

Introduction to Machine Learning

And it's applications in bioinformatics

email: marie-galadriel.briere@univ-amu.fr

Outline

01

Introduction

AI: History and Overview

02

The basics of ML

ML workflow, ML tasks,
ML performance

03

Main approaches

Some example of ML approaches,
spanning the five tribes of ML

04

Applications

Focus on bioinformatics

05

Tools

Languages and
frameworks

06

Conclusion

Limits, challenges and
future of ML

Objectives

Provide an overview of the field

Provide keys for further exploration

Develop critical thinking skills about AI

Evaluation

Written in-class exam (~30 min)

- you can consult the **notes** taken in class

Introduction

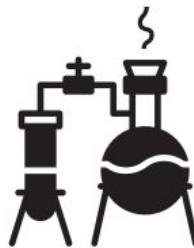
AI: History and Overview



FIDLE

<https://fidle.cnrs.fr>

1st paradigm



Experimental science

2nd paradigm

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle$$

$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$

Theoretical science

3rd paradigm

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle$$
$$\nabla \times H = J + \frac{\partial D}{\partial t}$$
$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$



Computational science

4th paradigm¹



Data-driven science



1600



1950



2000

Artificial Intelligence ?

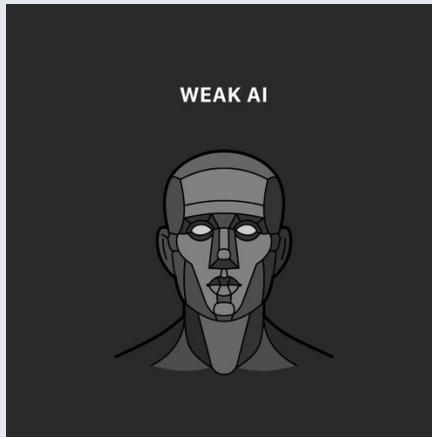
A machine that:

- Simulates human intelligent behavior?
- Thinks like a human?
- Perceives its environment and acts on it?
- Can solve complex tasks?
- Can adapt to a changing environment with incomplete information?

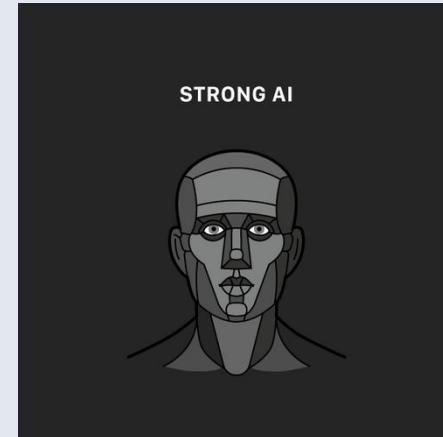
There is no universally accepted single definition...

For the European Parliament, artificial intelligence represents any tool used by a machine to 'reproduce human-related behaviors, such as reasoning, planning, and creativity' => **And beyond!**

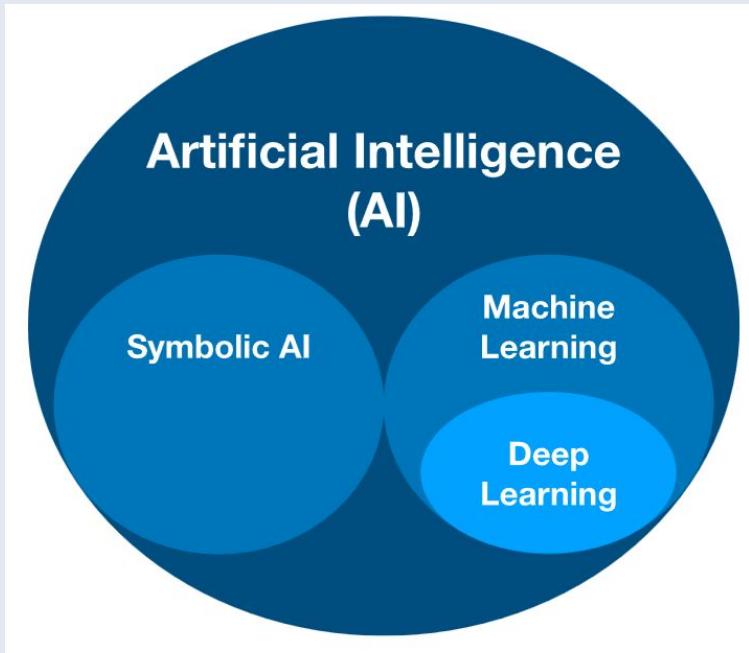
Weak AI and Strong AI



A machine that equals or
surpasses human intelligence,
but **for a specific task**

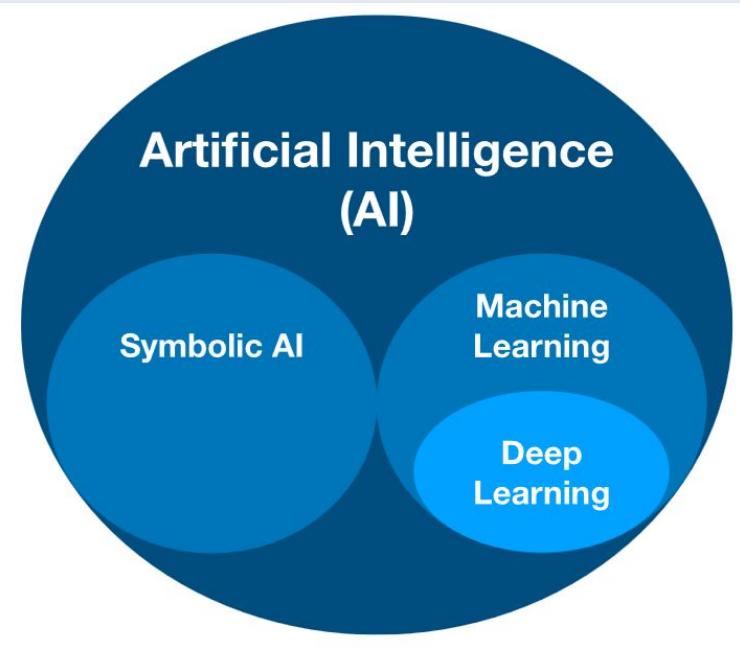


A machine with human-level
general intelligence, which can
perform **any new task**



Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

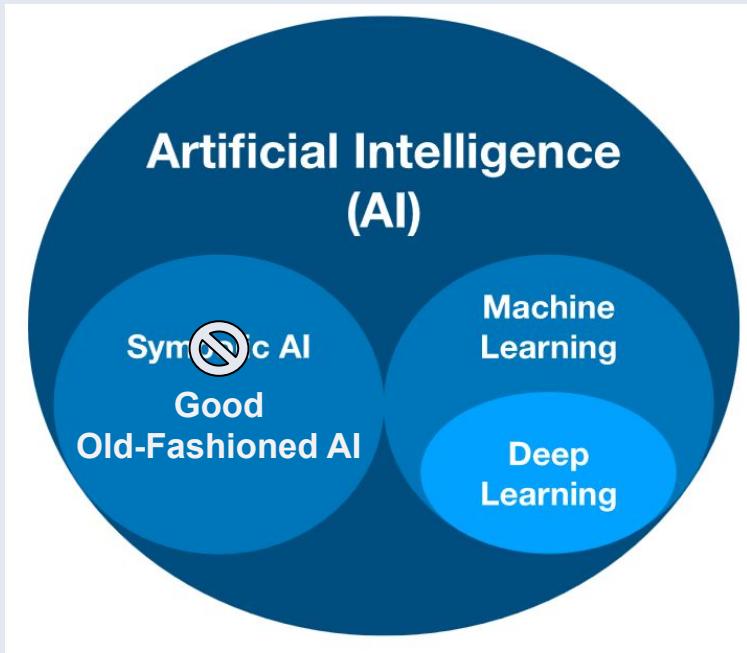
Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.



Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

Symbolic AI (GOFAI) is an approach to artificial intelligence that **uses logical and symbolic reasoning techniques** to manipulate abstract symbols and solve problems.

Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.



Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.

Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

Symbolic AI (GOFAI) is an approach to artificial intelligence that **uses logical and symbolic reasoning techniques** to manipulate abstract symbols and solve problems.

An example of GOFAI

Backward chaining (Chainage arrière)

Given a set of rules:

1. If X croaks and X eats flies – Then X is a frog
2. If X chirps and X sings – Then X is a canary
3. If X is a frog – Then X is green
4. If X is a canary – Then X is yellow

And an input:

- Fritz croaks
- Fritz eats flies

What color is Fritz ?

- 1) If X croaks and eats flies – Then X is a frog
- 2) If X chirps and sings – Then X is a canary
- 3) If X is a frog – Then X is green
- 4) If X is a canary – Then X is yellow

You are looking for what color your pet is there are two options.

- 1) If X croaks and eats flies – Then X is a frog
- 2) If X chirps and sings – Then X is a canary
- 3) If X is a frog – Then X is green
- 4) If X is a canary – Then X is yellow

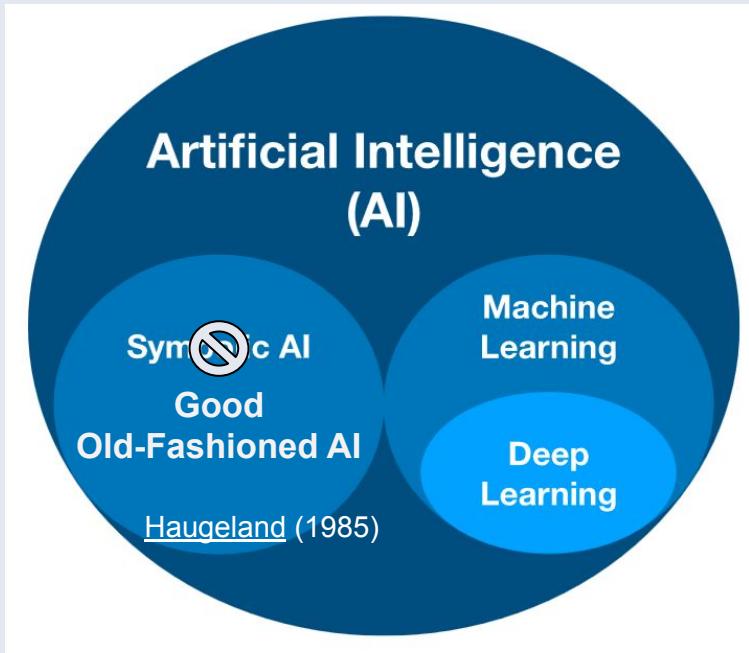
Try the first option.

- 1) If X croaks and eats flies – Then X is a frog
- 2) If X chirps and sings – Then X is a canary
- 3) If X is a frog – Then X is green
- 4) If X is a canary – Then X is yellow

Iterate through the list and see if you can find if X is a frog.

- 1) If X croaks and eats flies – Then X is a frog
- 2) If X chirps and sings – Then X is a canary
- 3) If X is a frog – Then X is green
- 4) If X is a canary – Then X is yellow

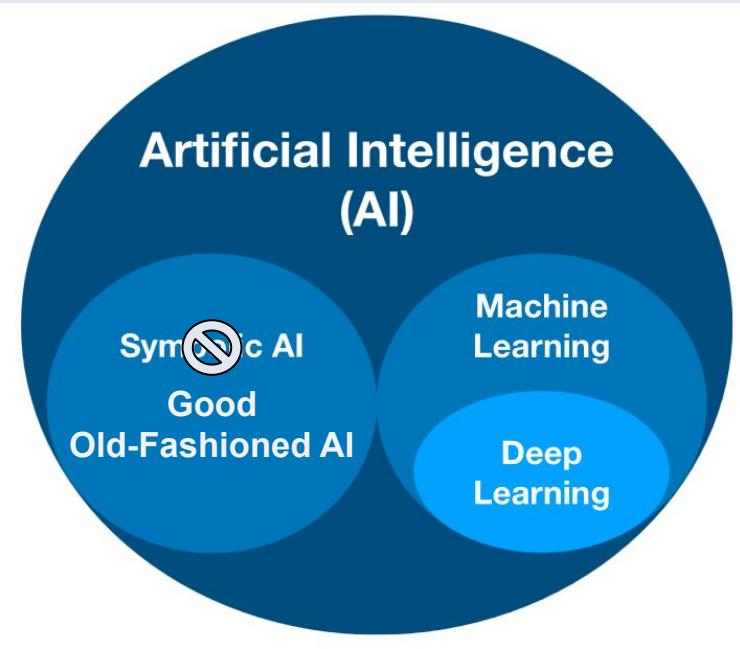
Repeat with step 1. X croaks and eats flies is given as true. Since X croaks and eats flies, X is a frog. Since X is a frog, X is green.



Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

Symbolic AI (GOFAI) is an approach to artificial intelligence that **uses logical and symbolic reasoning techniques** to manipulate abstract symbols and solve problems.

Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.

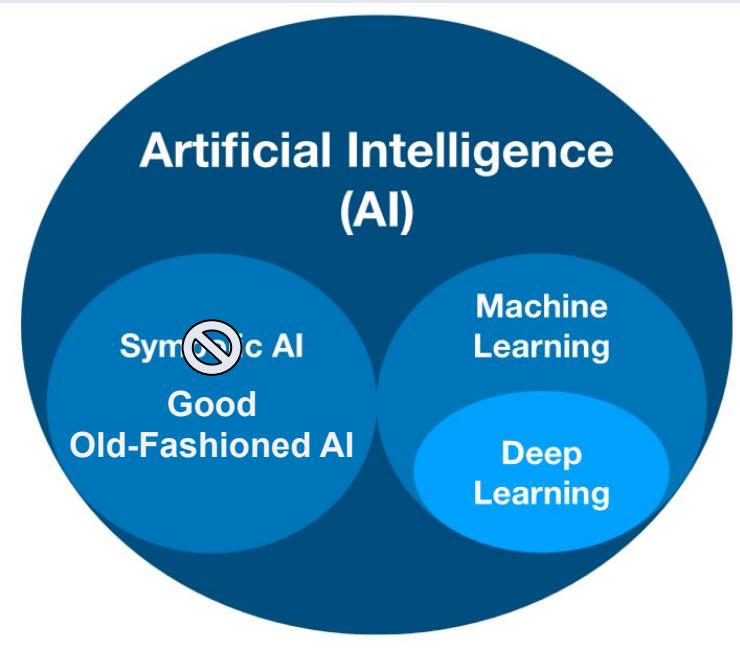


Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.

Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

Symbolic AI (GOFAI) is an approach to artificial intelligence that **uses logical and symbolic reasoning techniques** to manipulate abstract symbols and solve problems.

Machine learning focuses on developing computer algorithms and models capable of **learning from data** and **performing specific tasks without being explicitly programmed for each task**.



Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.

Artificial intelligence is the general field of computer science that aims to develop intelligent machines.

Symbolic AI (GOFAI) is an approach to artificial intelligence that **uses logical and symbolic reasoning techniques** to manipulate abstract symbols and solve problems.

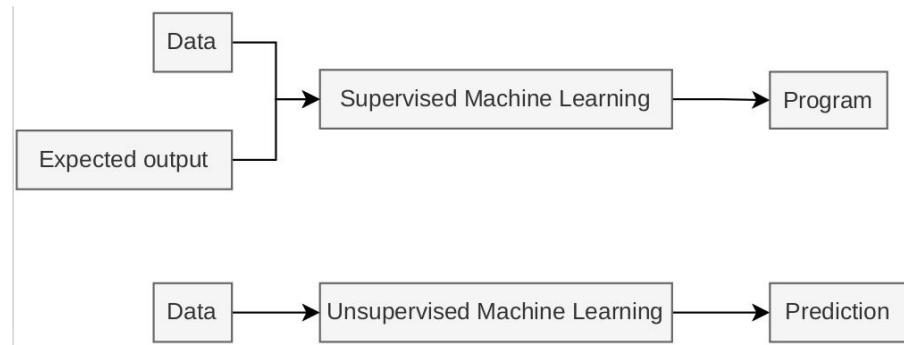
Machine learning focuses on developing computer algorithms and models capable of **learning from data** and **performing specific tasks without being explicitly programmed for each task**.

Deep learning is a subfield of machine learning that involves the use of **artificial neural networks => connectionist AI**.

GOFAI



Machine Learning

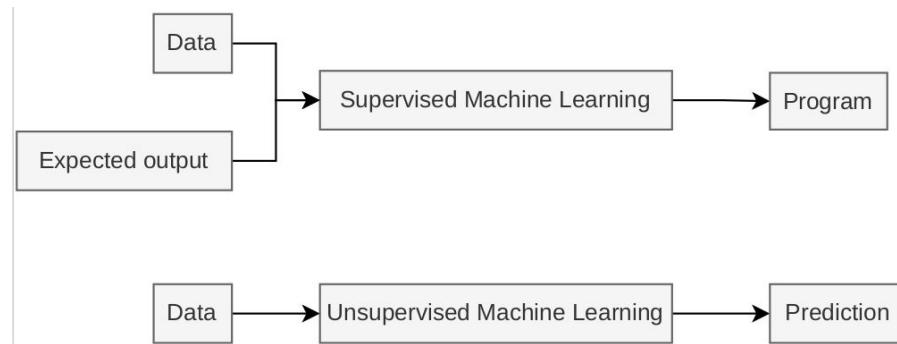


GOFAI



Deductive approach

Machine Learning

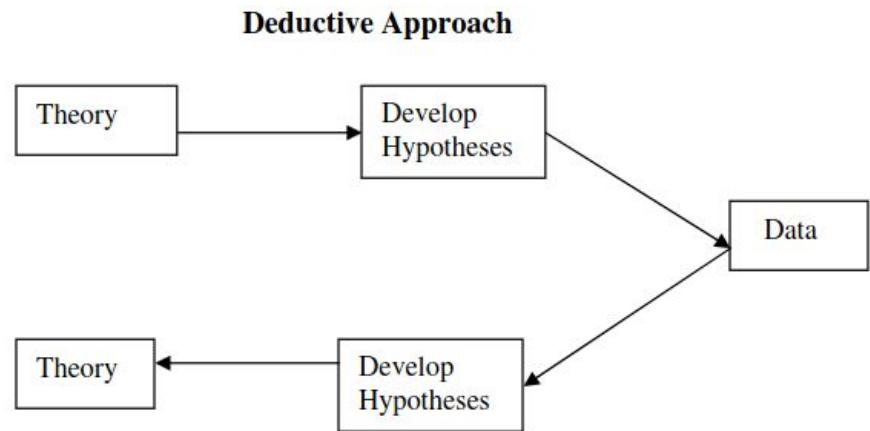


Inductive approach

GOFAI

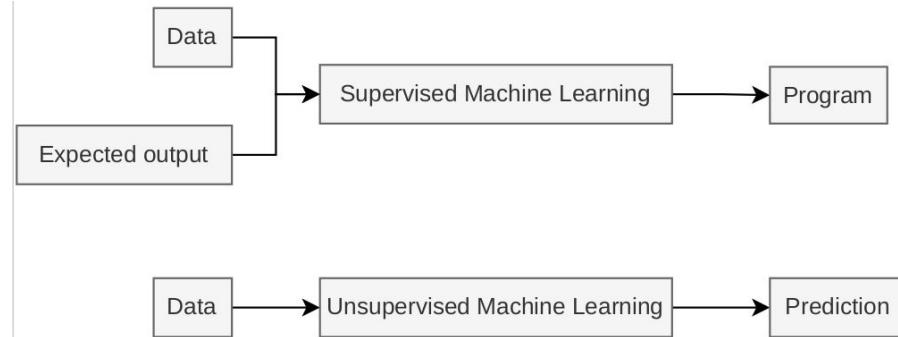


Deductive approach



Inductive Approach

Machine Learning



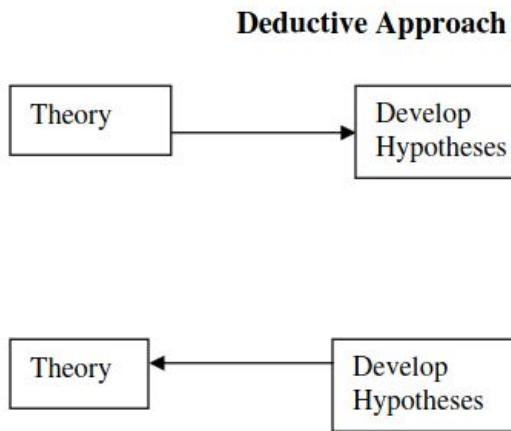
Inductive approach

Shah, Dr. Naimatullah. "Determinants of Employee Readiness for Organisational Change," April 27, 2023.

GOFAI

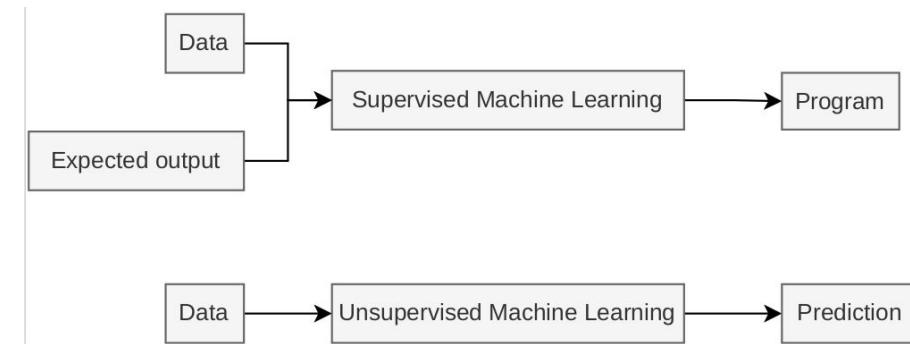


Deductive approach

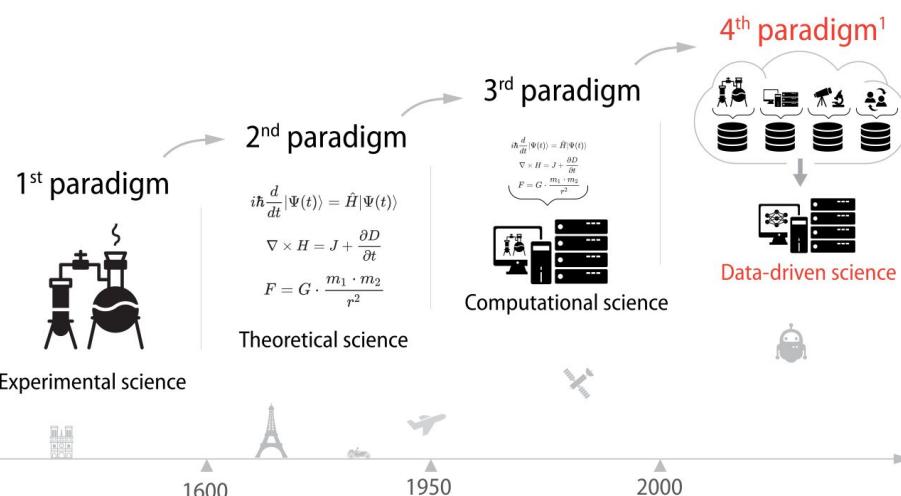


Deductive Approach

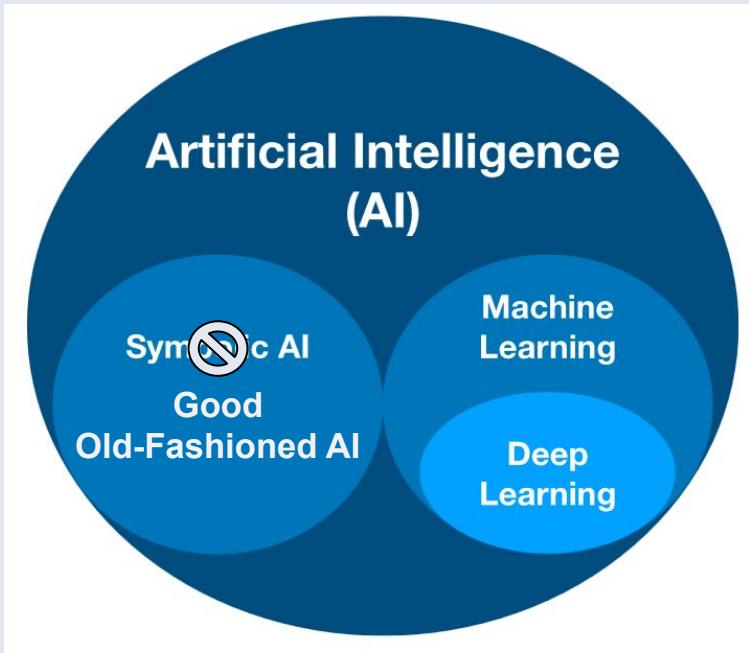
Machine Learning



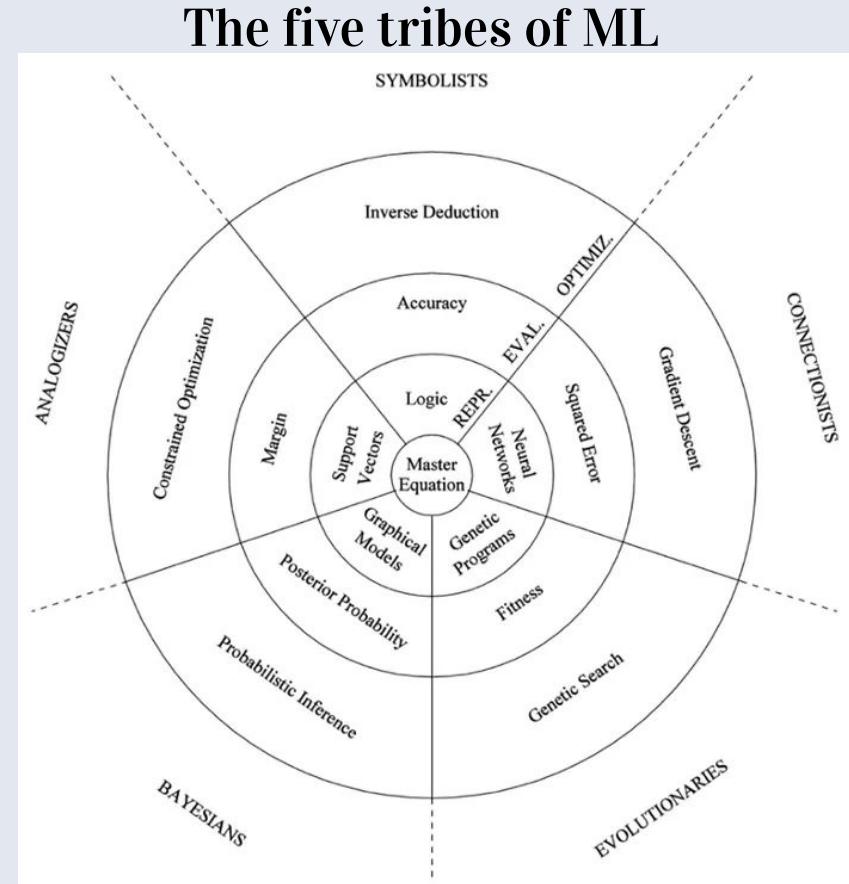
Inductive approach



Inductive Approach

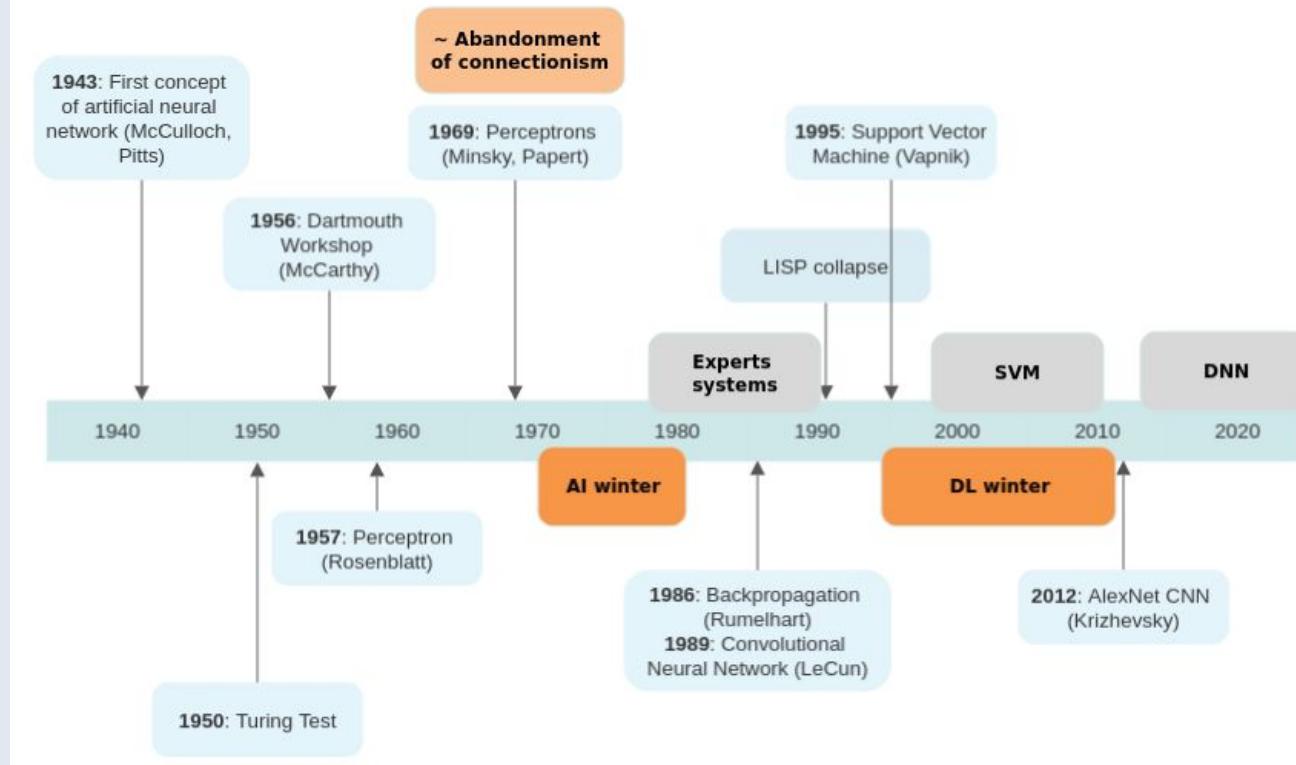


Graziani, Mara, et al. "A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences." *Artificial intelligence review* (2022): 1-32.



Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

Some AI landmarks



Take home message

- Data-driven science
 - deductive vs inductive approaches
 - Traditional programming vs machine learning
 - GOFAI vs Machine Learning
 - GOFAI vs Symbolist ML
- Weak AI and Strong AI
 - Task vs general intelligence

Basics of ML

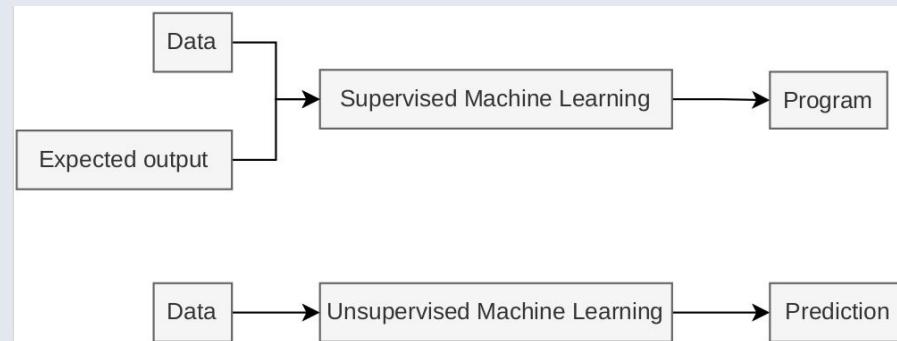
Approaches and tasks

GOFAI



Deductive approach

Machine Learning

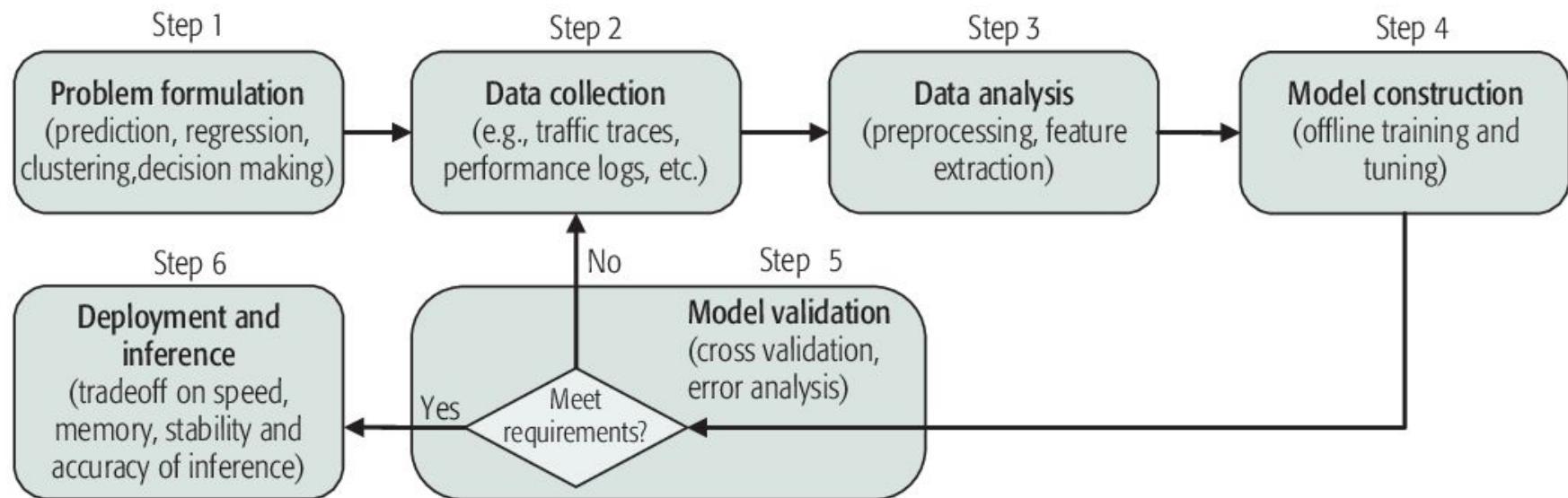


Inductive approach

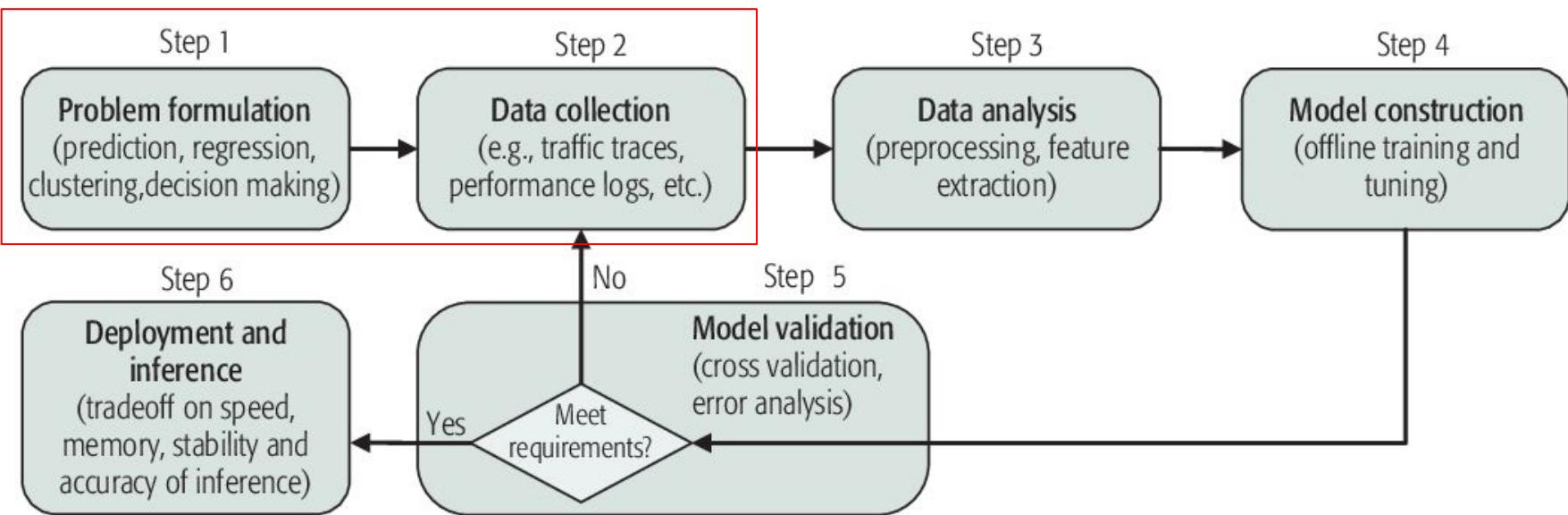
Learning: a modification of a behavior based on an experience - Fabien Benureau (2015).

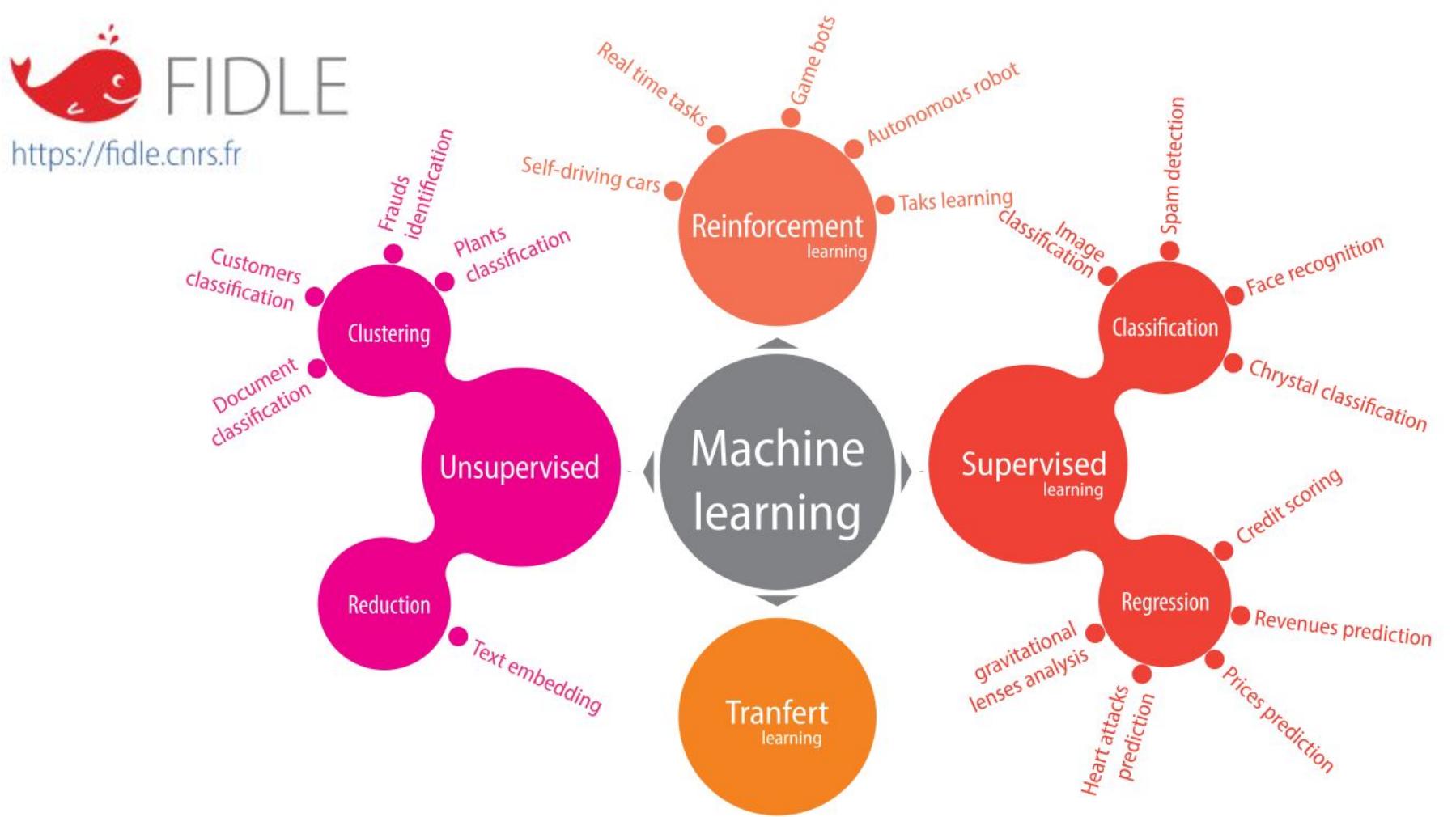
Machine learning: learn a task, that is, to improve its performance for that task, based on experience (data).

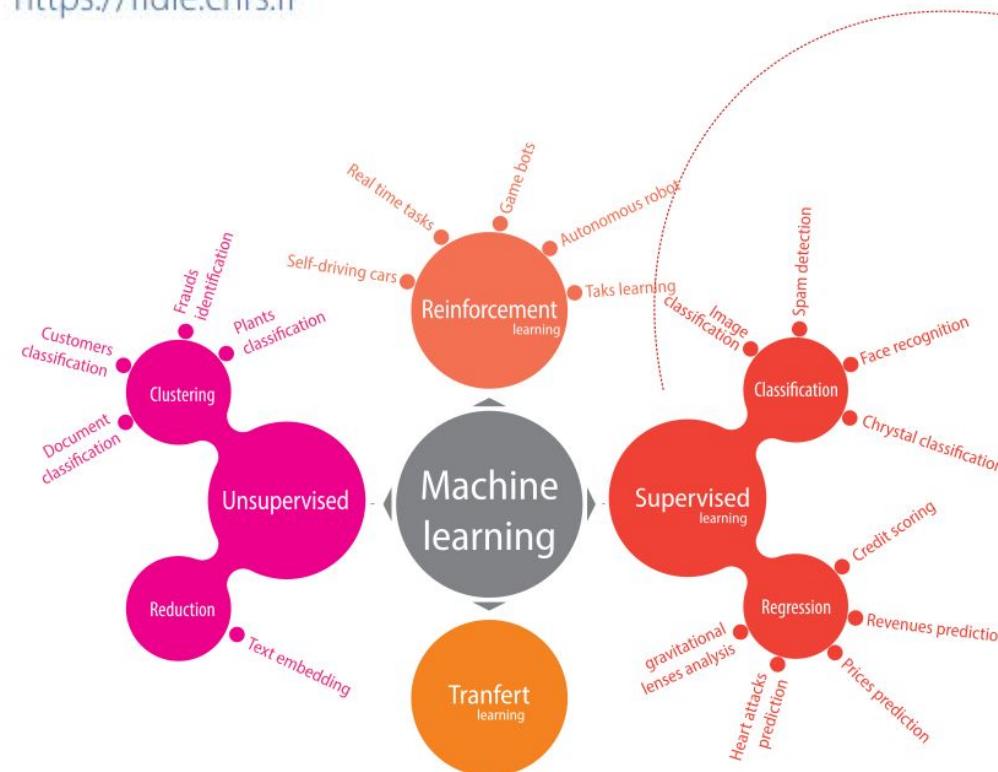
Machine Learning Workflow



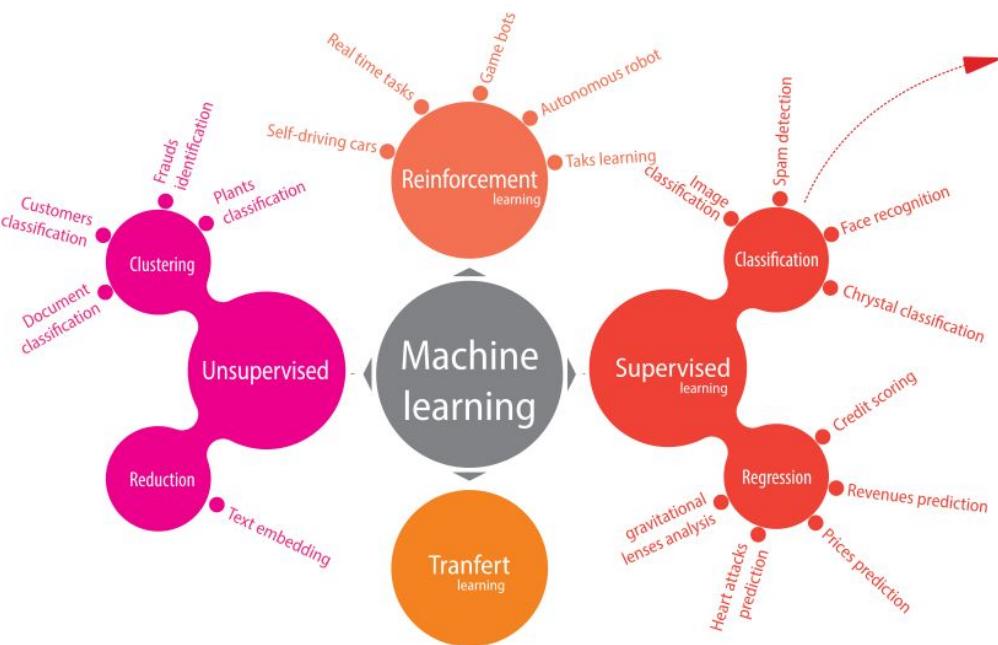
Machine Learning Workflow







Learning from examples



Classification :

Predict qualitative informations



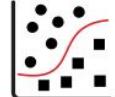
This is a cat

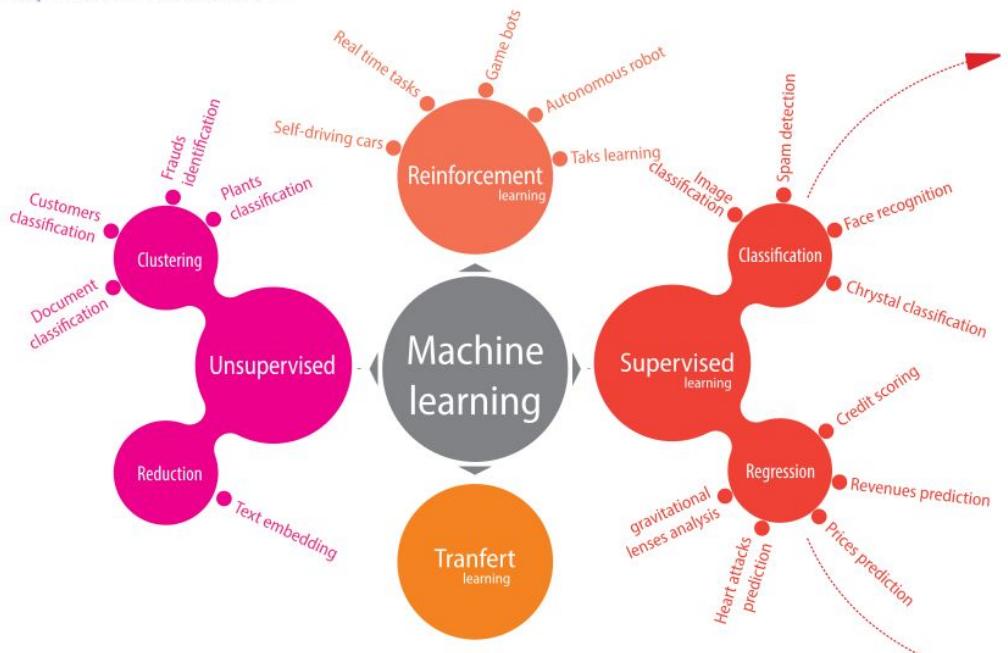


This is a rabbit



Tell me,
what is it ?





Classification :

Predict qualitative informations



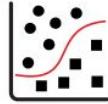
This is a cat



This is a rabbit



Tell me,
what is it ?



Régression :

Predict quantitative informations



150 K€



400 K€



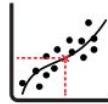
120 K€



100 K€



Tell me,
what's the
price ?

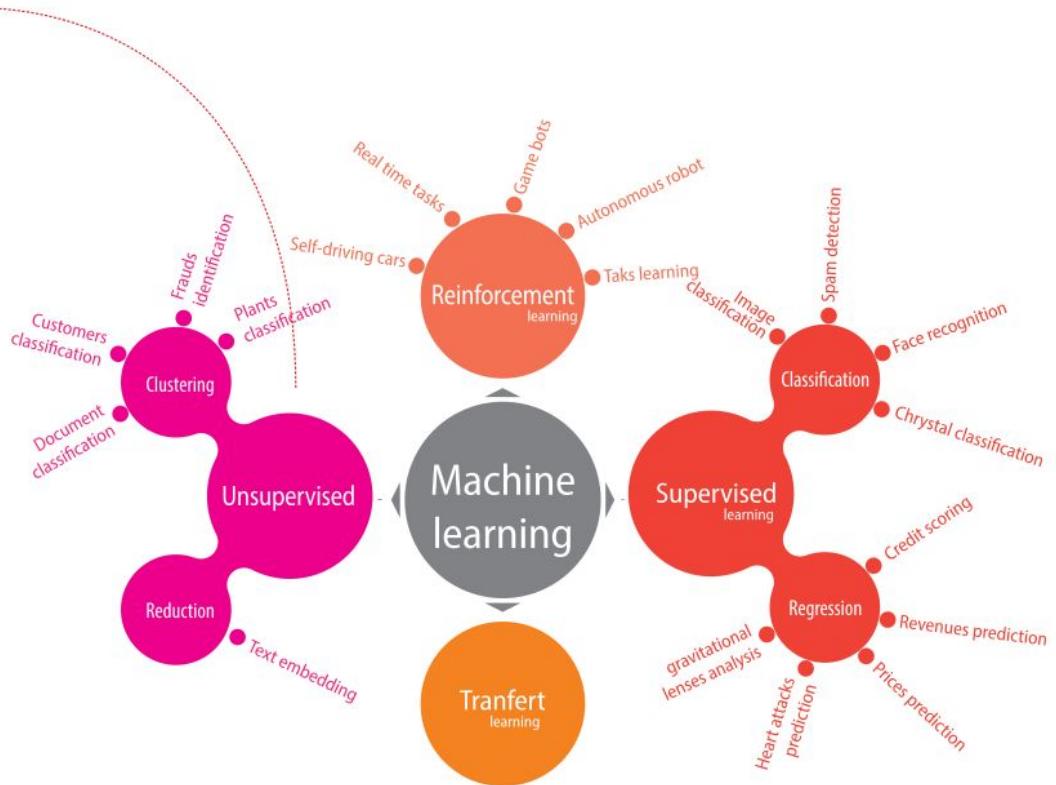


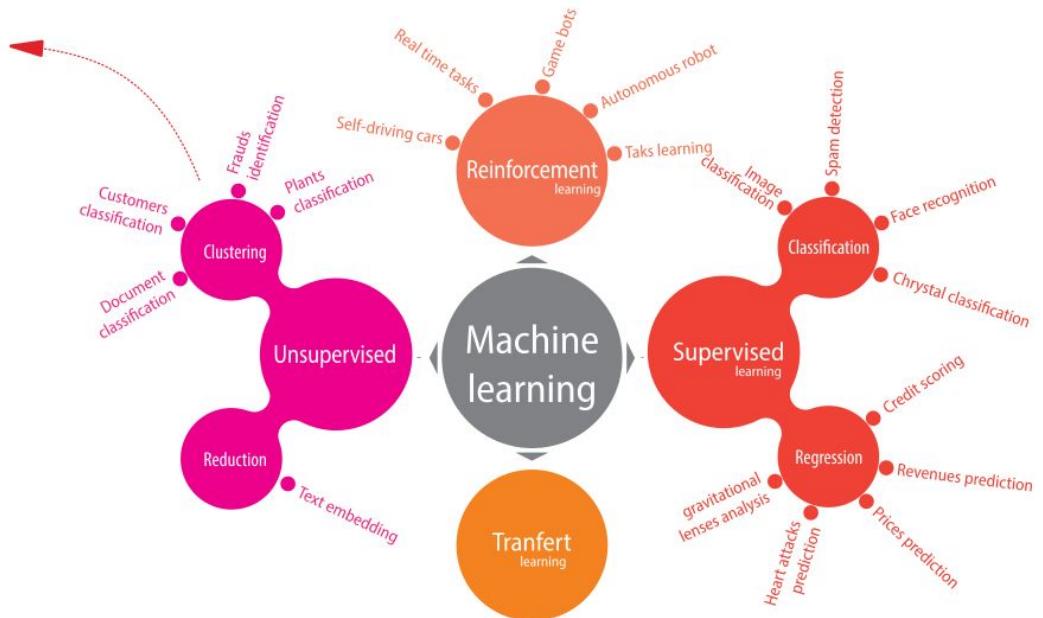
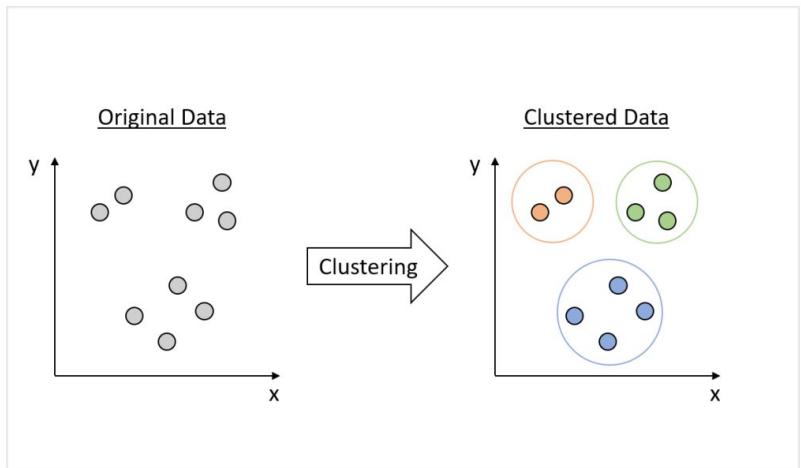


