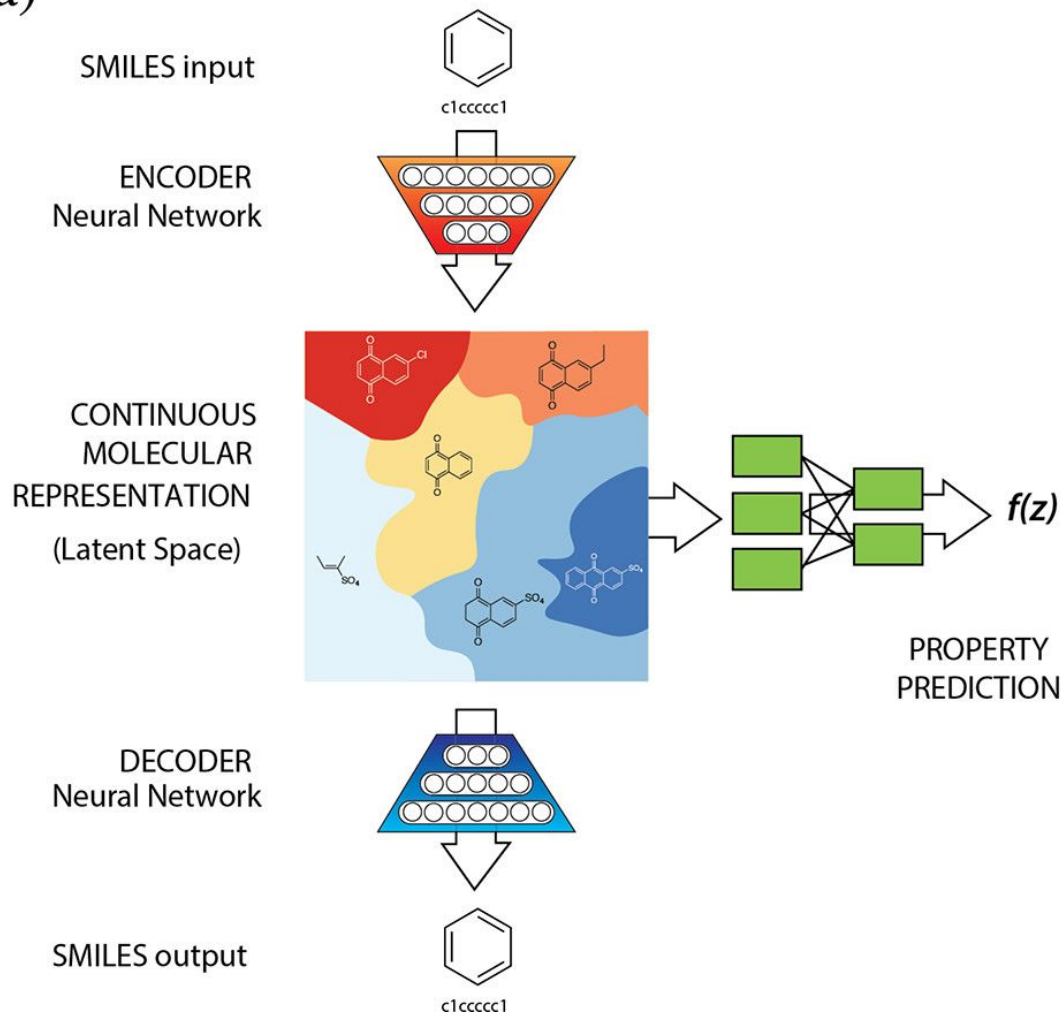
Generative



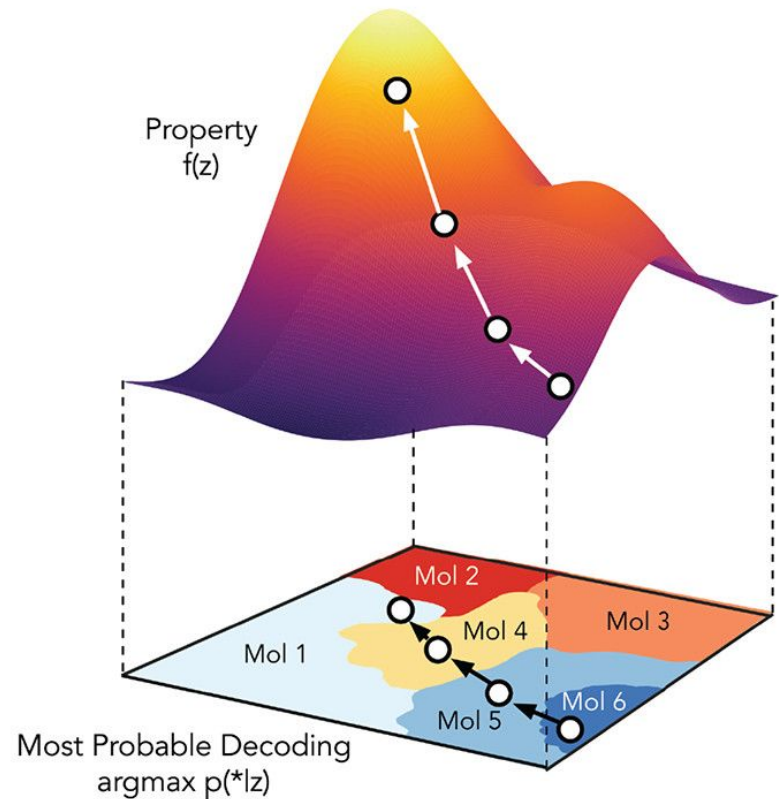
**Asume  $P(X|Y)$ ,  $P(Y)$   
Calcula  $P(Y|X)$  a través de Teorema de Bayes.**



(a)



(b)



**Gómez-Bombarelli, 2017.**

# Generación de moléculas

- Recurrent Neural Networks (RNN)
- Autoencoders -> VAE
- Generative Adversarial Networks (GANs)
- Reinforcement Learning
- CNNs (como parte de Autoencoders o GANs)

# Desafíos del tema

- Programación (diversidad: R, Python, Tensorflow, PyTorch, OpenAI Gym)
- Matemática (Estadística).
- Campo en constante evolución.
- Cantidad de material disponible crece exponencialmente.
- Cambio en el paradigma de la computación / estadística (*En discusión*)