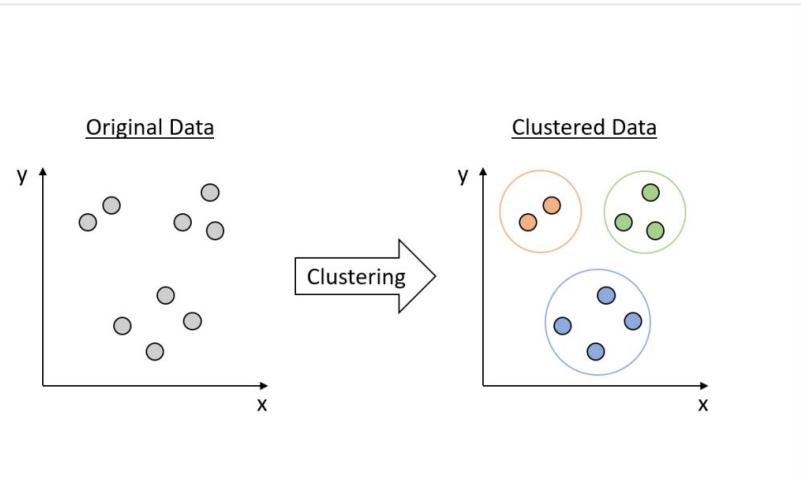
FIDLE

<https://fidle.cnrs.fr>

Learning from data alone





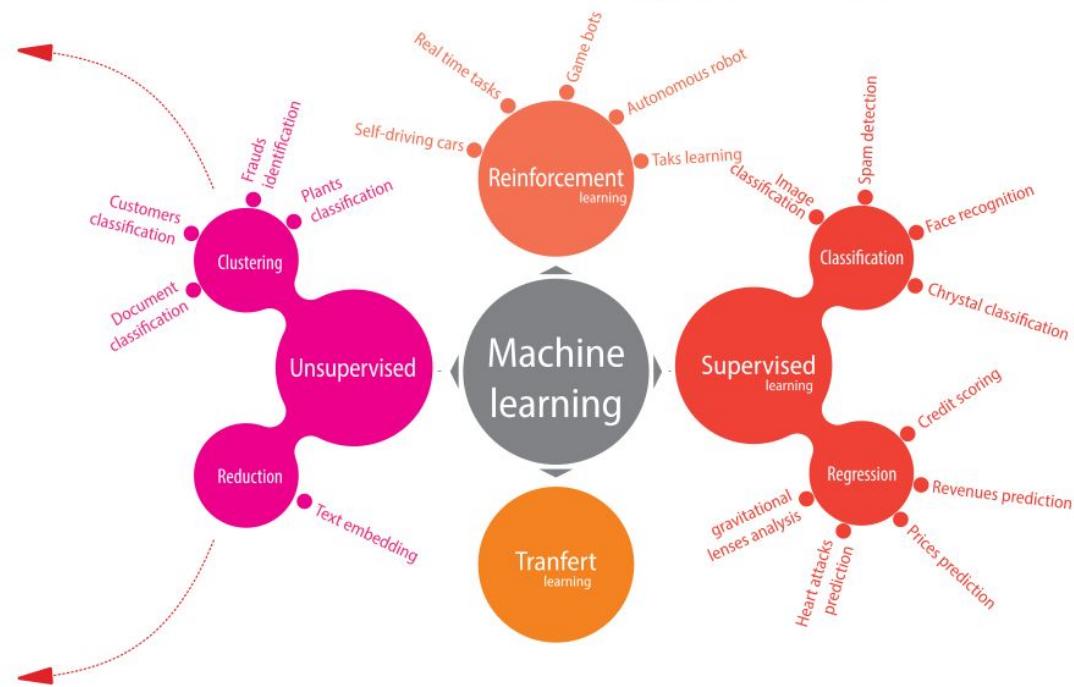


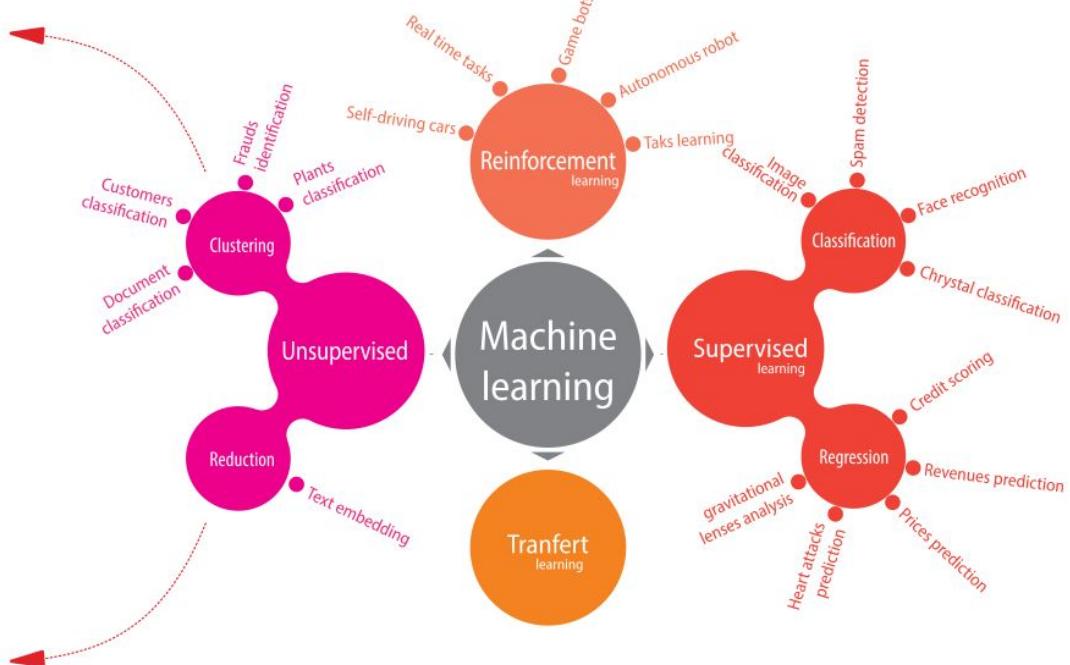
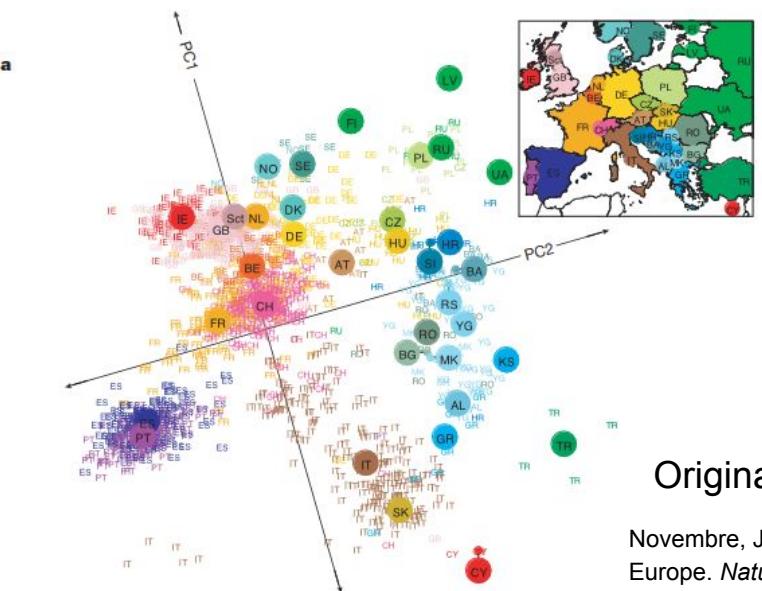
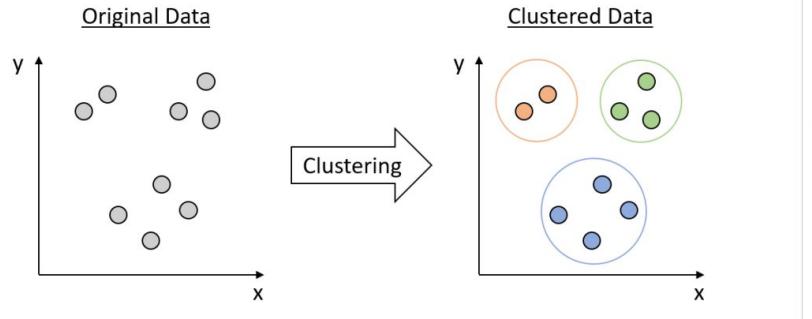
Reduction :

Reduce the number of dimensions



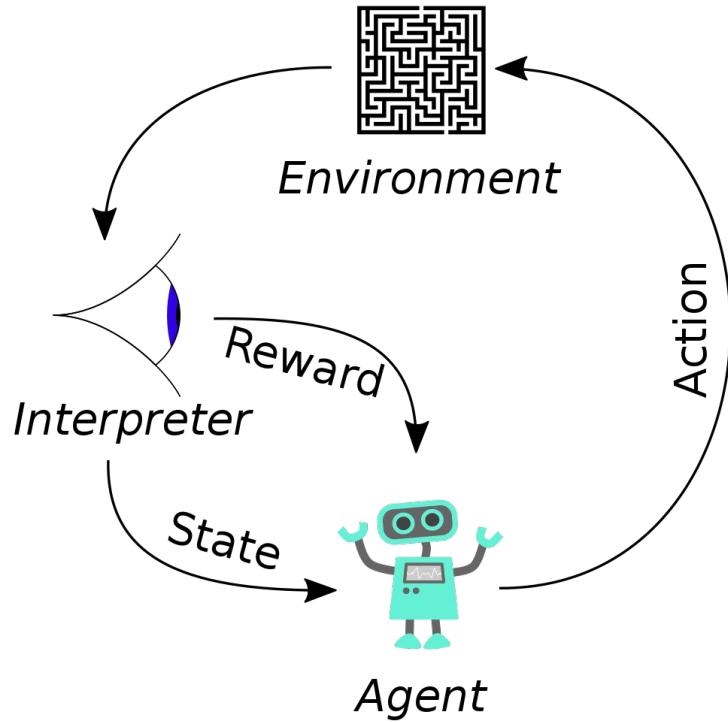
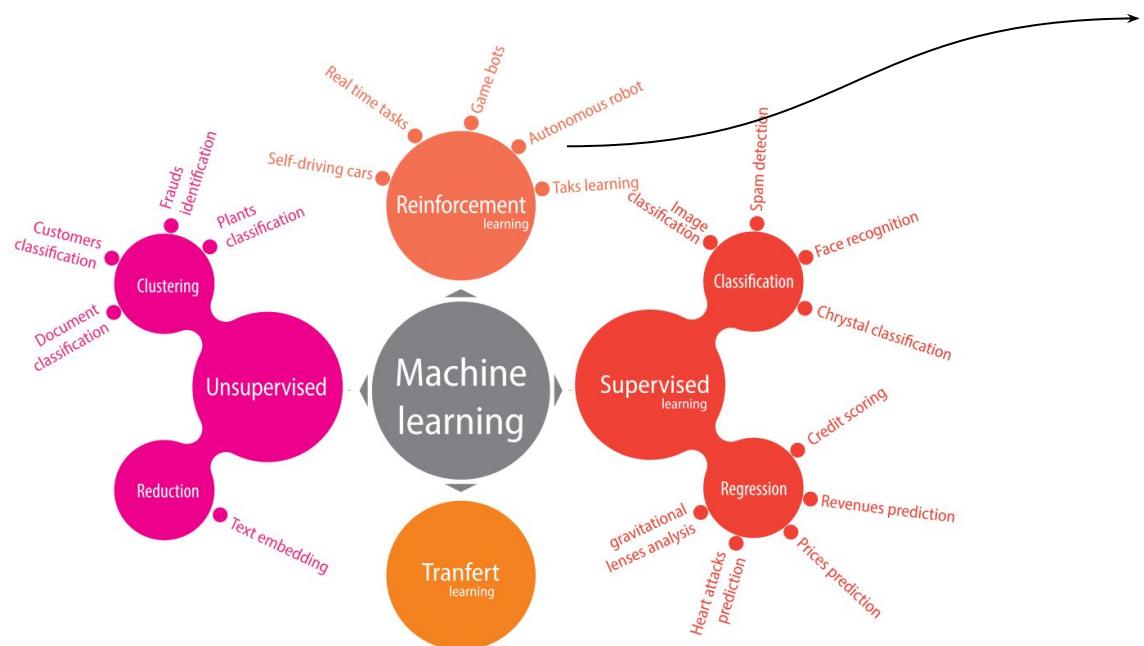
Simplify
while
keeping
meaning

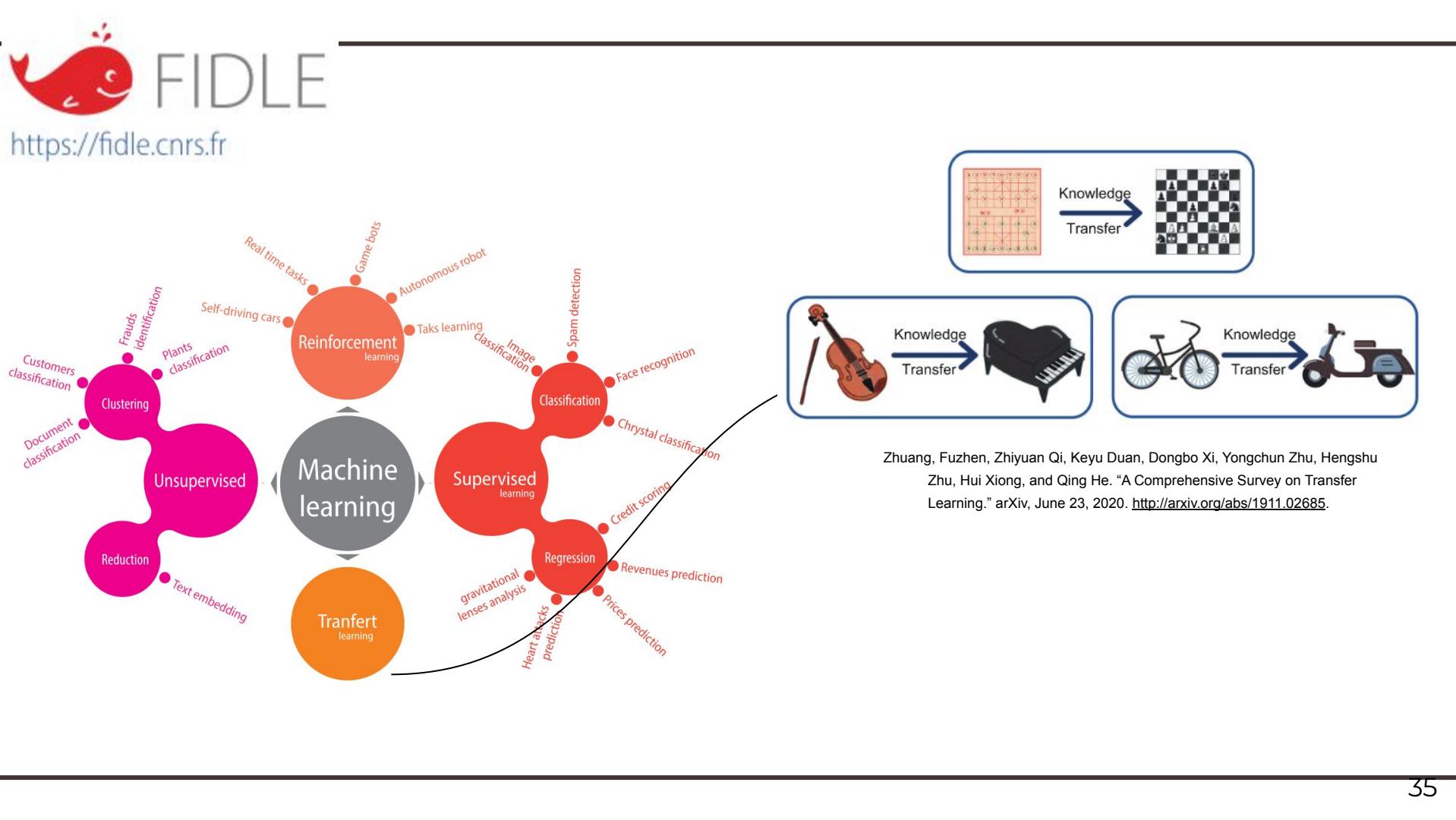




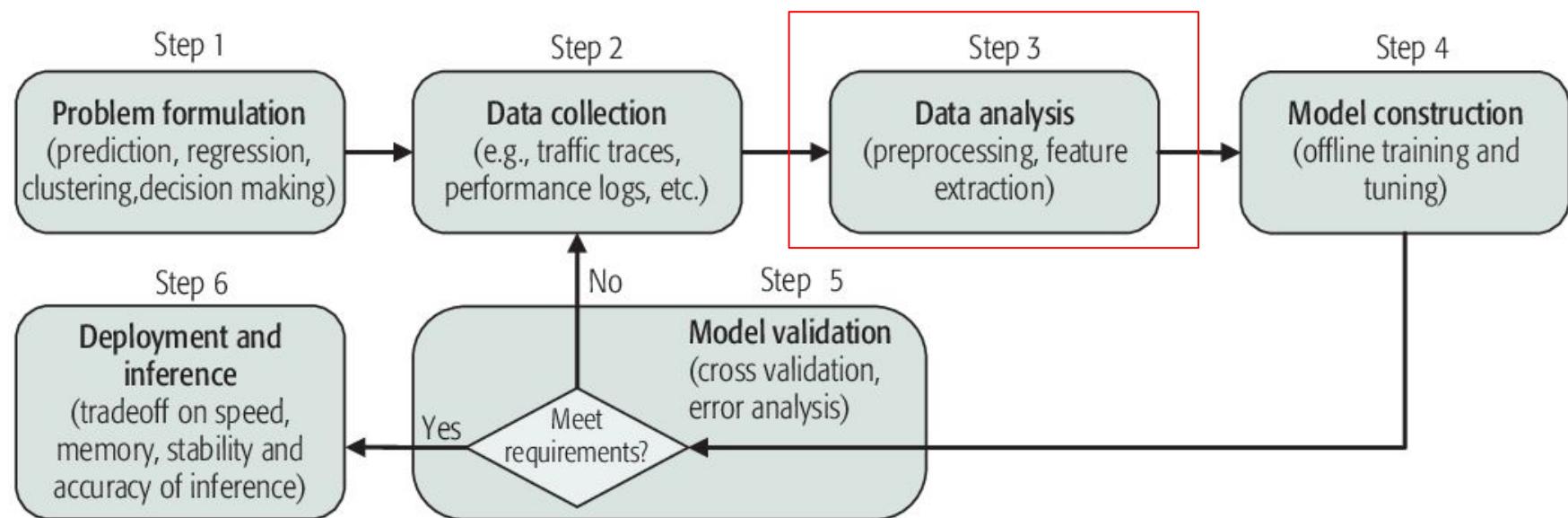
Original data = 500 000 variables !!

Novembre, J., Johnson, T., Bryc, K. et al. Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008). <https://doi.org/10.1038/nature07331>

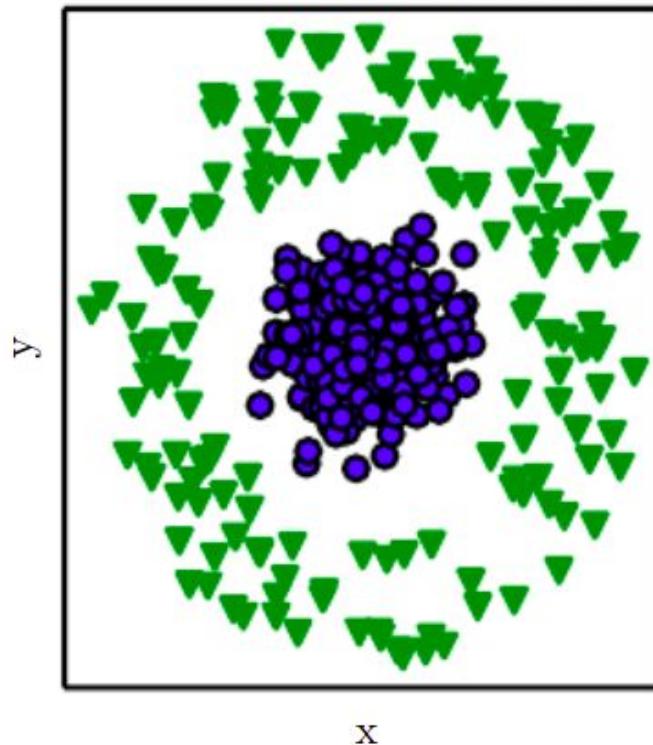




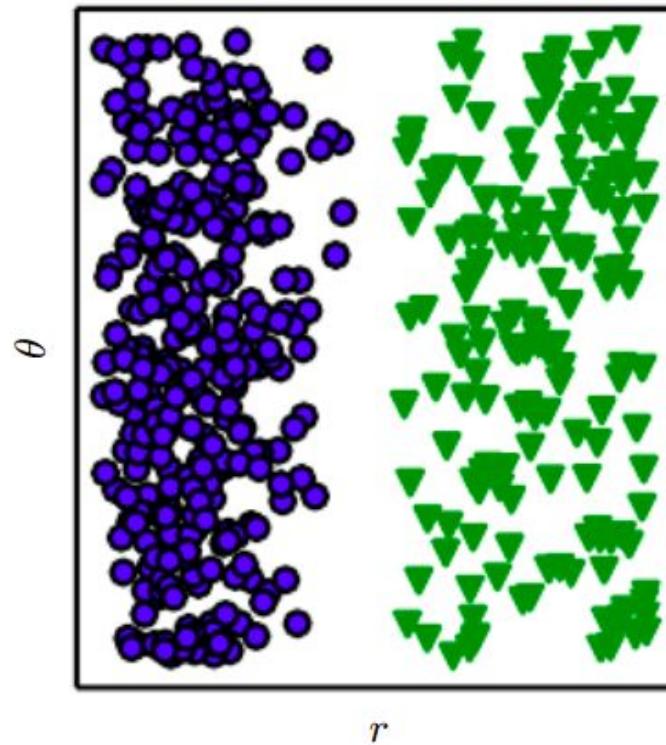
Machine Learning Workflow



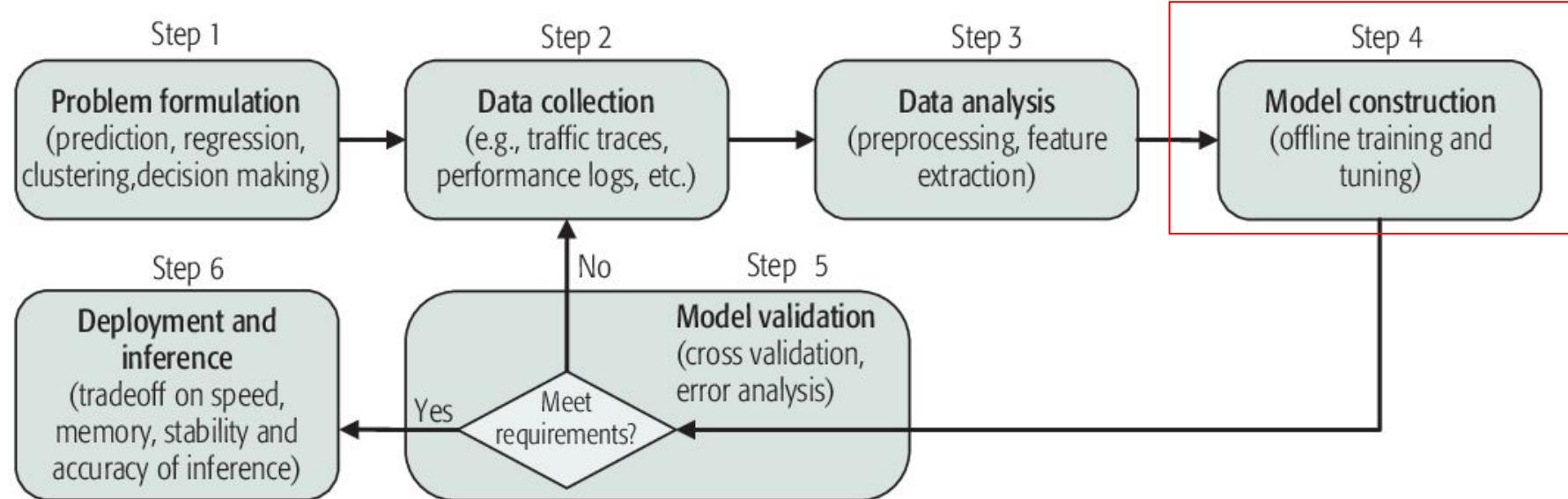
Cartesian coordinates



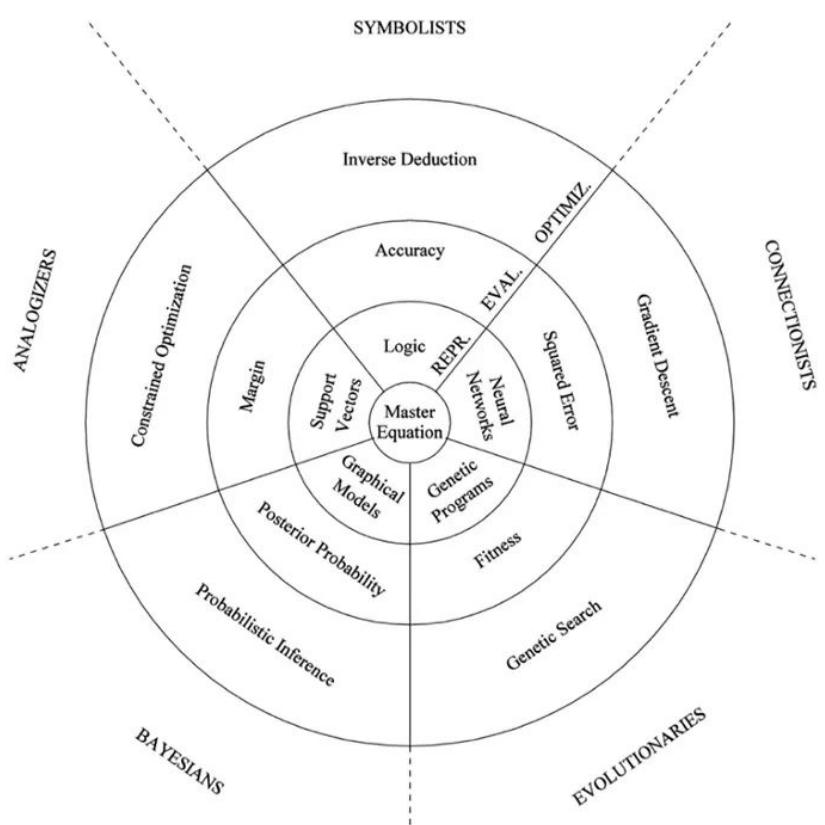
Polar coordinates



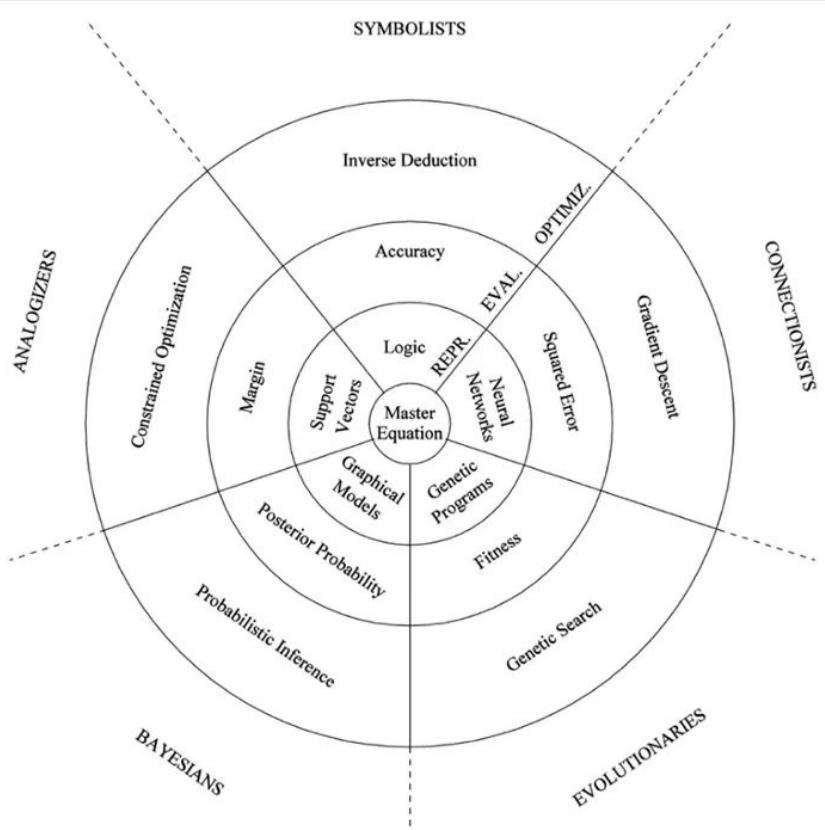
Machine Learning Workflow



The five tribes of Machine Learning



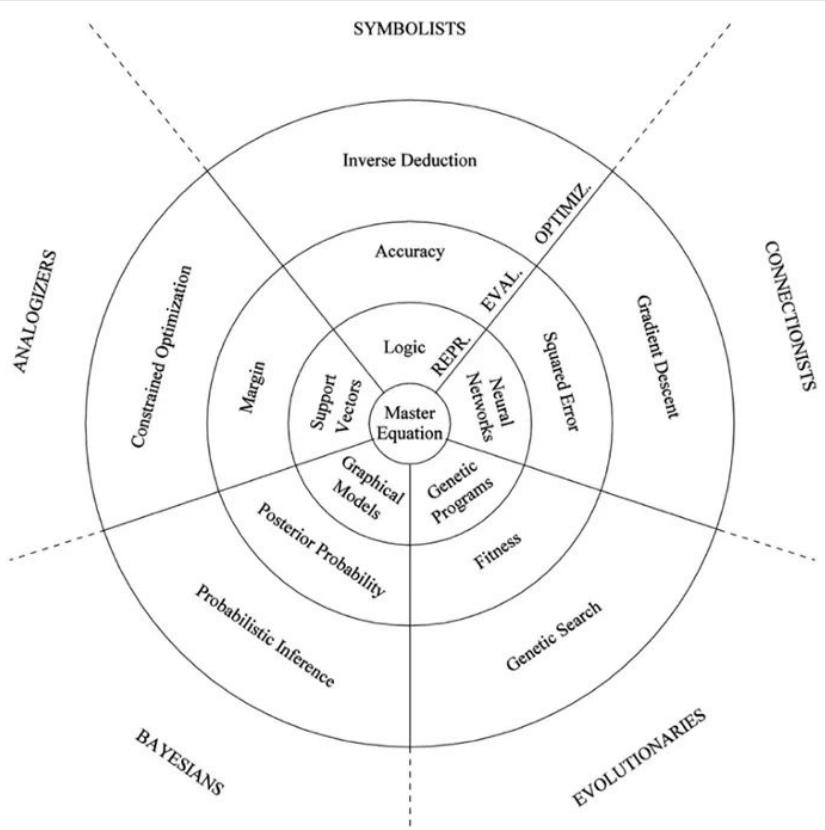
Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



The five tribes of Machine Learning

Symbolists: based on logic and inverse deduction

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

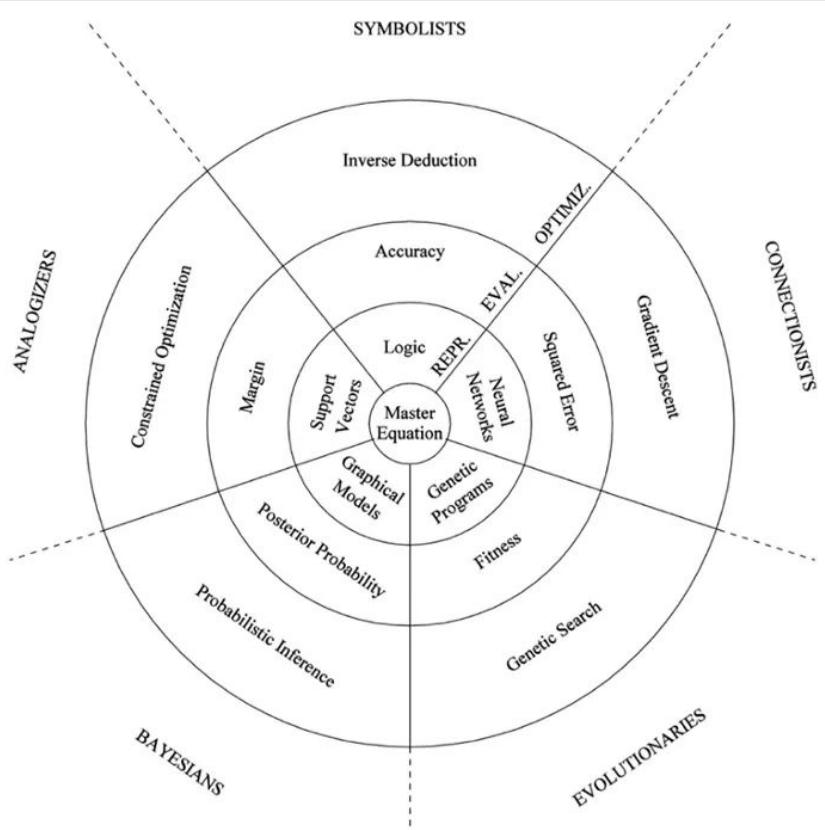


The five tribes of Machine Learning

Symbolists: based on logic and inverse deduction

Connectionists: based on neural networks

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



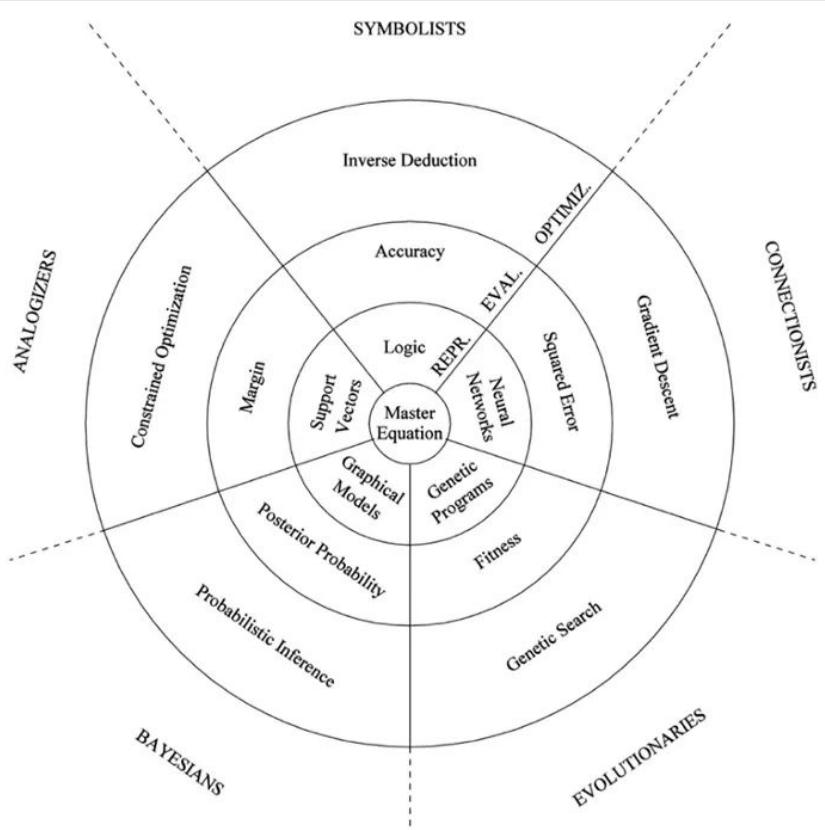
The five tribes of Machine Learning

Symbolists: based on logic and inverse deduction

Connectionists: based on neural networks

Evolutionaries: mimic genetic programs (evolution by mutation)

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



The five tribes of Machine Learning

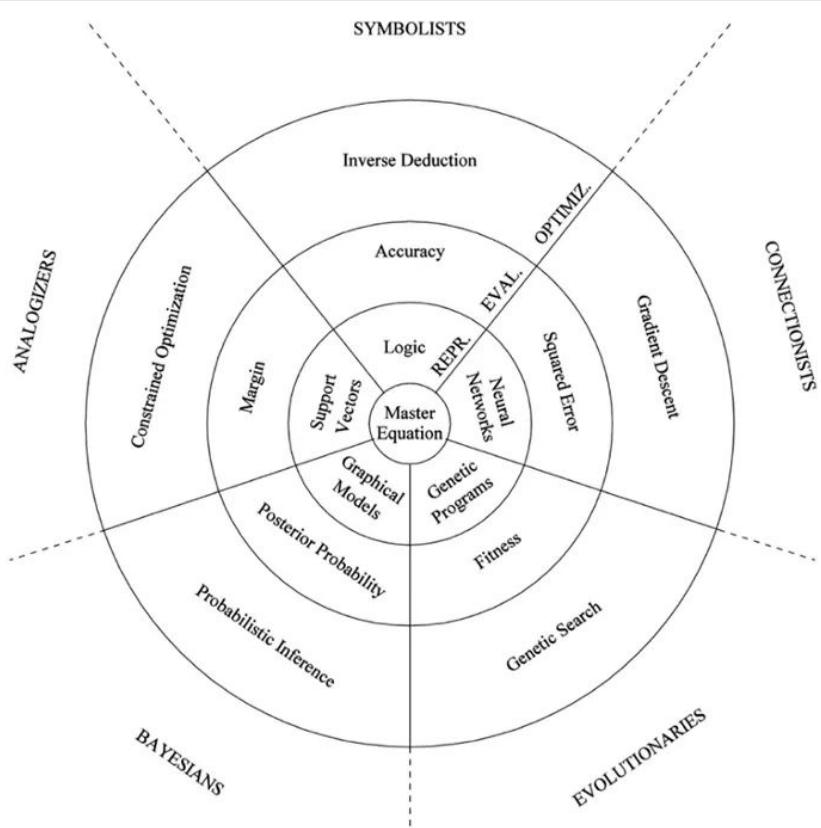
Symbolists: based on logic and inverse deduction

Connectionists: based on neural networks

Evolutionaries: mimic genetic programs (evolution by mutation)

Bayesians: based on probabilistic inference

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



The five tribes of Machine Learning

Symbolists: based on logic and inverse deduction

Connectionists: based on neural networks

Evolutionaries: mimic genetic programs (evolution by mutation)

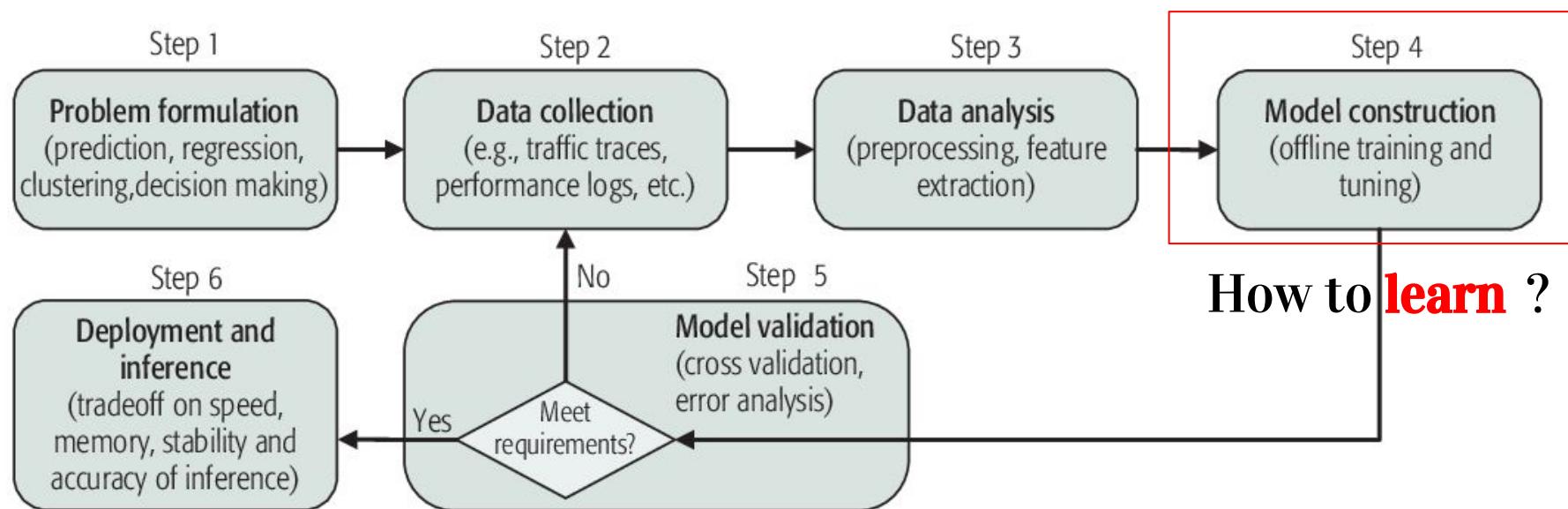
Bayesians: based on probabilistic inference

Analogizers: based on computing distances and similarities

More on that later!

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

Machine Learning Workflow



How to **learn** ?

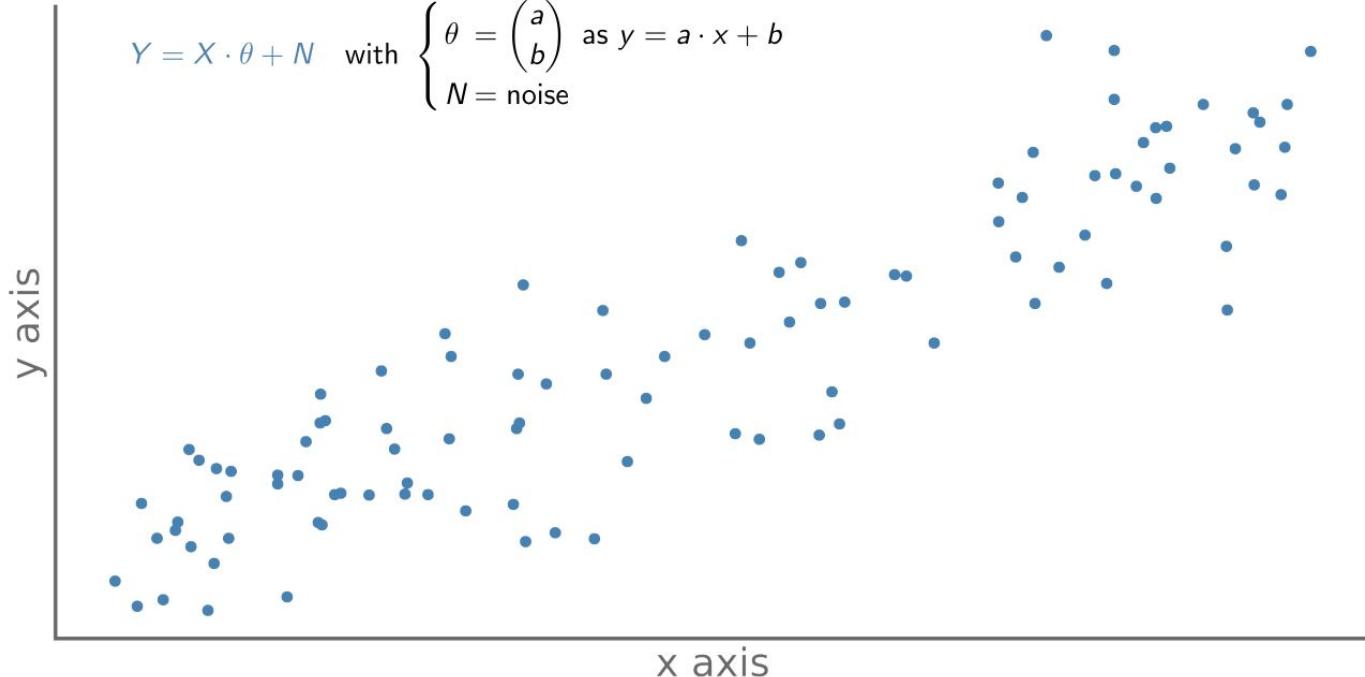
- Optimizing an **objective function**

A few examples ...

Linear regression

We have a phenomenon, for which we have observations

$$Y = X \cdot \theta + N \quad \text{with} \quad \begin{cases} \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ as } y = a \cdot x + b \\ N = \text{noise} \end{cases}$$

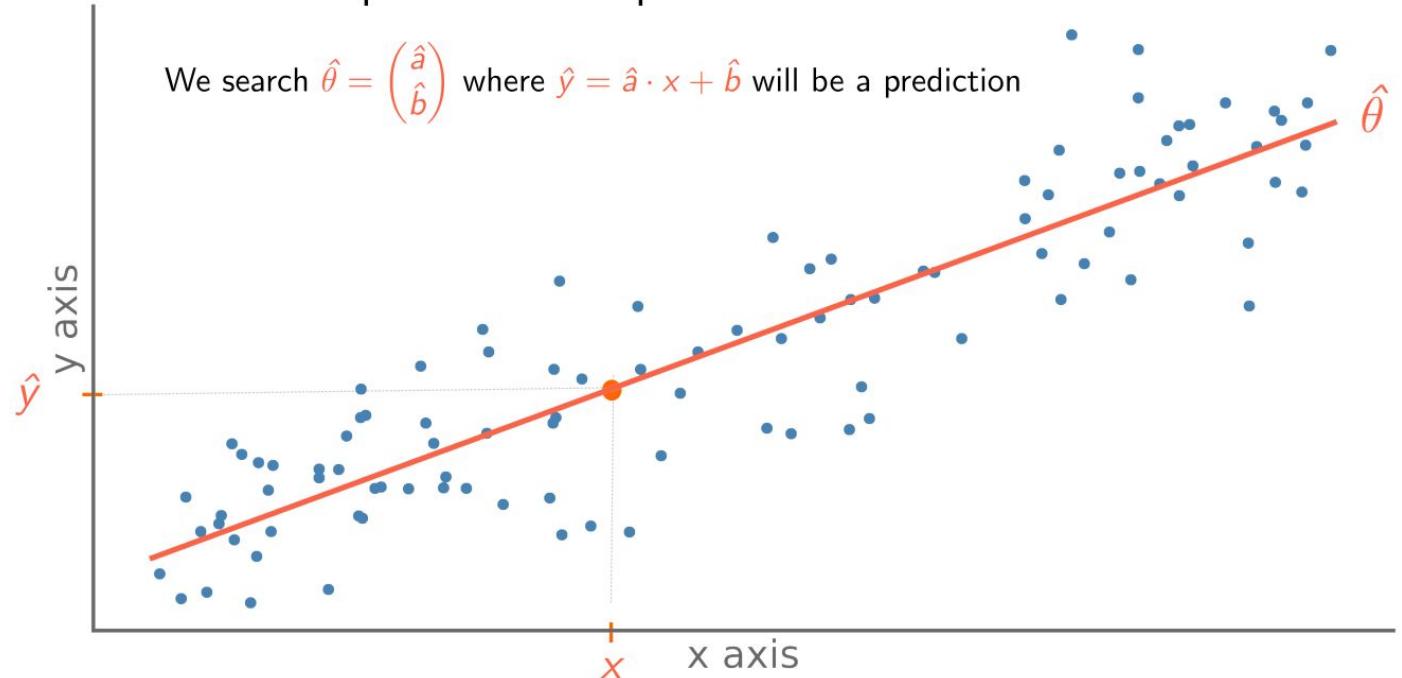




Linear regression

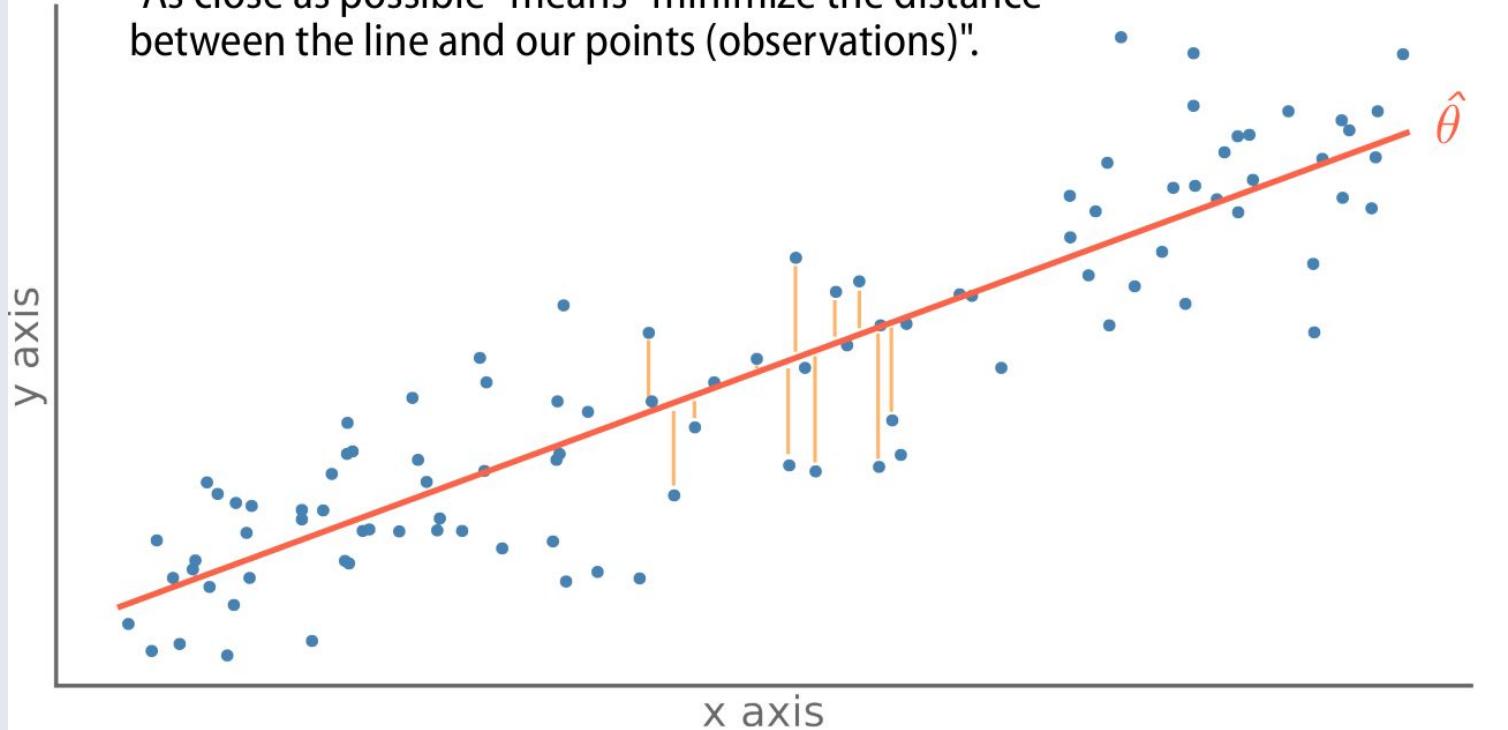
We are looking for a straight line that passes « as close as possible » to our points.

We search $\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ where $\hat{y} = \hat{a} \cdot x + \hat{b}$ will be a prediction



Linear regression

"As close as possible" means "minimize the distance between the line and our points (observations)".

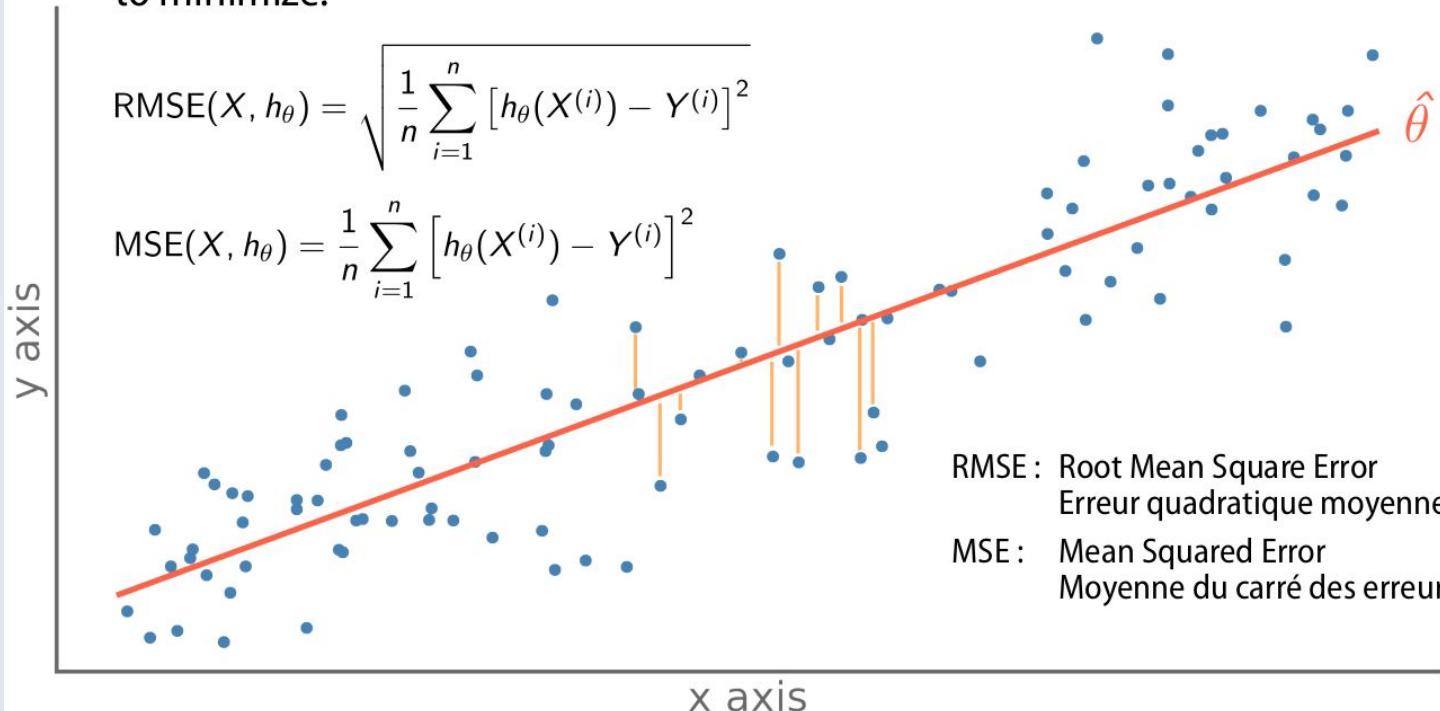


Linear regression

For this, we will use an «loss function », which we will try to minimize.

$$\text{RMSE}(X, h_\theta) = \sqrt{\frac{1}{n} \sum_{i=1}^n [h_\theta(X^{(i)}) - Y^{(i)}]^2}$$

$$\text{MSE}(X, h_\theta) = \frac{1}{n} \sum_{i=1}^n [h_\theta(X^{(i)}) - Y^{(i)}]^2$$

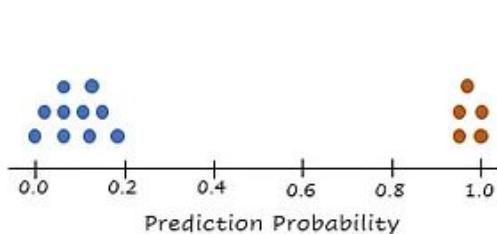


Classification

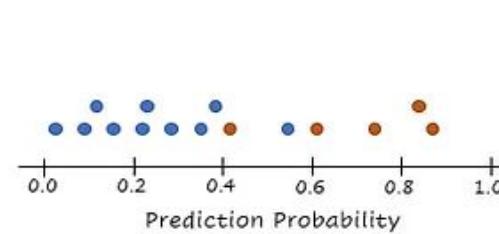
$$Logloss_i = -[y_i \ln p_i + (1 - y_i) \ln(1 - p_i)]$$

Email Seq	Content	Actual Class	Prediction Probability	Prediction Class
121	Hi, If you want to earn quick \$1000 bucks, click on the link below.	1	0.95	1
343	Hi, Hope you are doing fine. I just wanted to circle back and confirm if you had chance to think about the gardening services that our firm provides. Let me know if you have any questions. Thanks, Dave.	0	0.08	0
876	Hello Charlene, I am from IT department. I noticed that your computer is over-utilizing CPU and throwing warning signals to the remote server. Could you please connect with me through the following link if you intend to rectify the issue? I am available from 8AM-5PM. https://www.connectIT.com .	1	0.25	0

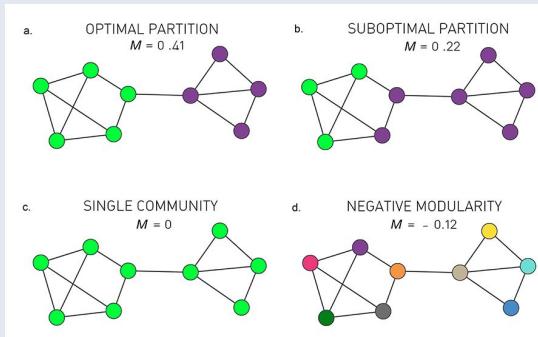
Lower log-loss score



Higher log-loss score



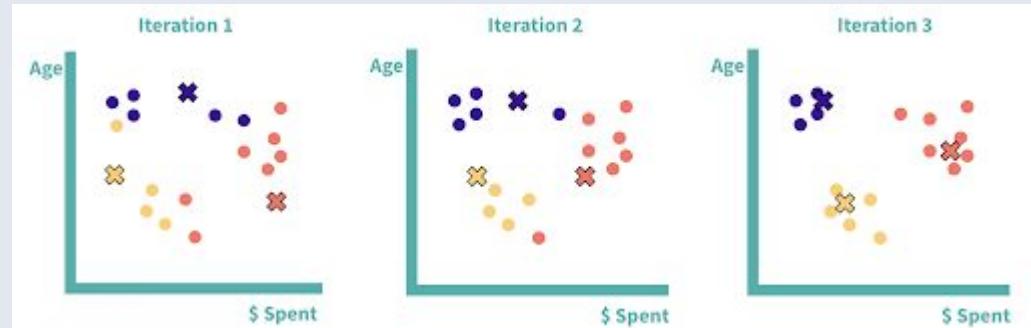
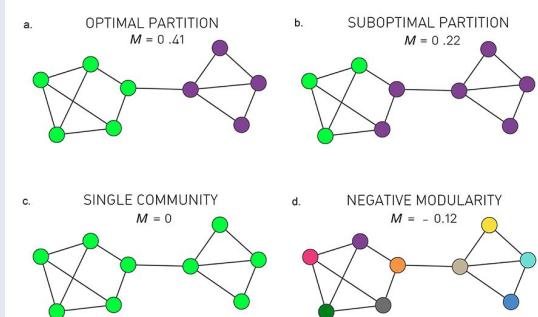
Clustering



Source: Network Science by Albert-László Barabási. Accessed August 29, 2023.

<http://networksciencebook.com/>.

Clustering



Source: Network Science by Albert-László Barabási. Accessed August 29, 2023.

<http://networksciencebook.com/>.

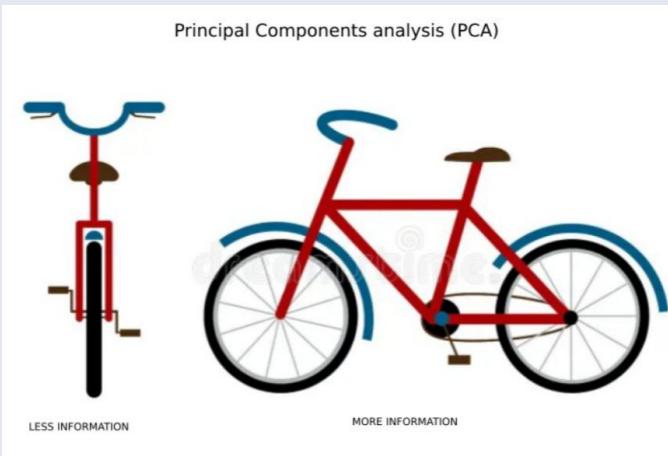
number of clusters number of cases centroid for cluster j

case i

objective function $\rightarrow J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$

distance function

Dimension reduction

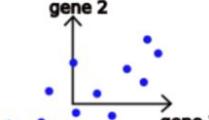
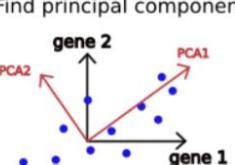


Source: Courtesy of Solène Song

OUTLINE : ON A DUMMY EXAMPLE TO GO FROM 2D TO 1D

- 1) Data

	Features	
	gene 1	gene 2
cell 1	0.1	0.4
cell 2	0.3	0.8
cell 3	0.4	0.5
cell


- 2) Find principal components
 

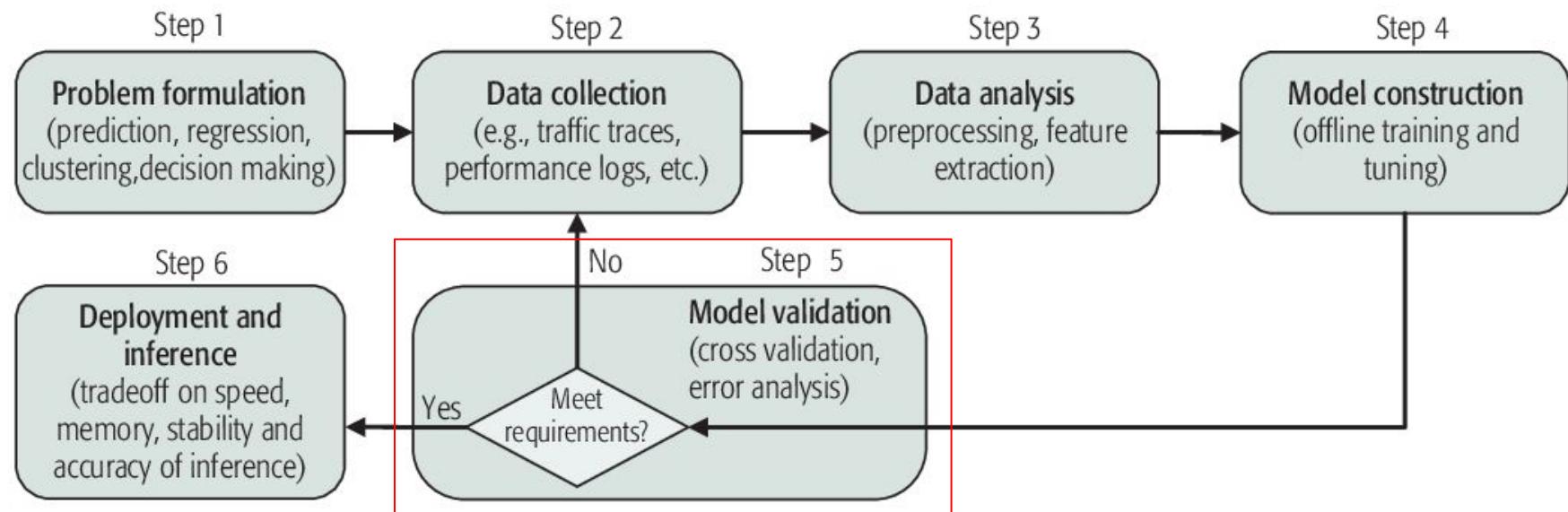
PCA1 accounts for 91% for the total variation (points are maximally spread out)
 PCA2 accounts for 9% for the total variation
- 3) Reduce dimensionality
 


Low dimension representation

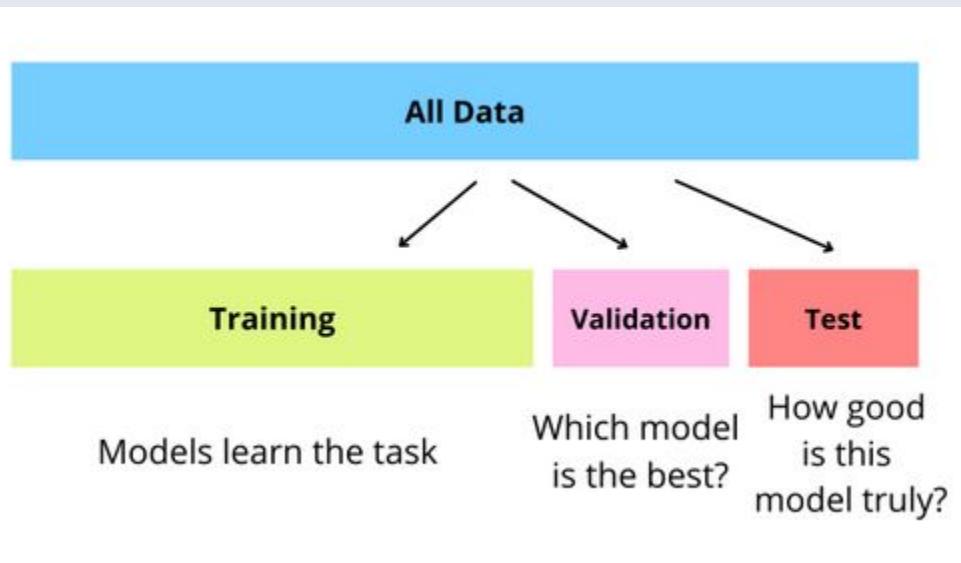
How to **learn** ?

- Optimizing an **objective function**

Machine Learning Workflow

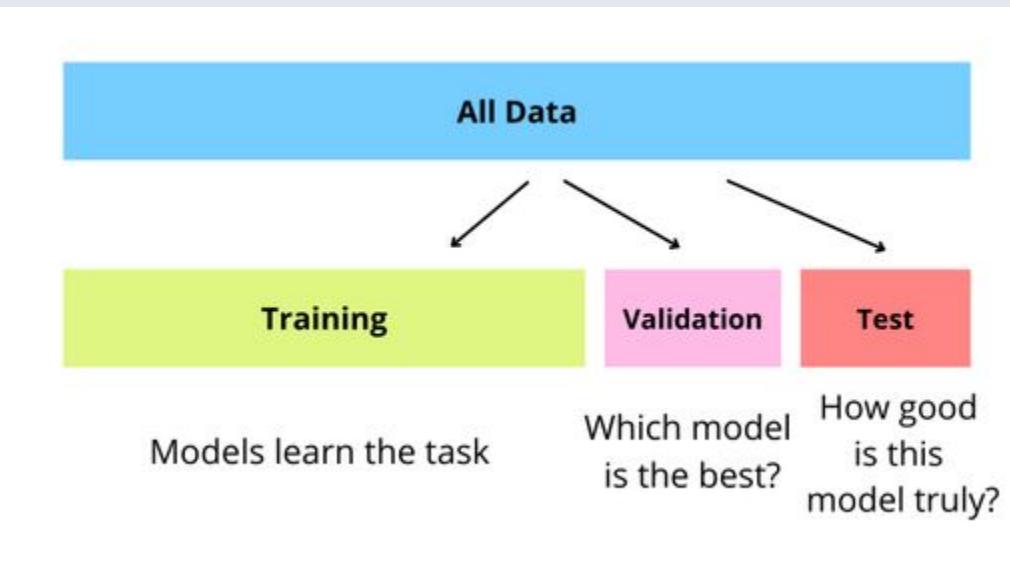


Supervised ML



Training Dataset: The sample of data used to fit the model.

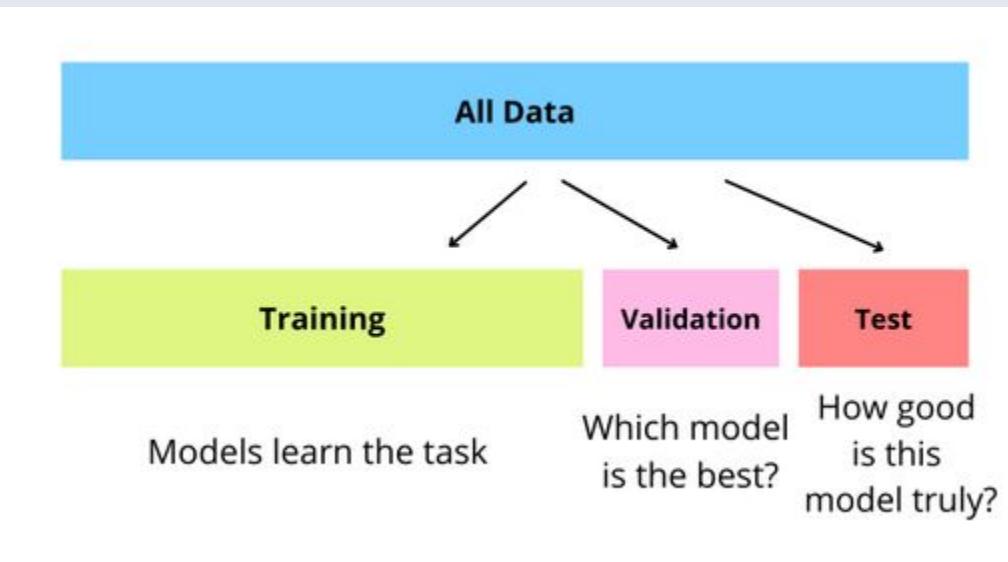
Supervised ML



Training Dataset: The sample of data used to fit the model.

Validation Dataset: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.

Supervised ML

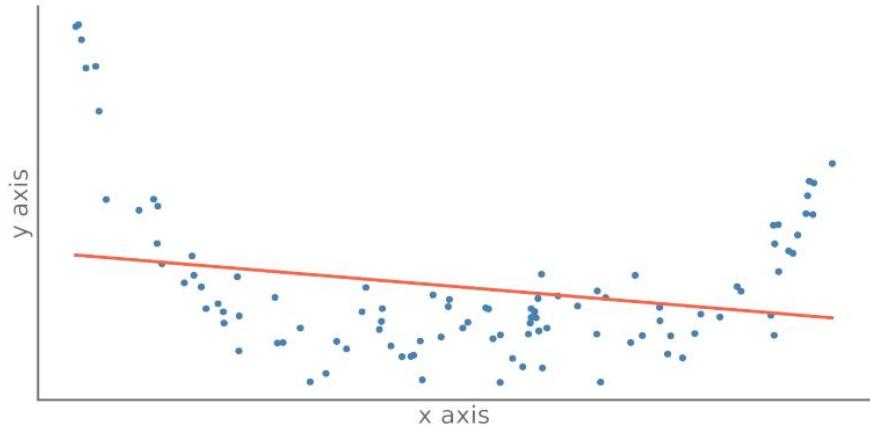


Training Dataset: The sample of data used to fit the model.

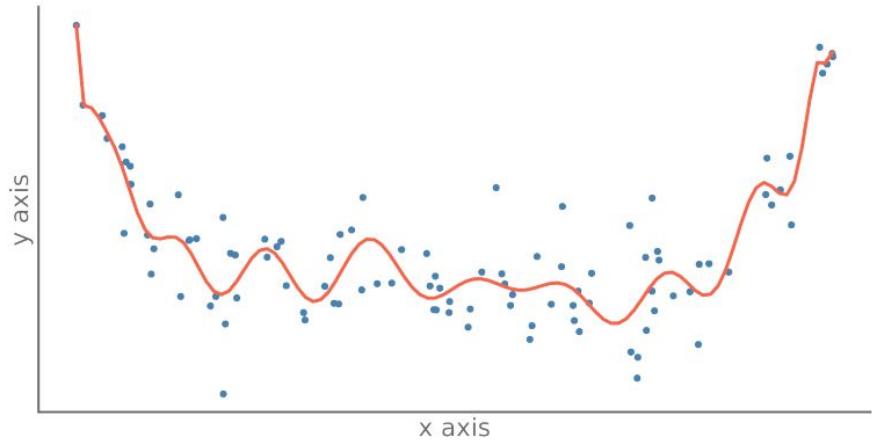
Validation Dataset: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.

Test Dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

Underfitting and overfitting



Underfitting



Overfitting

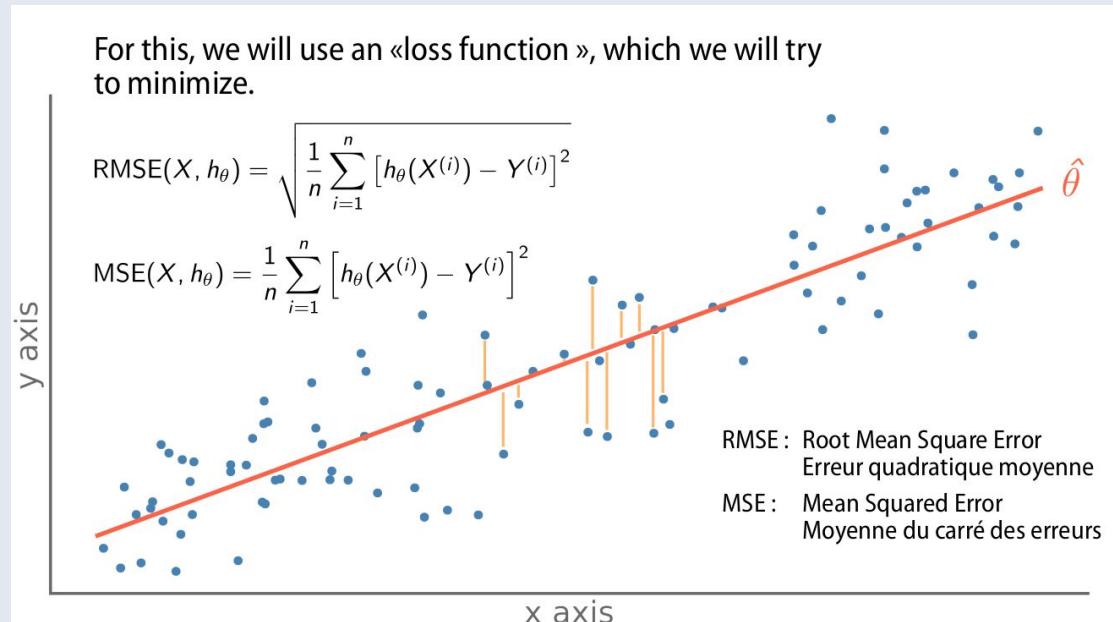
Performance metrics

- Regression
 - MSE, RMSE, ...

For this, we will use an «loss function », which we will try to minimize.

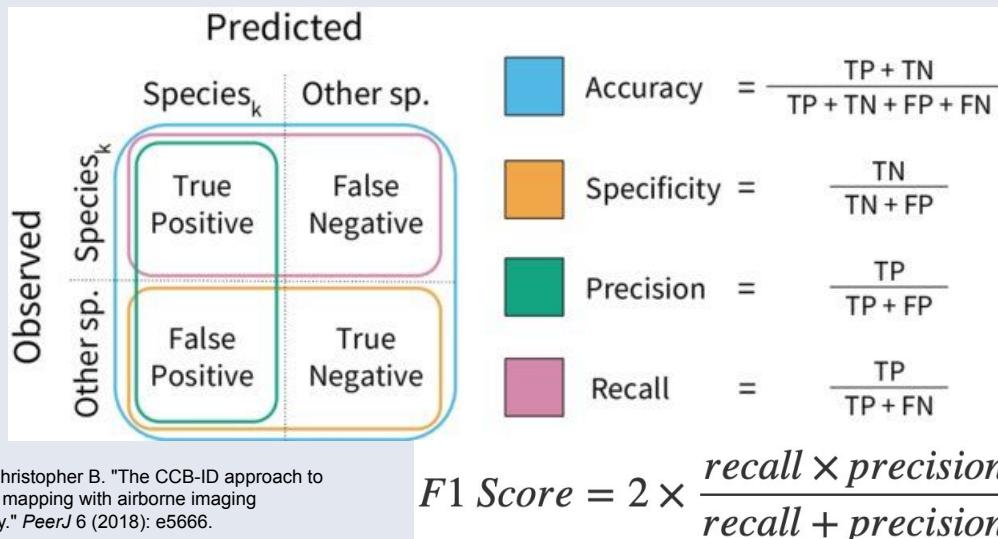
$$\text{RMSE}(X, h_{\theta}) = \sqrt{\frac{1}{n} \sum_{i=1}^n [h_{\theta}(X^{(i)}) - Y^{(i)}]^2}$$

$$\text{MSE}(X, h_{\theta}) = \frac{1}{n} \sum_{i=1}^n [h_{\theta}(X^{(i)}) - Y^{(i)}]^2$$

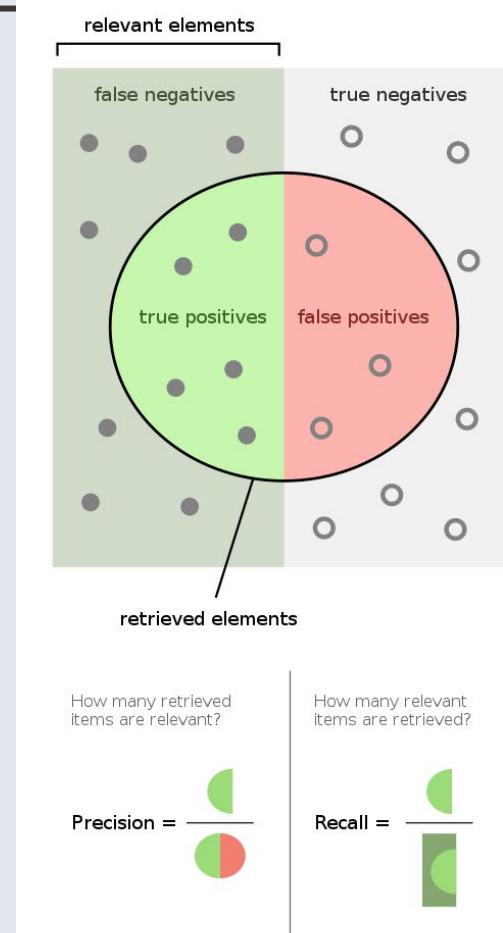


Performance metrics

- Regression
 - MSE, RMSE
- Classification
 - Precision, Recall, F1-score, AUROC

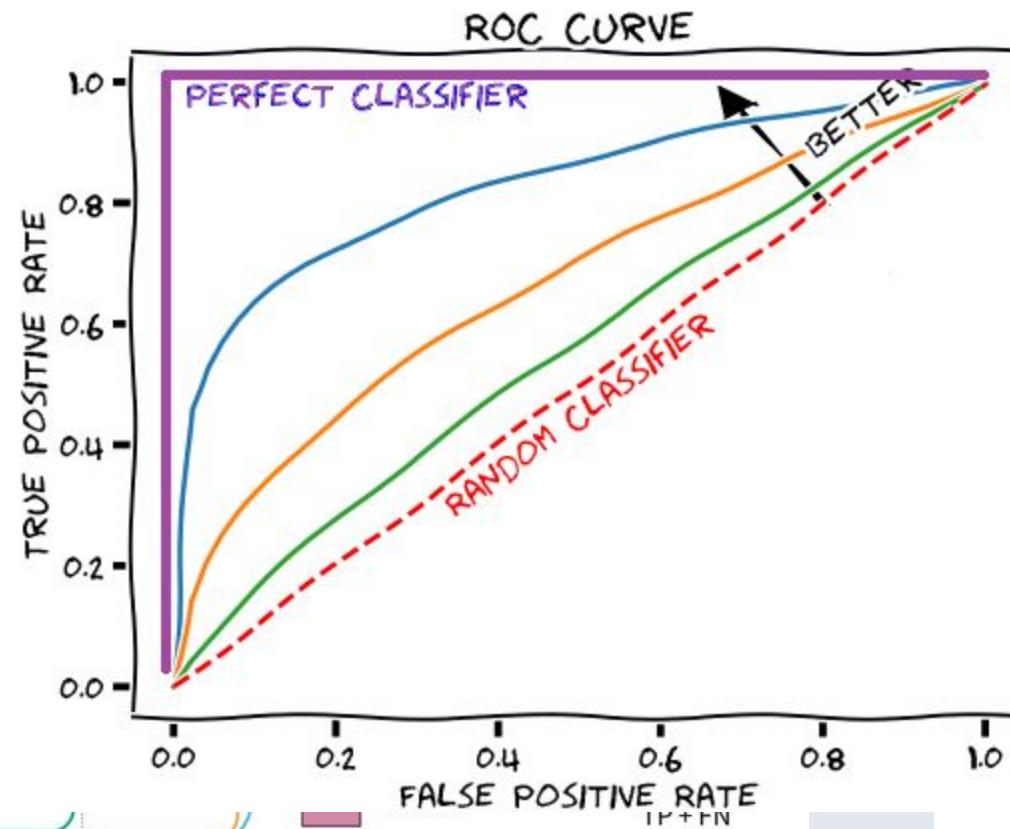
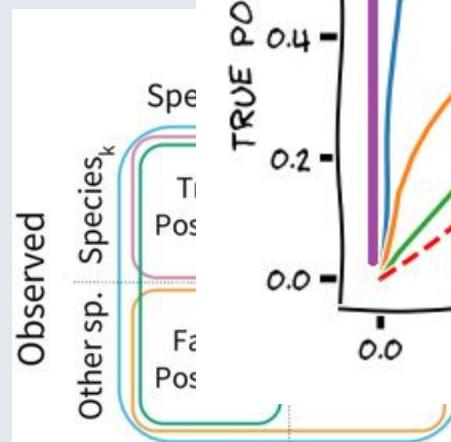


Anderson, Christopher B. "The CCB-ID approach to tree species mapping with airborne imaging spectroscopy." *PeerJ* 6 (2018): e5666.

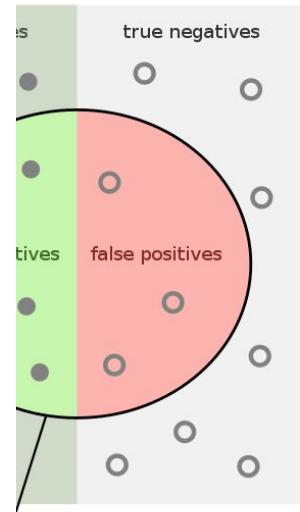


Performance

- Regression
 - MSE, R²
- Classification
 - Precision, Recall, F1



relevant elements



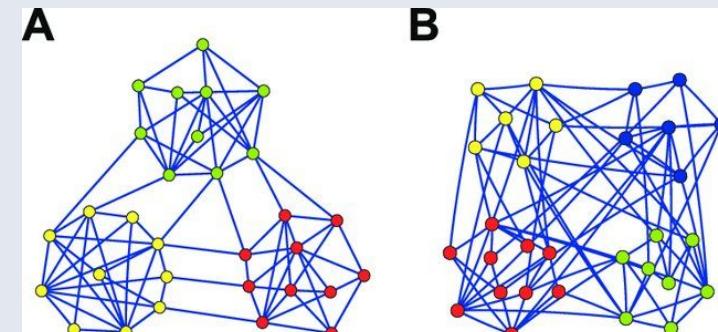
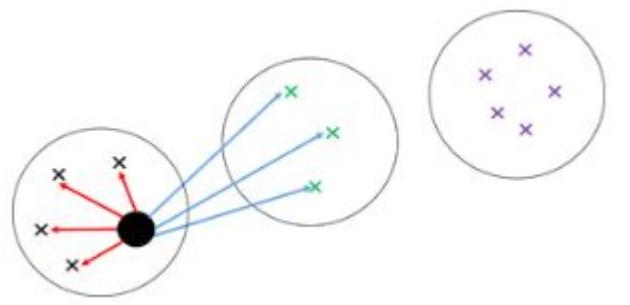
elements

How many relevant items are retrieved?

$$\text{Recall} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items}}$$

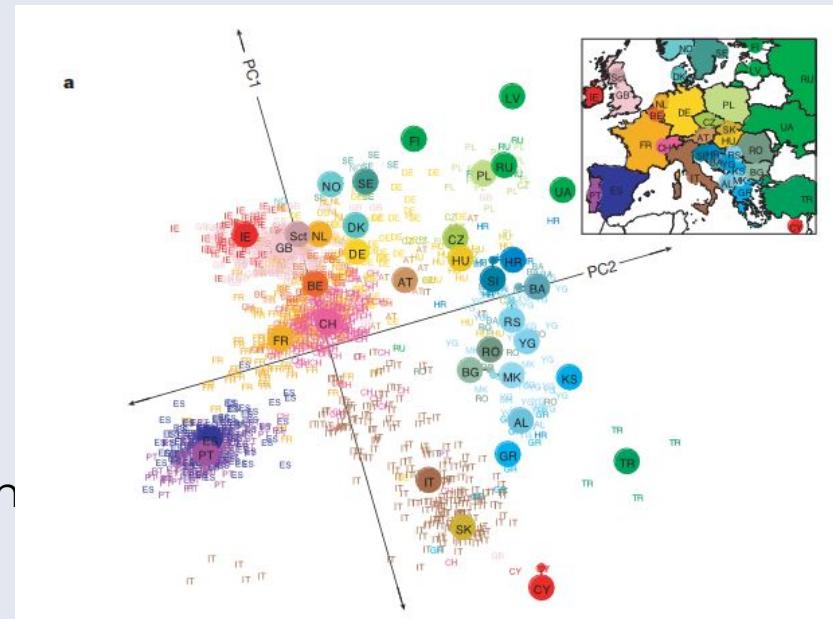
Performance metrics

- Regression
 - MSE, RMSE
- Classification
 - Accuracy, recall, F1-score, AUROC
- Clustering
 - Silhouette score (cohésion, séparation), modularity, ...



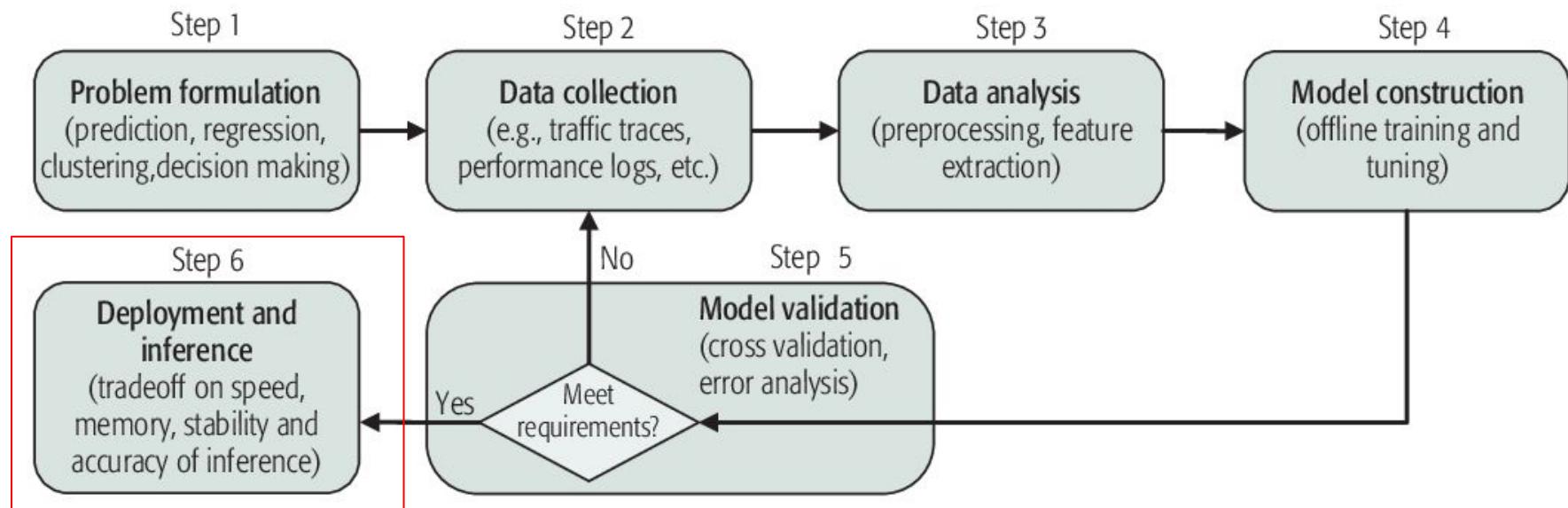
Performance metrics

- Regression
 - MSE, RMSE
- Classification
 - Accuracy, recall, F1-score, AUROC
- Clustering
 - Silhouette score (cohésion, séparation)
- Dimension reduction
 - Visual inspection, classification quality, data reconstruction, ...



Novembre, J., Johnson, T., Bryc, K. et al. Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
<https://doi.org/10.1038/nature07331>

Machine Learning Workflow



Take home message

➤ ML approaches and tasks

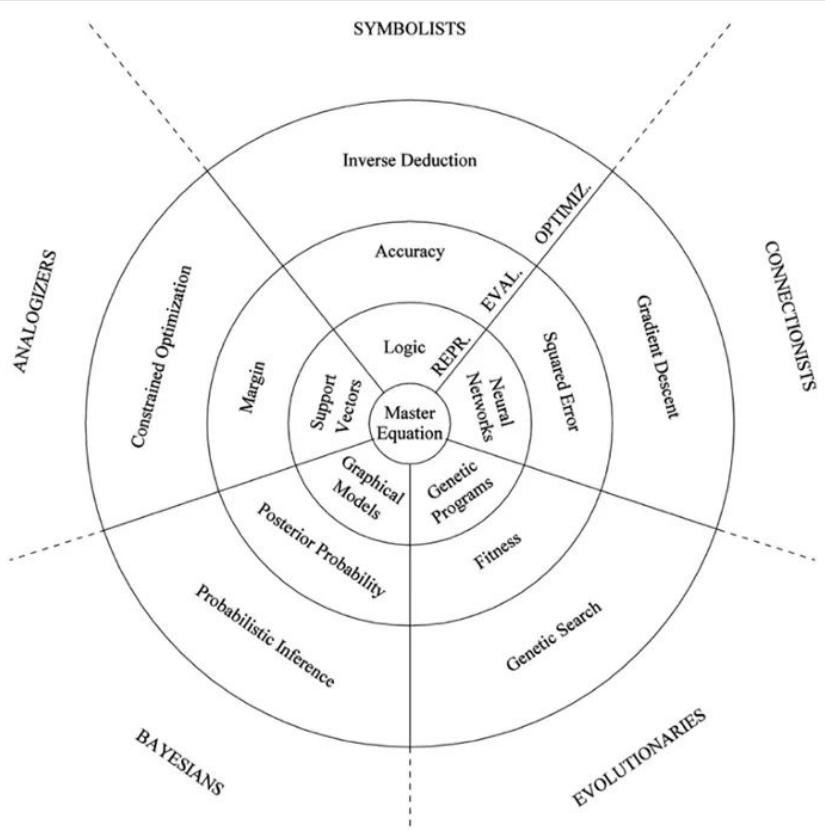
- Supervised
 - Regression, classification
- Unsupervised
 - Clustering, dimension reduction
- Reinforcement learning
- Transfert learning

➤ ML Workflow

- Define the question, translate into a ML task
- Gather and preprocess the data
- Define the model
 - The five tribes of ML
 - Learning algorithm
- Evaluate the model
 - Train and test sets
 - Evaluation metrics
 - Overfitting and underfitting
- Deploy the model

Main approaches

The five tribes of Machine Learning



Connectionists

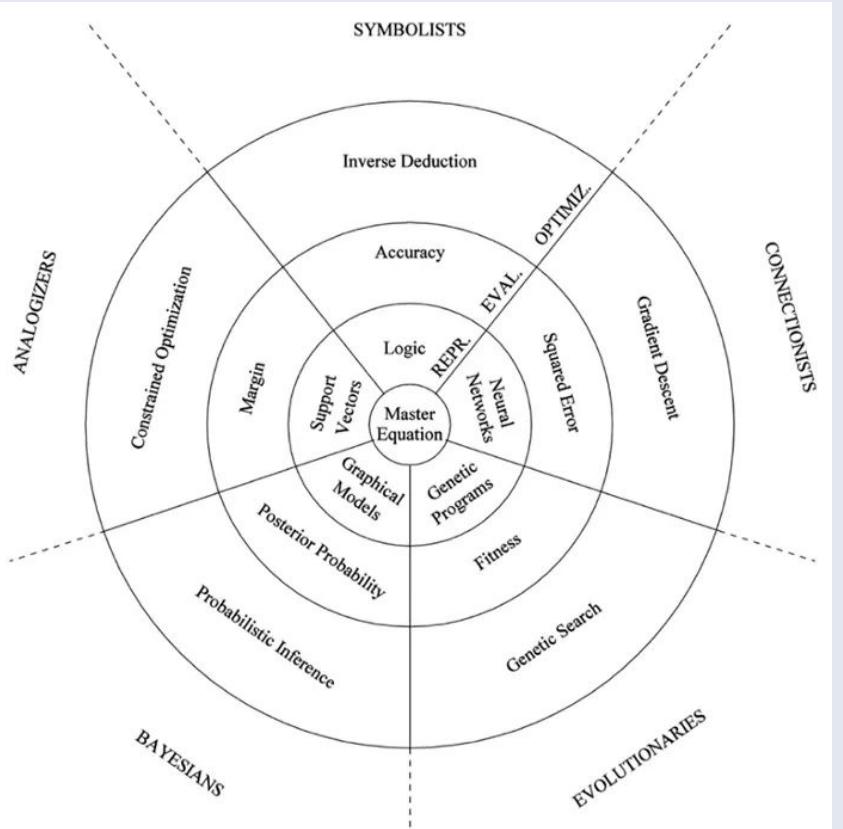
Deep Neural Networks (DNNs)

Convolutional Neural Networks (CNNs)

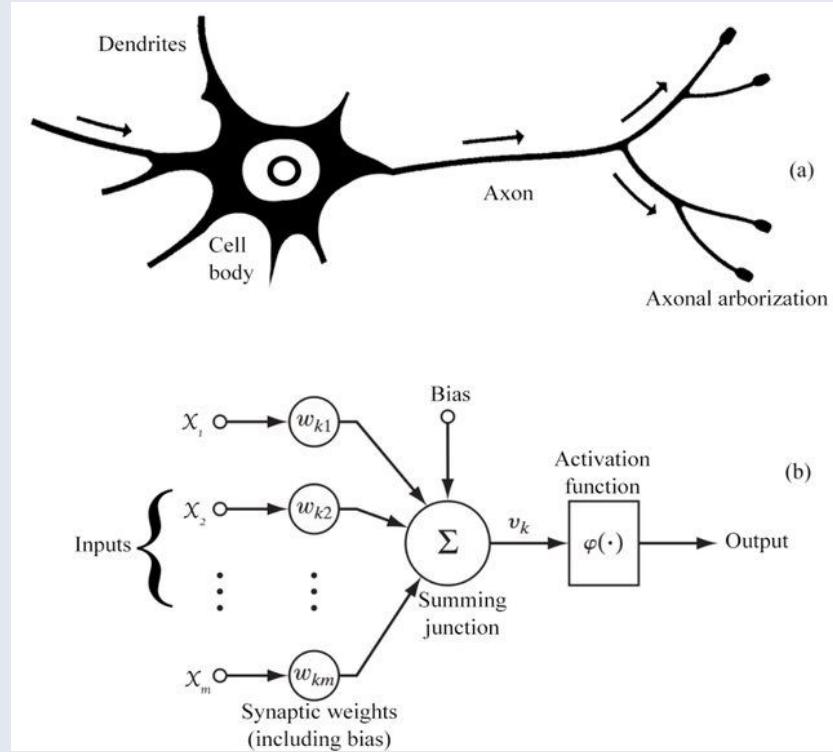
Recurrent Neural Networks (RNNs)

etc ...

Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"



Warren McCulloch and Walter Pitts, 1943

Akgün, Ergün, and Metin Demir. "Modeling course achievements of elementary education teacher candidates with artificial neural networks." *International Journal of Assessment Tools in Education* 5.3 (2018): 491-509.

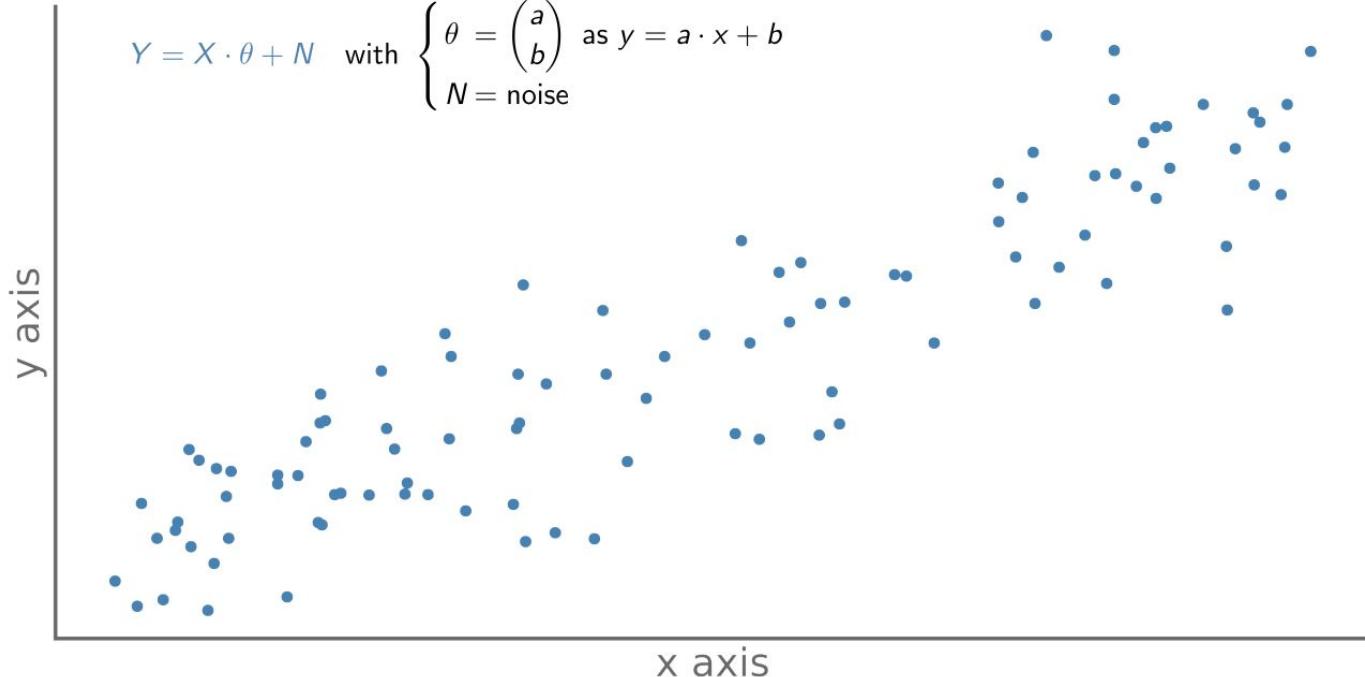
How to **learn** ?

- Optimizing a **objective function** (usually, a **cost** function)
 - Gradient descent

Linear regression

We have a phenomenon, for which we have observations

$$Y = X \cdot \theta + N \quad \text{with} \quad \begin{cases} \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ as } y = a \cdot x + b \\ N = \text{noise} \end{cases}$$

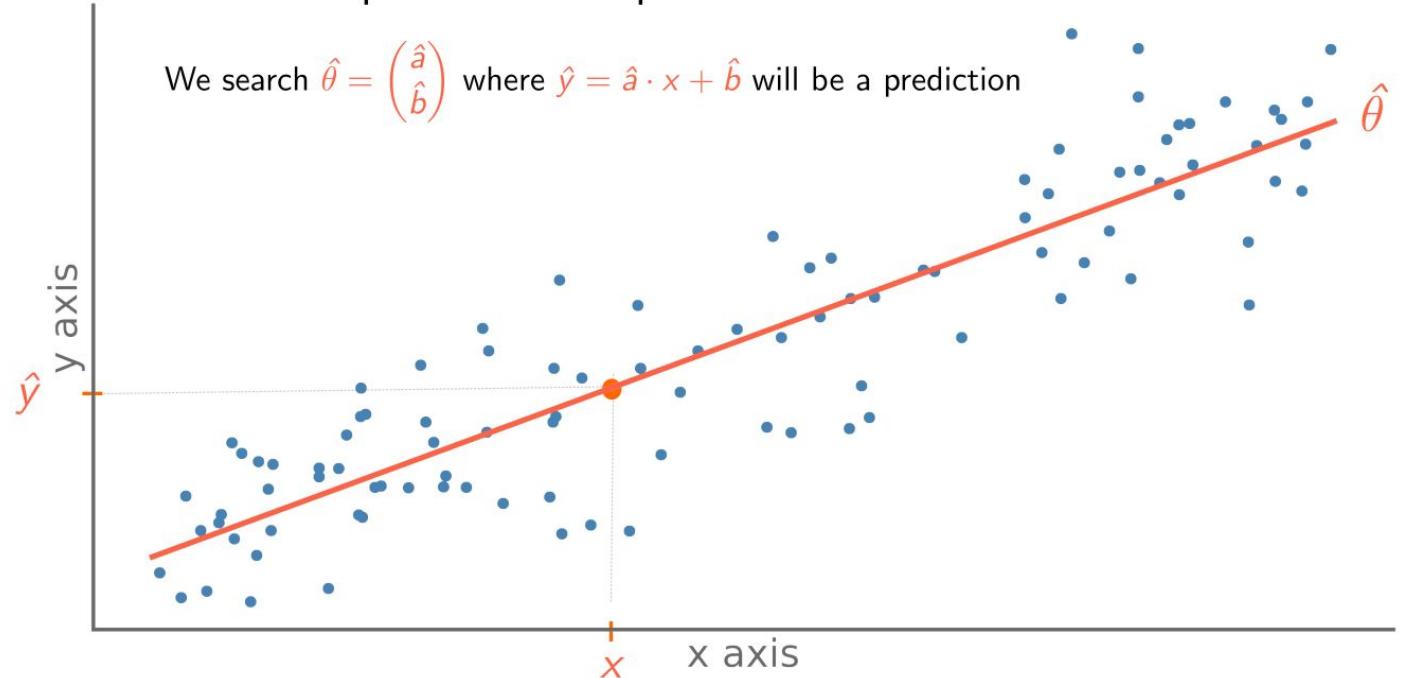




Linear regression

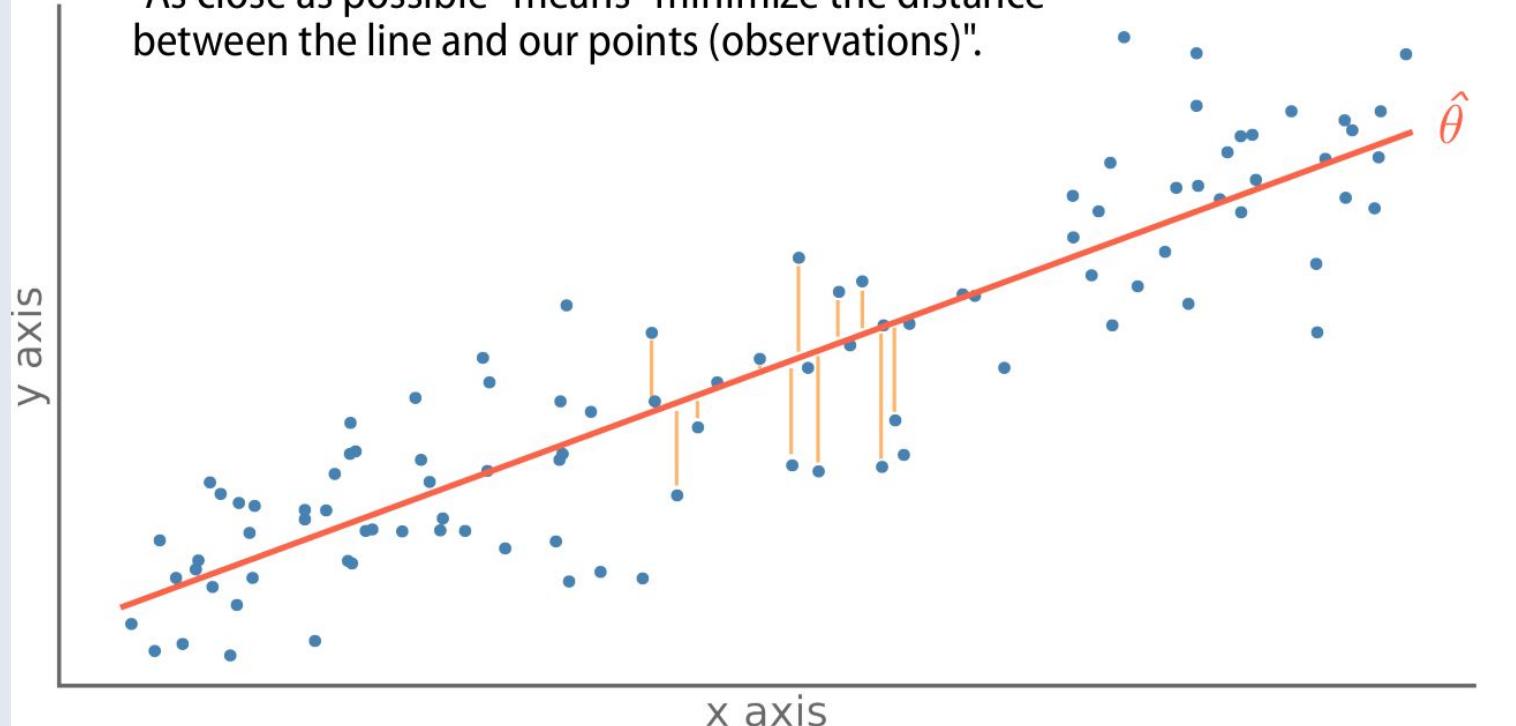
We are looking for a straight line that passes « as close as possible » to our points.

We search $\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ where $\hat{y} = \hat{a} \cdot x + \hat{b}$ will be a prediction



Linear regression

"As close as possible" means "minimize the distance between the line and our points (observations)".

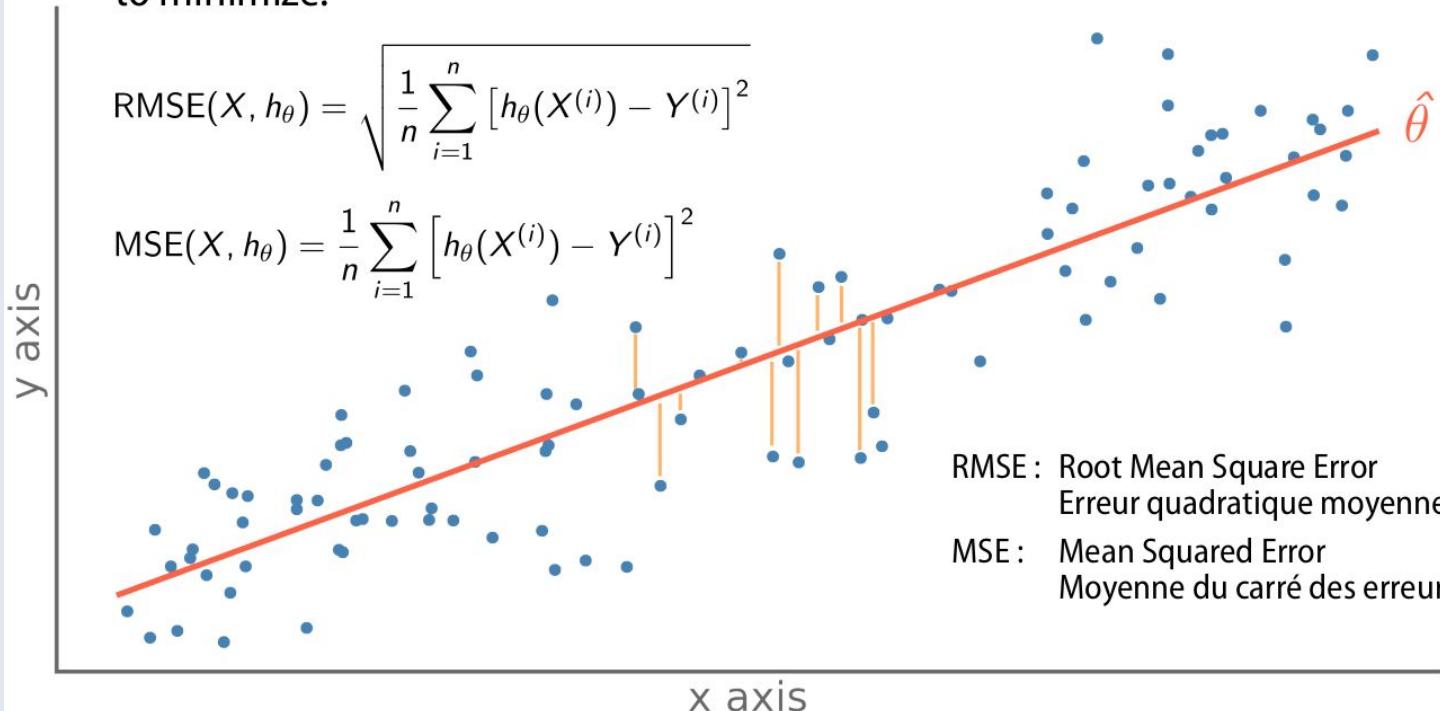


Linear regression

For this, we will use an «loss function », which we will try to minimize.

$$\text{RMSE}(X, h_\theta) = \sqrt{\frac{1}{n} \sum_{i=1}^n [h_\theta(X^{(i)}) - Y^{(i)}]^2}$$

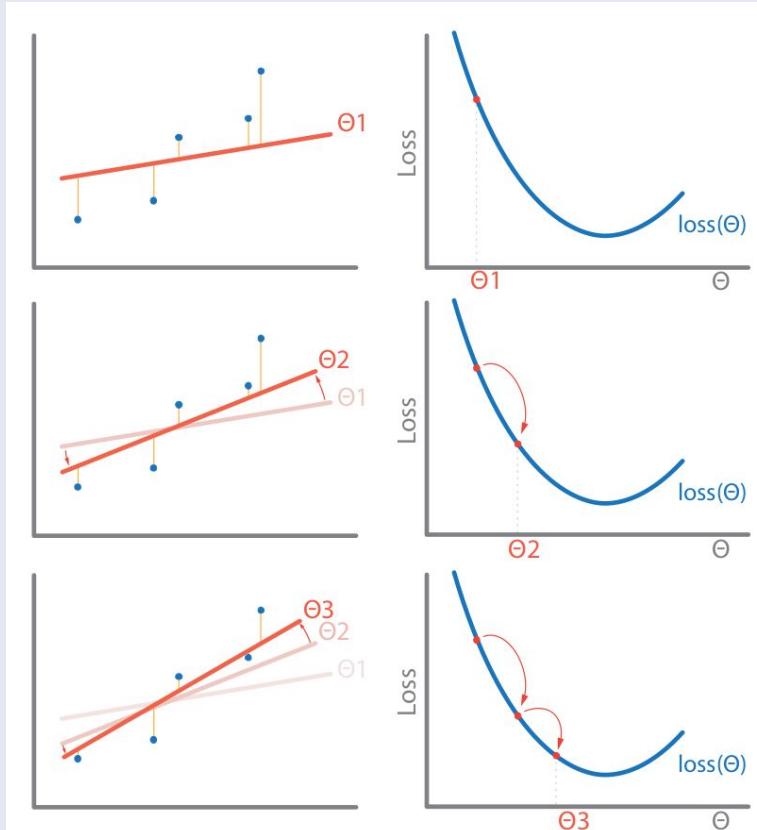
$$\text{MSE}(X, h_\theta) = \frac{1}{n} \sum_{i=1}^n [h_\theta(X^{(i)}) - Y^{(i)}]^2$$



RMSE : Root Mean Square Error
 Erreur quadratique moyenne
 MSE : Mean Squared Error
 Moyenne du carré des erreurs



Gradient descent



We will iteratively look for the best position of our line, by varying its parameters (Θ).



But how can we efficiently vary our parameters (Θ)?

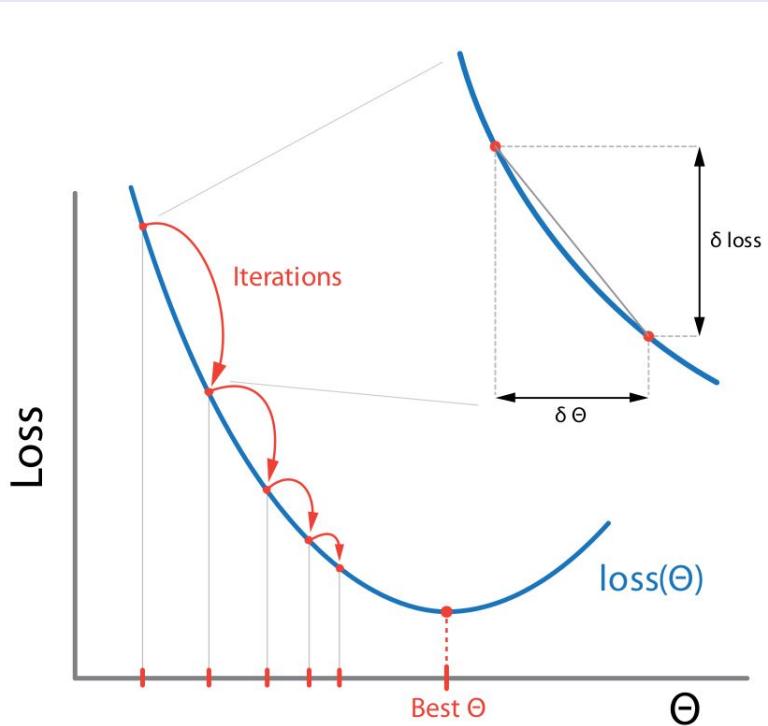
Note :

Loss functions could be :

$$RMSE(X, h_{\theta}) = \sqrt{\frac{1}{n} \sum_{i=1}^n [h_{\theta}(X^{(i)}) - Y^{(i)}]^2}$$

$$MSE(X, h_{\theta}) = \frac{1}{n} \sum_{i=1}^n [h_{\theta}(X^{(i)}) - Y^{(i)}]^2$$

Gradient descent



By changing Θ from $\delta\Theta$
 We improve $\text{loss}(\Theta)$ of δloss

The gradient is the slope we will follow
 to minimize our loss function.

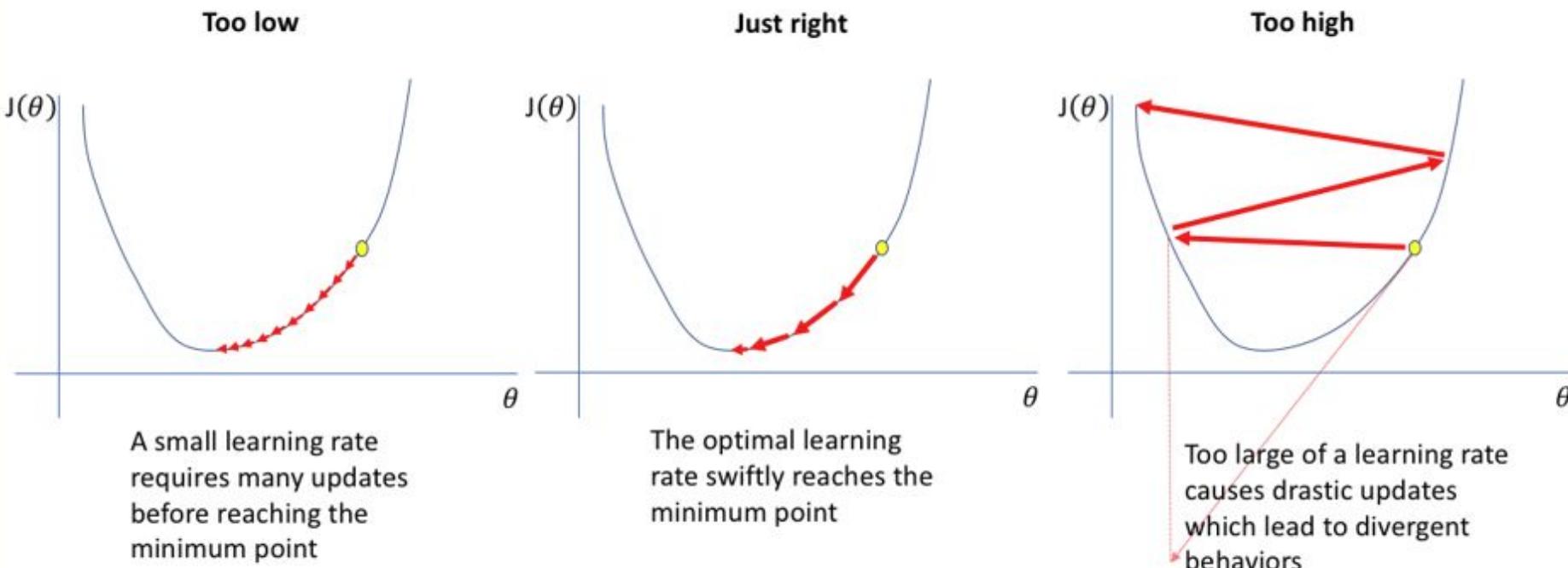
$$\text{gradient} = \frac{\delta \text{loss}}{\delta \theta}$$

One iterative solution is : $\theta \leftarrow \theta - \eta \cdot \frac{\delta \text{loss}}{\delta \theta}$

where η is the learning rate

This process is called **gradient descent** and
 the function used to optimize the descent,
optimization function

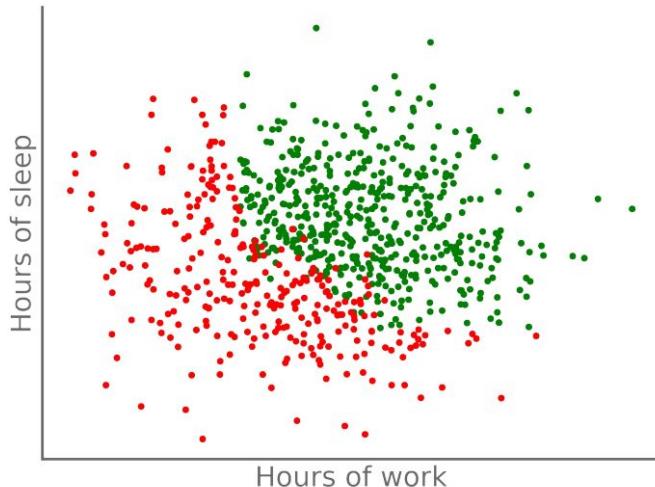
Gradient descent



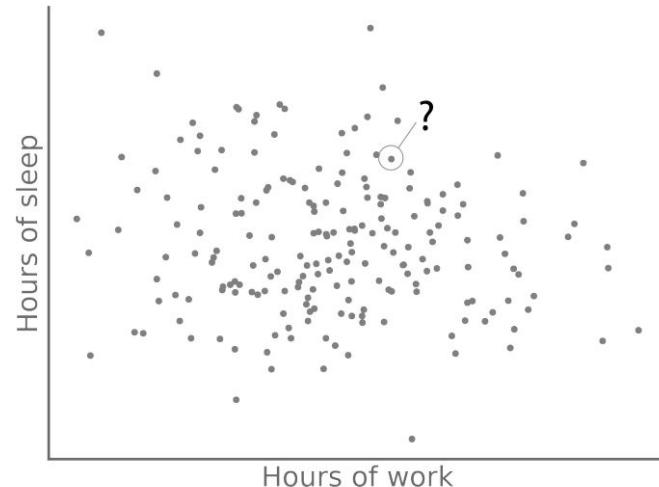
Logistic regression

A logistic regression is intended to provide a probability of belonging to a class.

Dataset : X Observations
y Classe

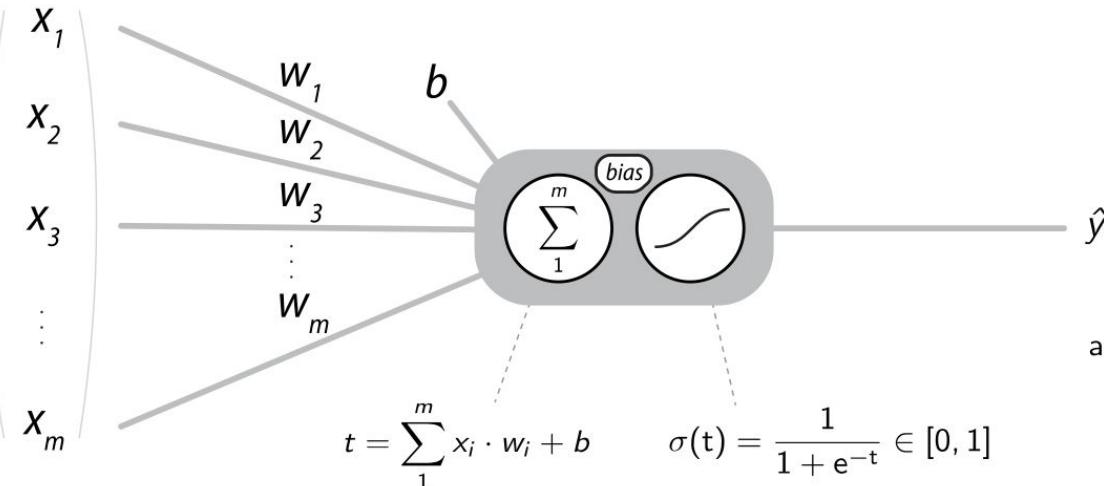


Objective : Predict the class
x given, we want to predict y
 $y_{\text{pred}} = f(x)$
where f is a linear function



Logistic regression

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



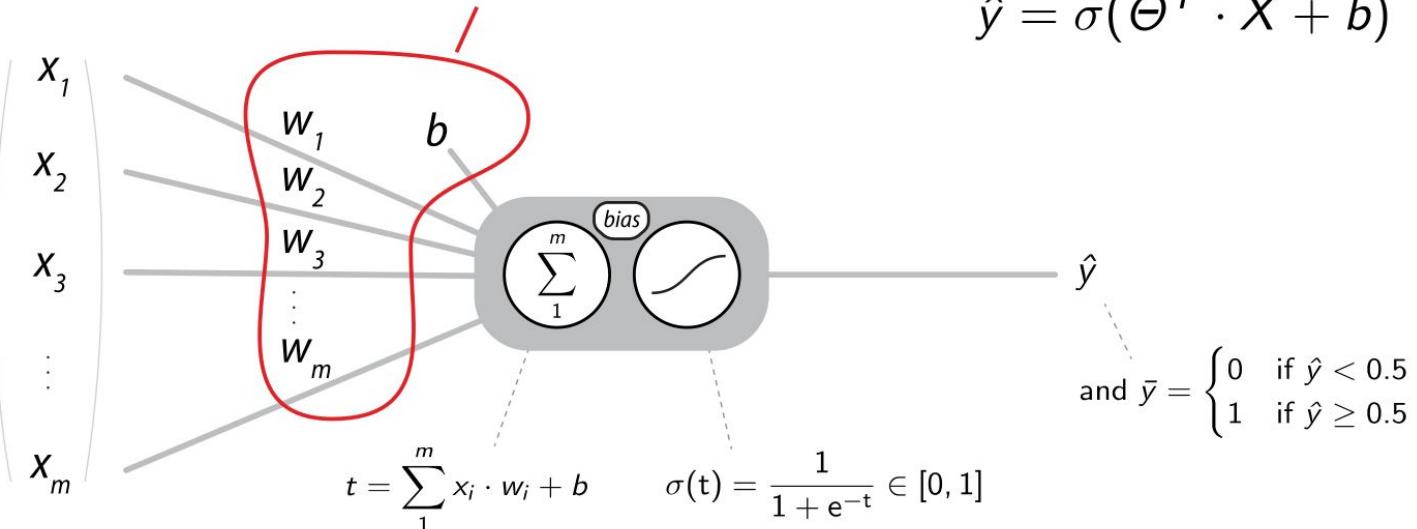
and $\bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \\ 1 & \text{if } \hat{y} \geq 0.5 \end{cases}$

Input	Bias / Weight	Activation function	Output
X	Θ	$\sigma(t)$	\hat{y}

Logistic regression

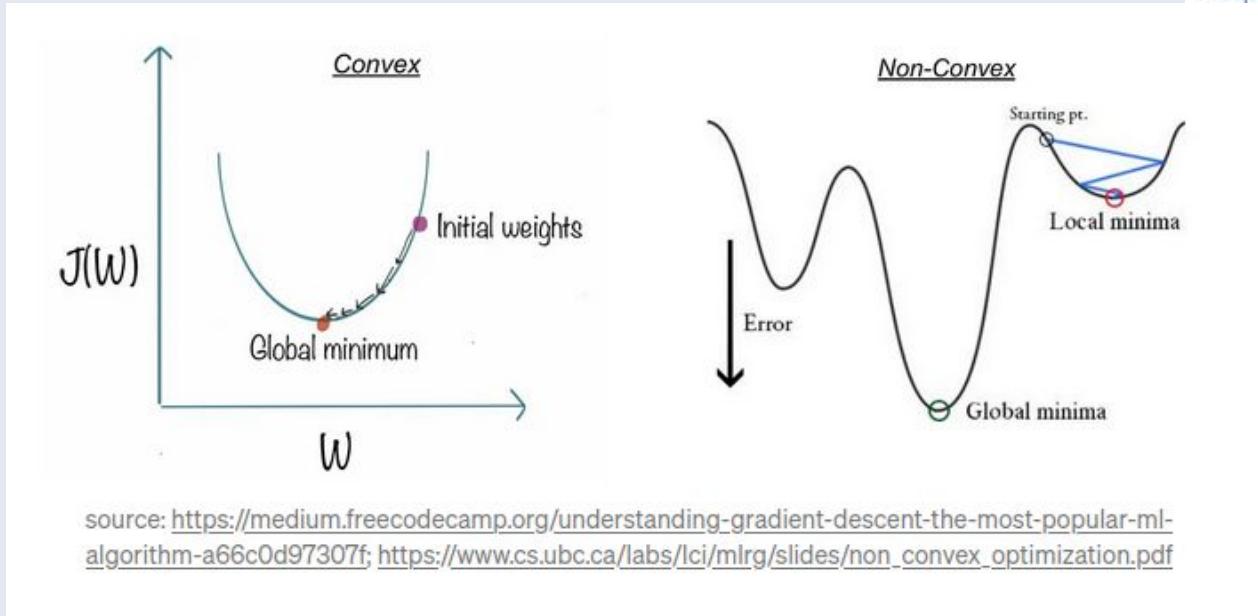
Determined by the minimisation
of a cost function $J(\Theta)$

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$

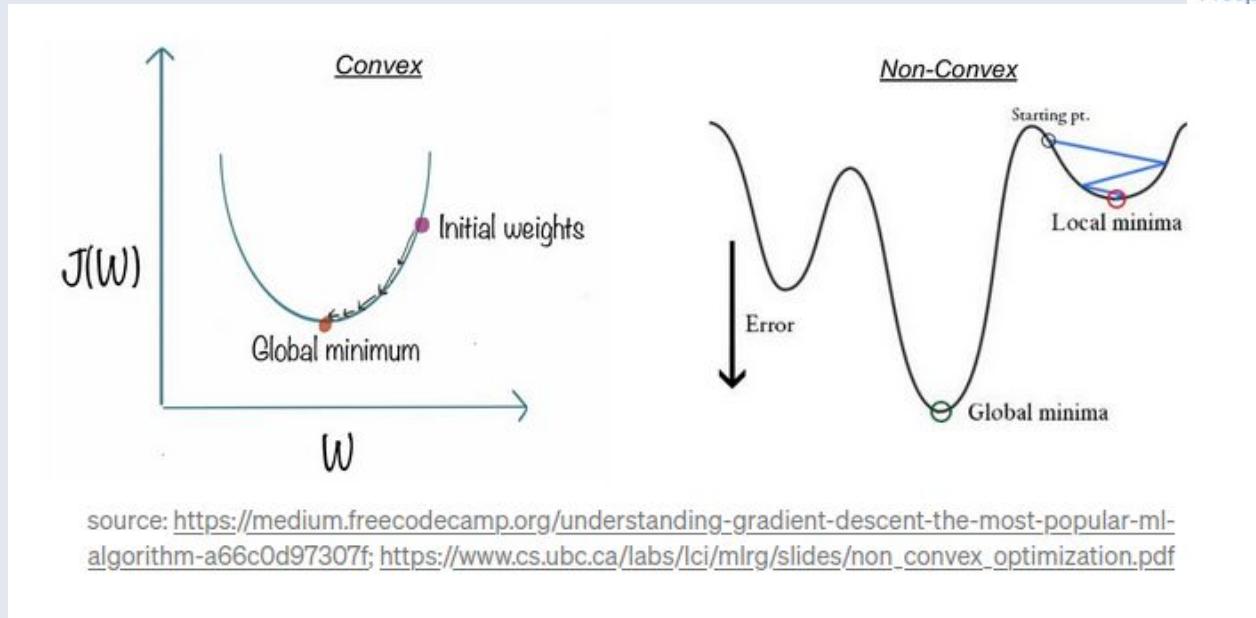


Input	Bias / Weight	Activation function	Output
X	Θ	$\sigma(t)$	\hat{y}

Logistic regression



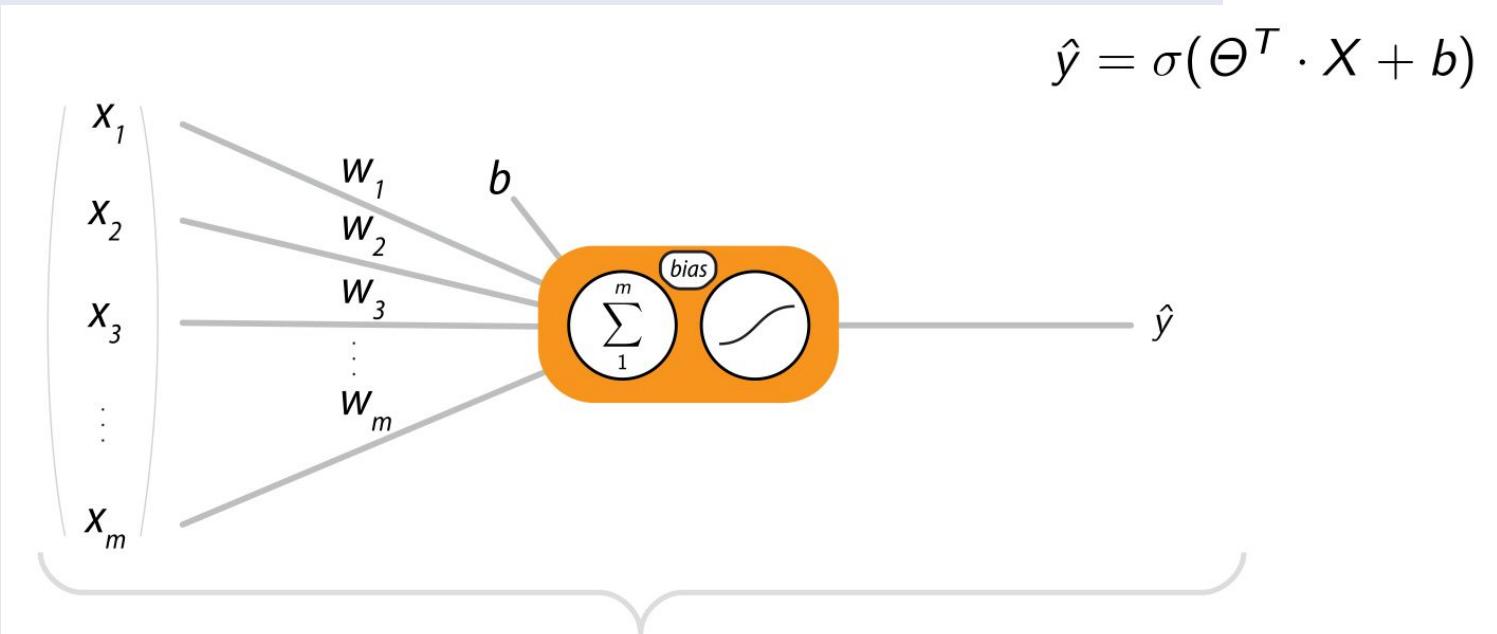
Logistic regression



$$\text{Cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1 - y) \log(1 - h_\theta(x))$$

Logistic regression

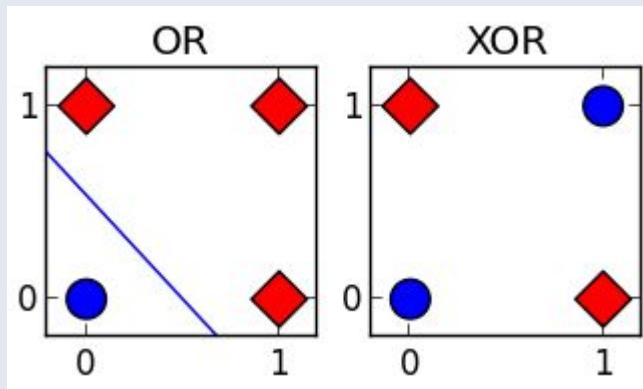
$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



That's an « **artificial neuron** » !

So, we have a neural network of... 1 neuron !

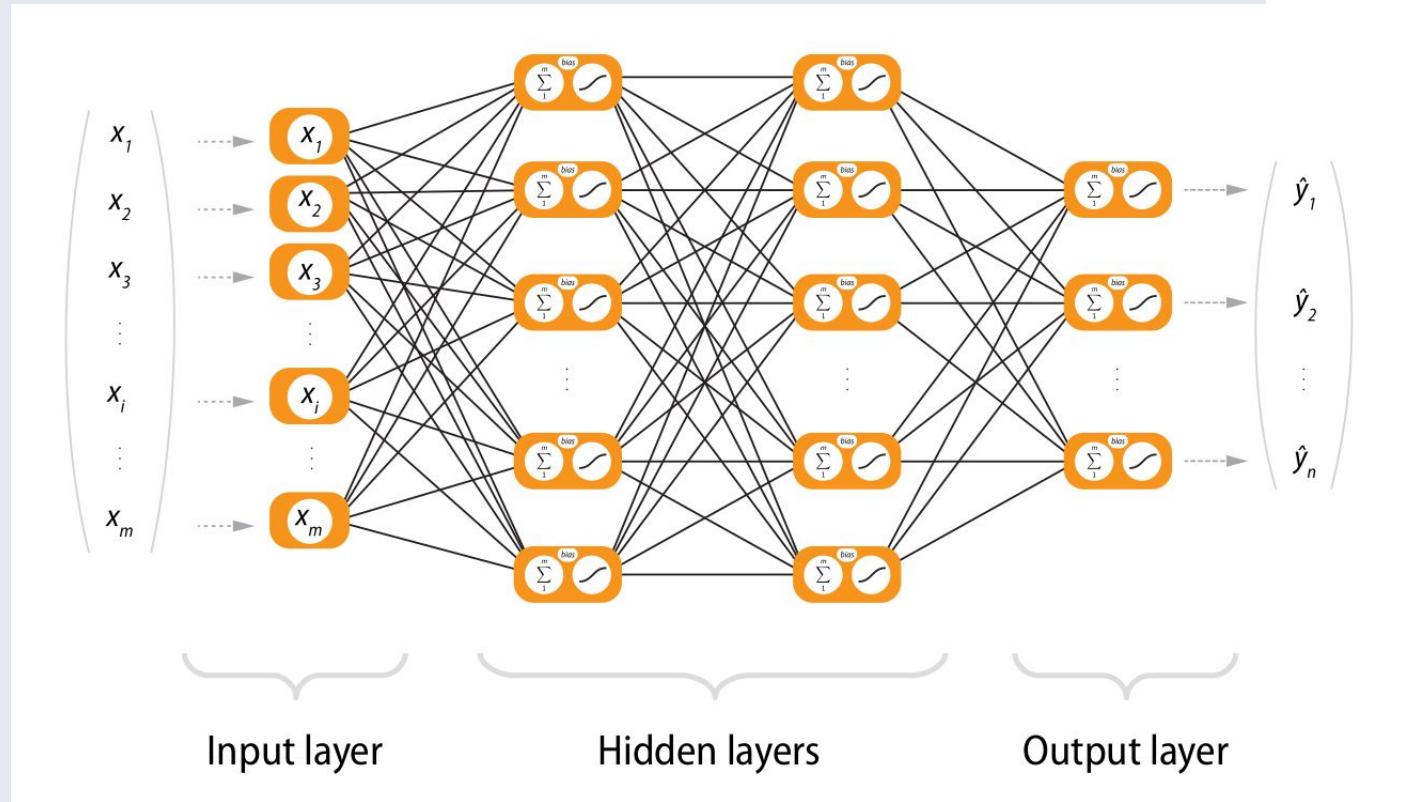
Limits of the perceptron



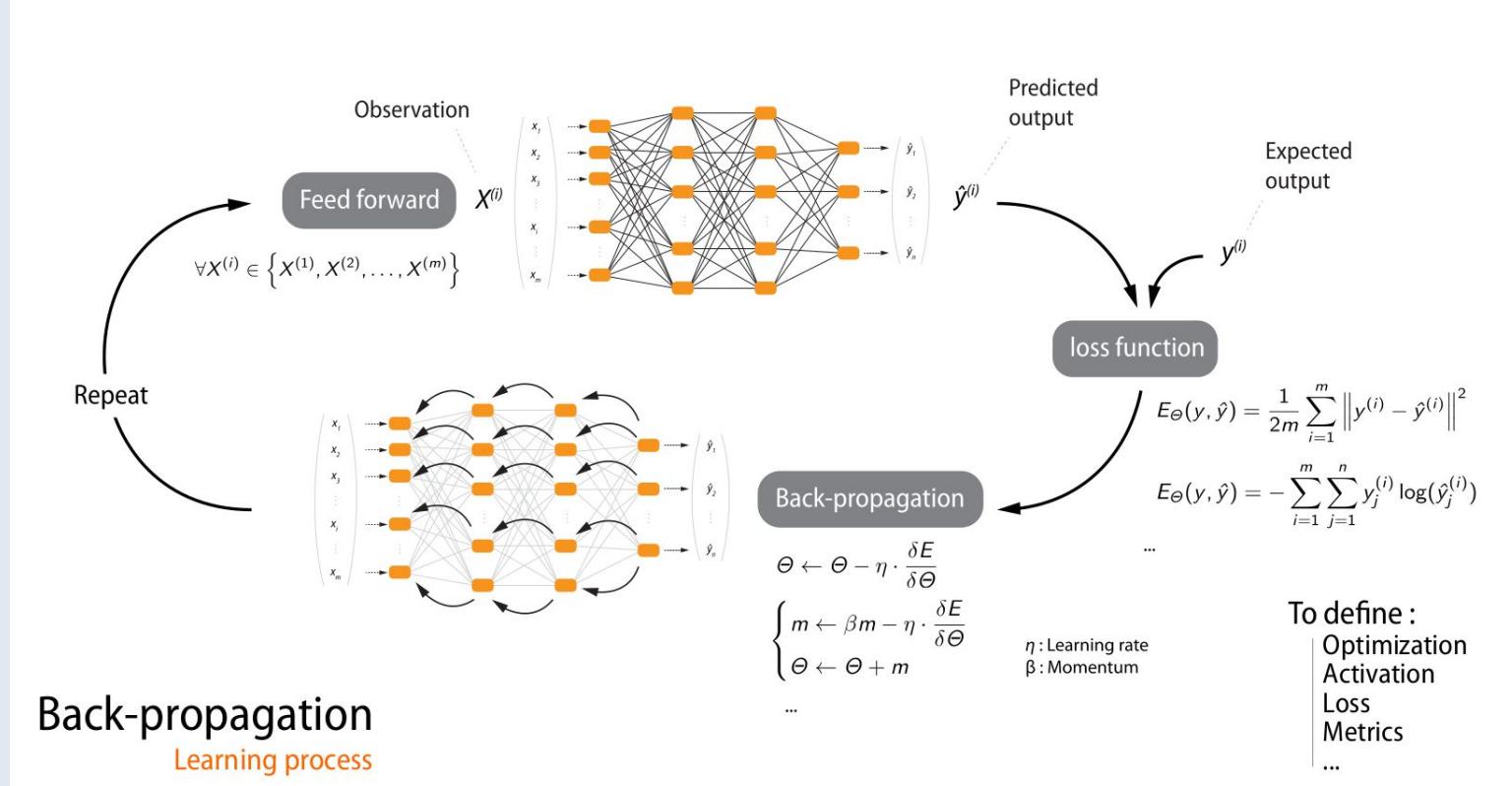
A	B	$A \oplus B$
False	False	False
False	True	True
True	False	True
True	True	False

=> Artificial neural networks

Neural networks

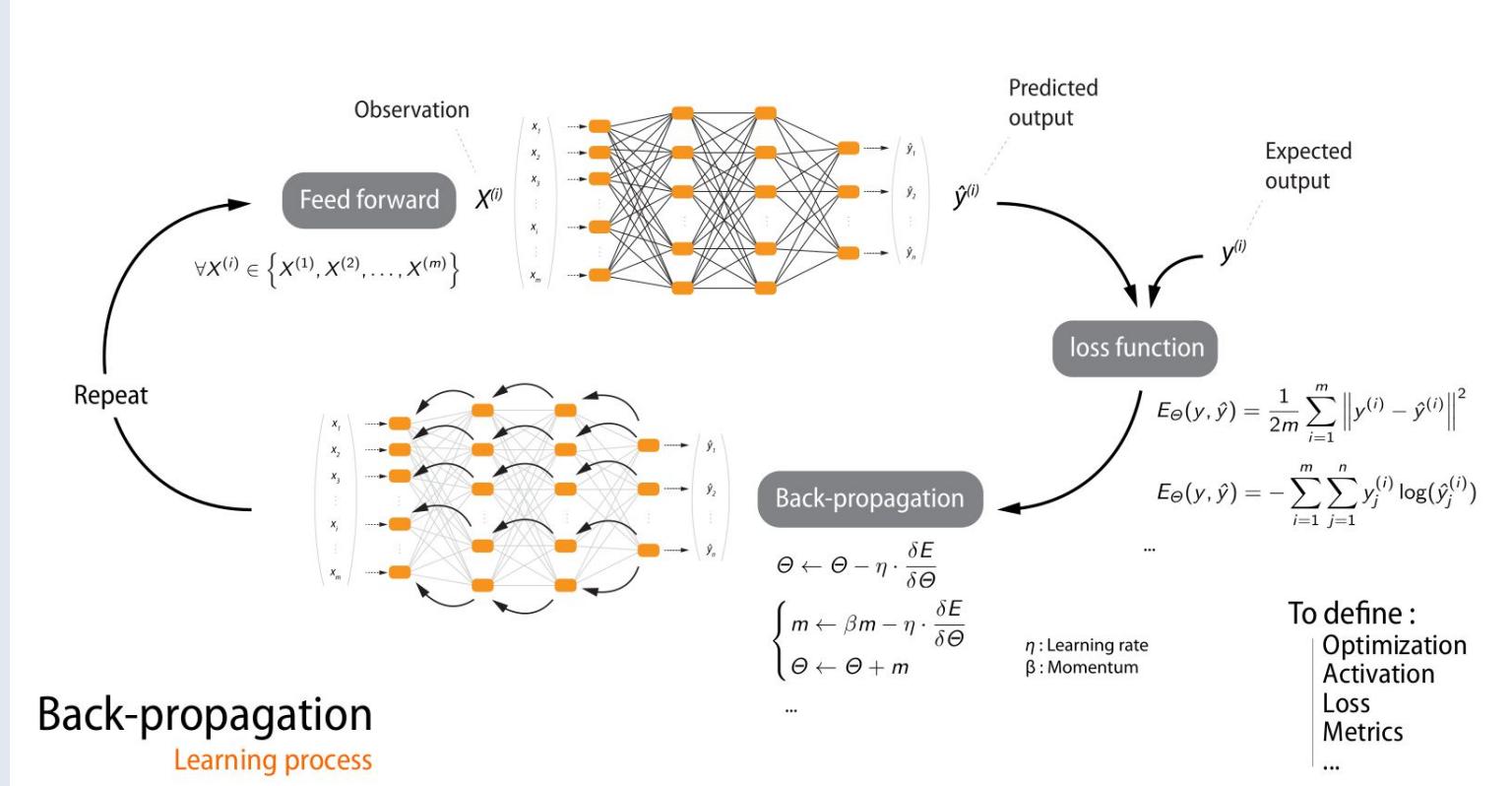


Back-propagation

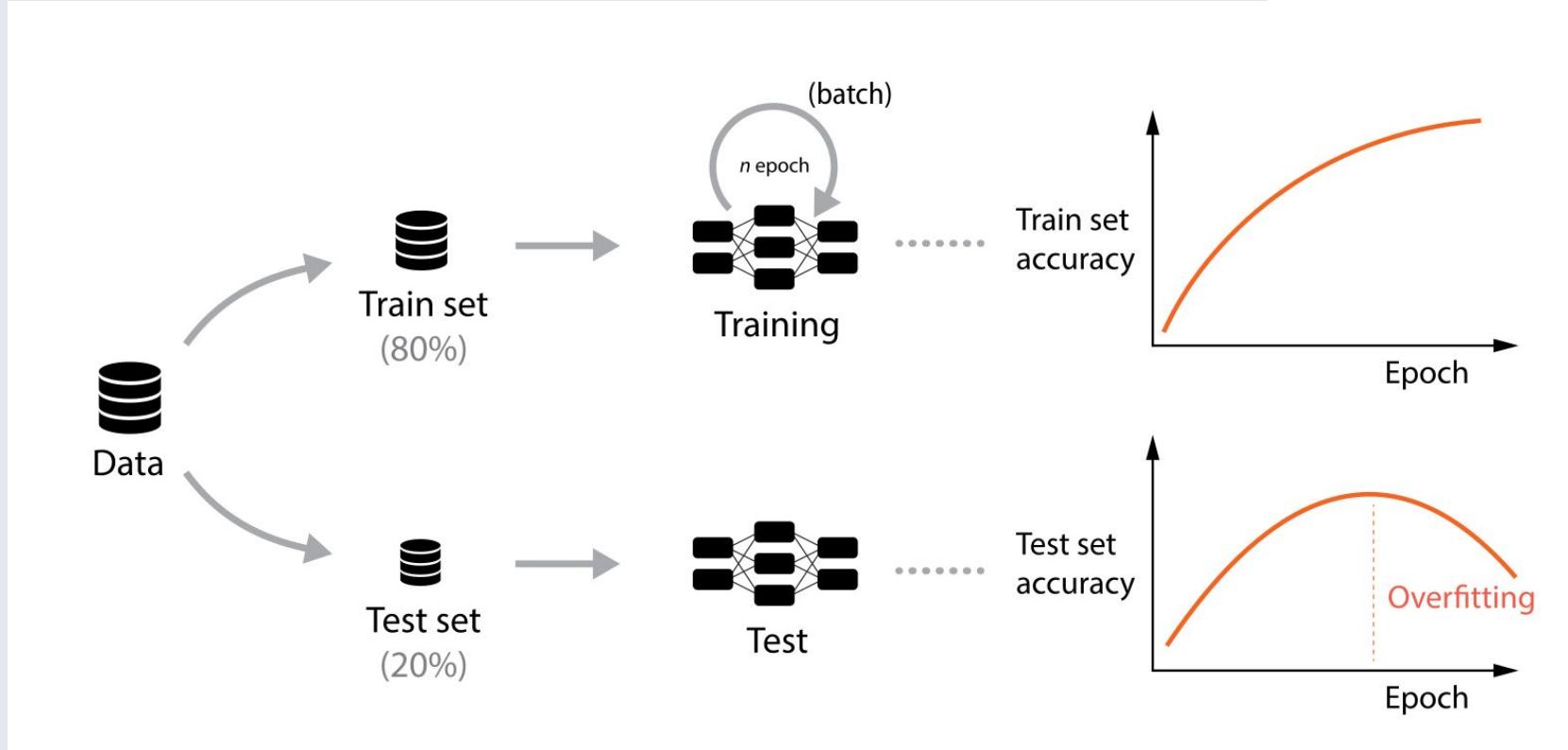


Back-propagation
Learning process

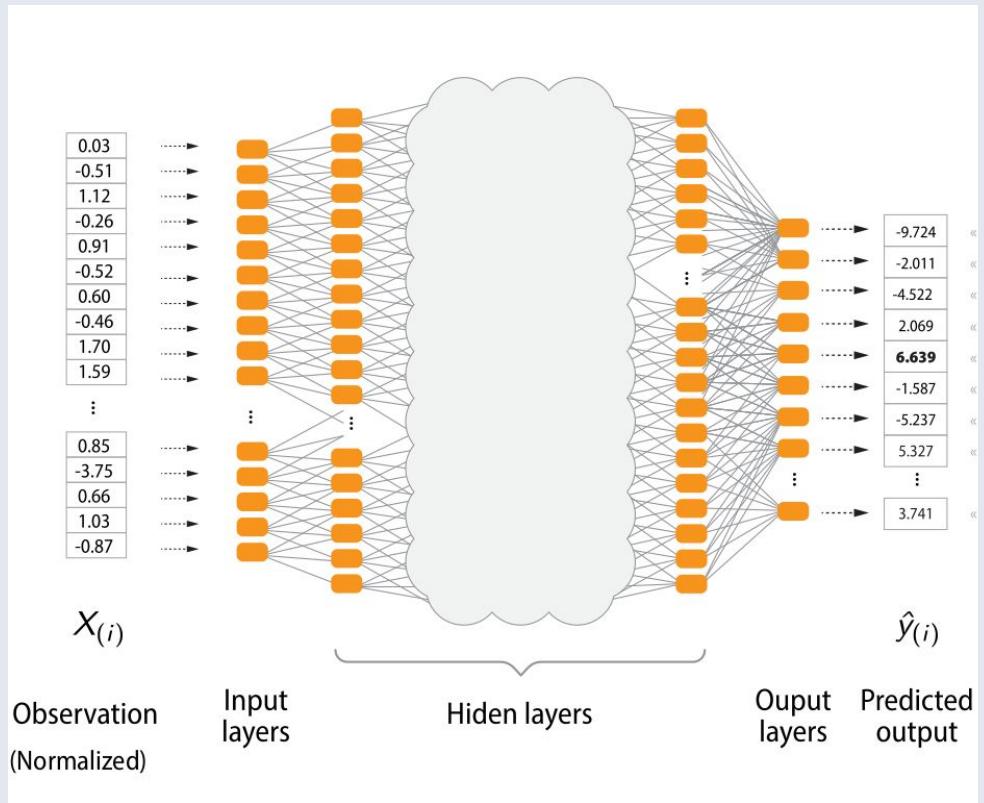
Back-propagation



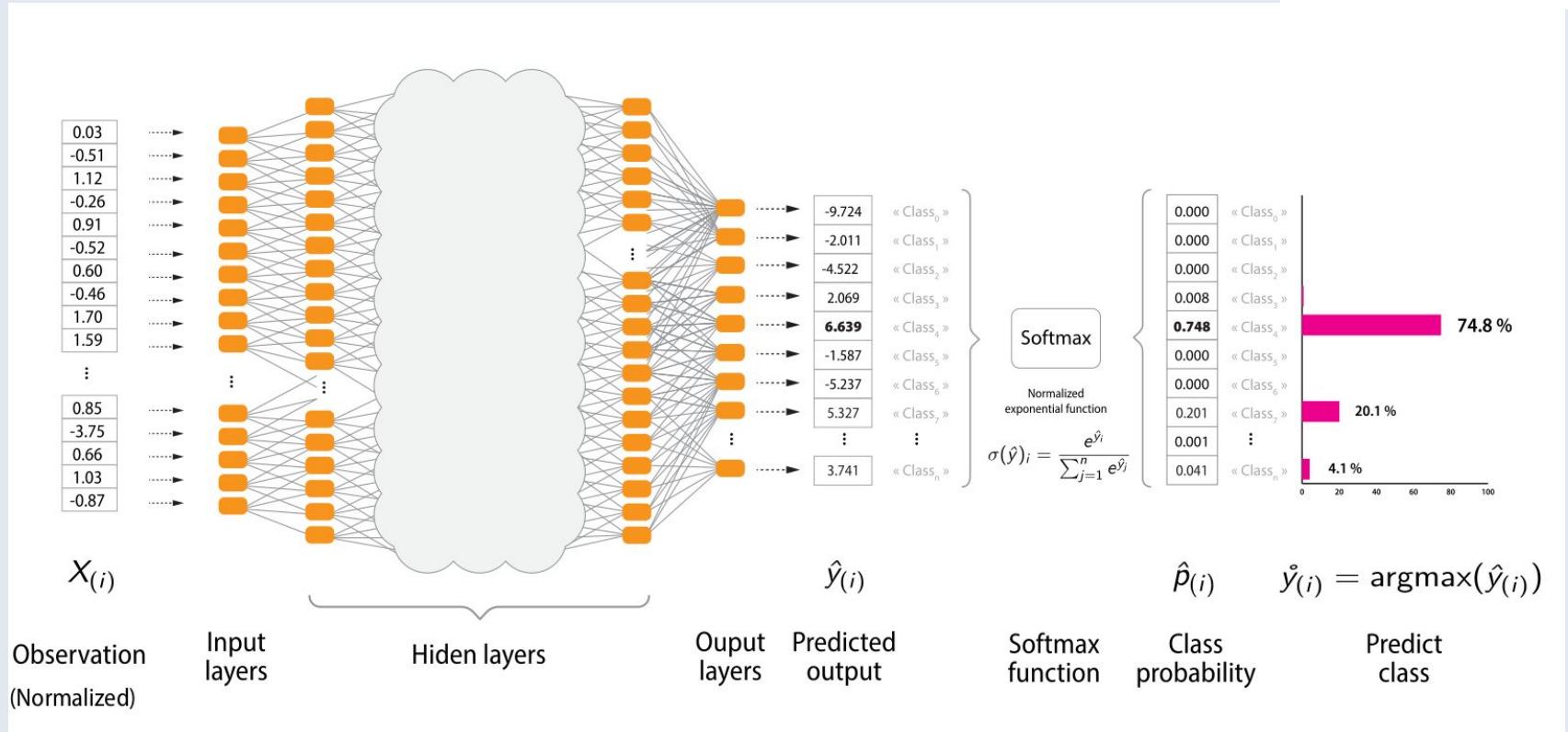
Training process



Regression VS classification



Regression VS classification



Convolutional NN

For a fully connected layer of (only)
±1000 neurons, we would need to



0.0008 M pixels
28x28, 8 bits

→ 785.000 params



24 M pixels
(r,v,b) 3x8 bits

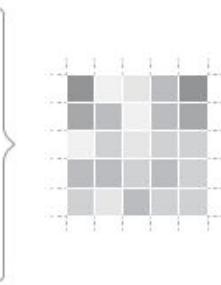
→ 72. 10E9 params...



Convolution ?



By Jan Kroon, from Pexels.com



5	2	1	3	5
4	3	2	3	4
0	2	1	2	2
3	3	2	3	2
2	1	3	2	2

Image piece				
5	2	1		
4	3	2		
0	2	1		
3	3	2		
2	1	3		

X

Image piece

Kernel 3x3

1	0	1
0	1	0
1	0	1

w

=	10
---	----

y

$$y = 5 \cdot 1 + 4 \cdot 0 + 0 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 + 2 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 + 1 \cdot 1 = 10$$

$$y = \sum_{i=1}^n \sum_{j=1}^m x_{i,j} \cdot w_{i,j} \quad \text{with } \begin{cases} n & \text{kernel width} \\ m & \text{kernel height} \end{cases}$$

⊗ Hadamard product

2D convolution



Convolution ?



By Jan Kroon, from Pexels.com

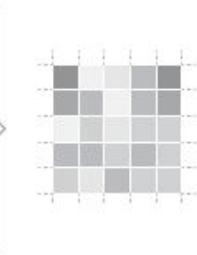


Image piece				
5	2	1	3	5
4	3	2	3	4
0	2	1	2	2
3	3	2	3	2

Kernel 3x3		
1	0	1
0	1	0
1	0	1

$$= \boxed{10}$$

y

Image piece				
5	2	1	3	5
4	3	2	3	4
0	2	1	2	2
3	3	2	3	2

Kernel 3x3		
1	0	1
0	1	0
1	0	1

$$= \boxed{10 \ 11}$$

y

Image piece				
5	2	1	3	5
4	3	2	3	4
0	2	1	2	2
3	3	2	3	2

Kernel 3x3		
1	0	1
0	1	0
1	0	1

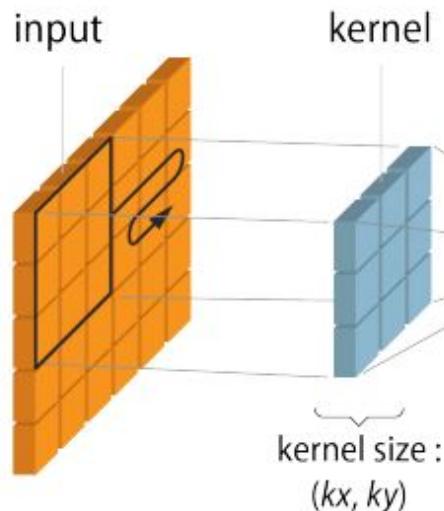
$$= \boxed{10 \ 11 \ 12}$$

y



We can perform convolutions in 1, 2, 3 ...or n-dimensional spaces !

Convolutional layer



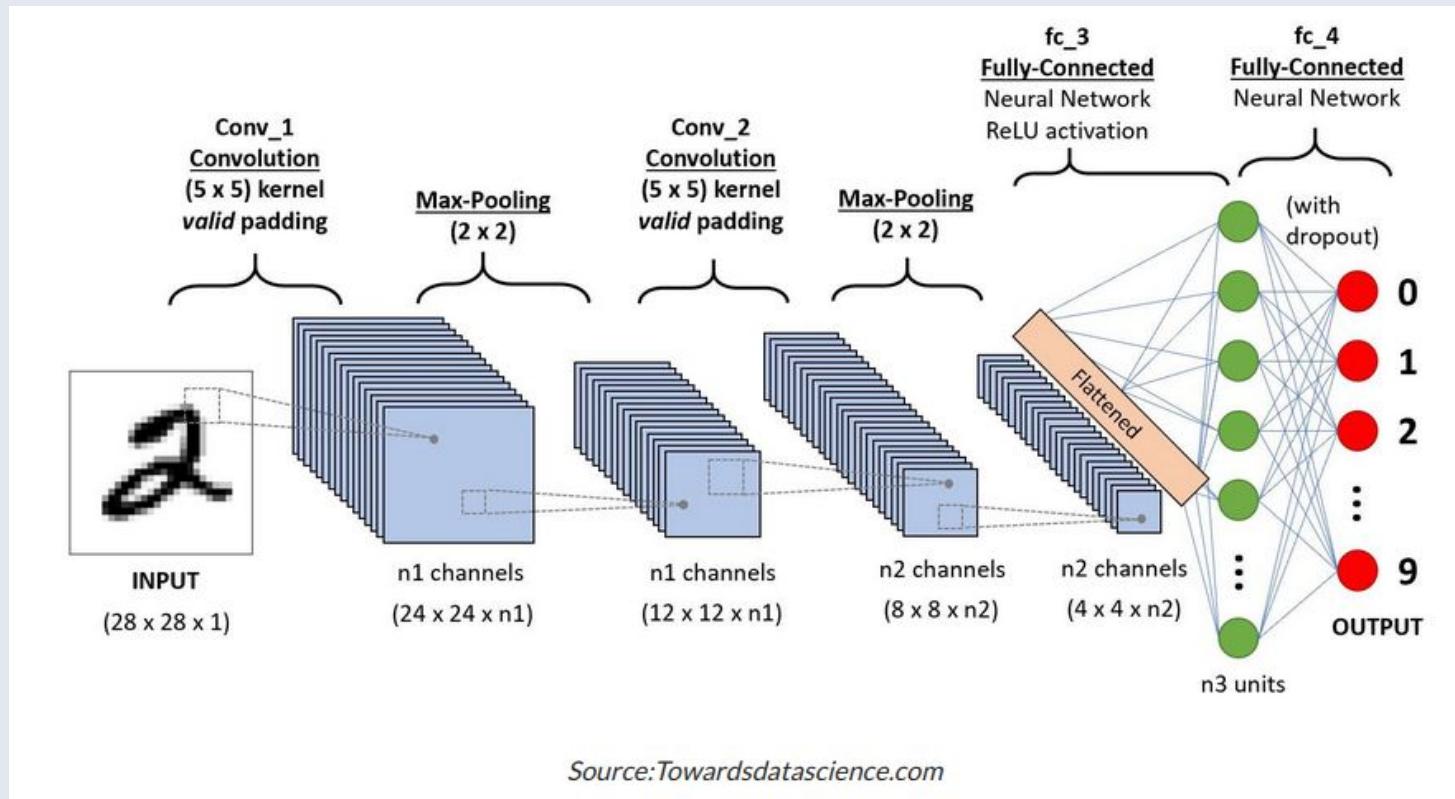
$$y = \sigma(\sum \text{input} \otimes \text{kernel} + b)$$

Where :

σ : is an activation function
 b : is a bias

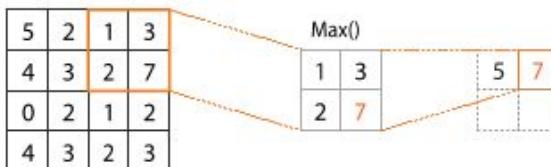
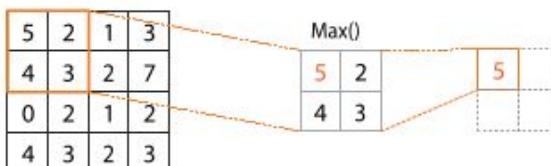
Number of parameters for a convolutional layer : $k_x \cdot k_y + 1$

Image classification



Pooling ?

Principle of Max Pooling :



It is possible to set the **window size**,
padding mode and **strides**.

By default, the strides correspond to
 the size of the window.

A window (2,2) generates an image
 twice as small.

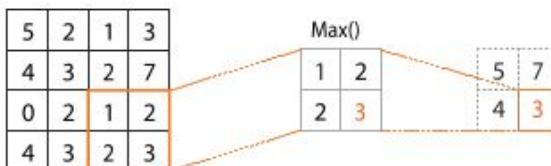
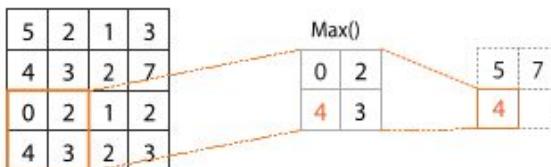
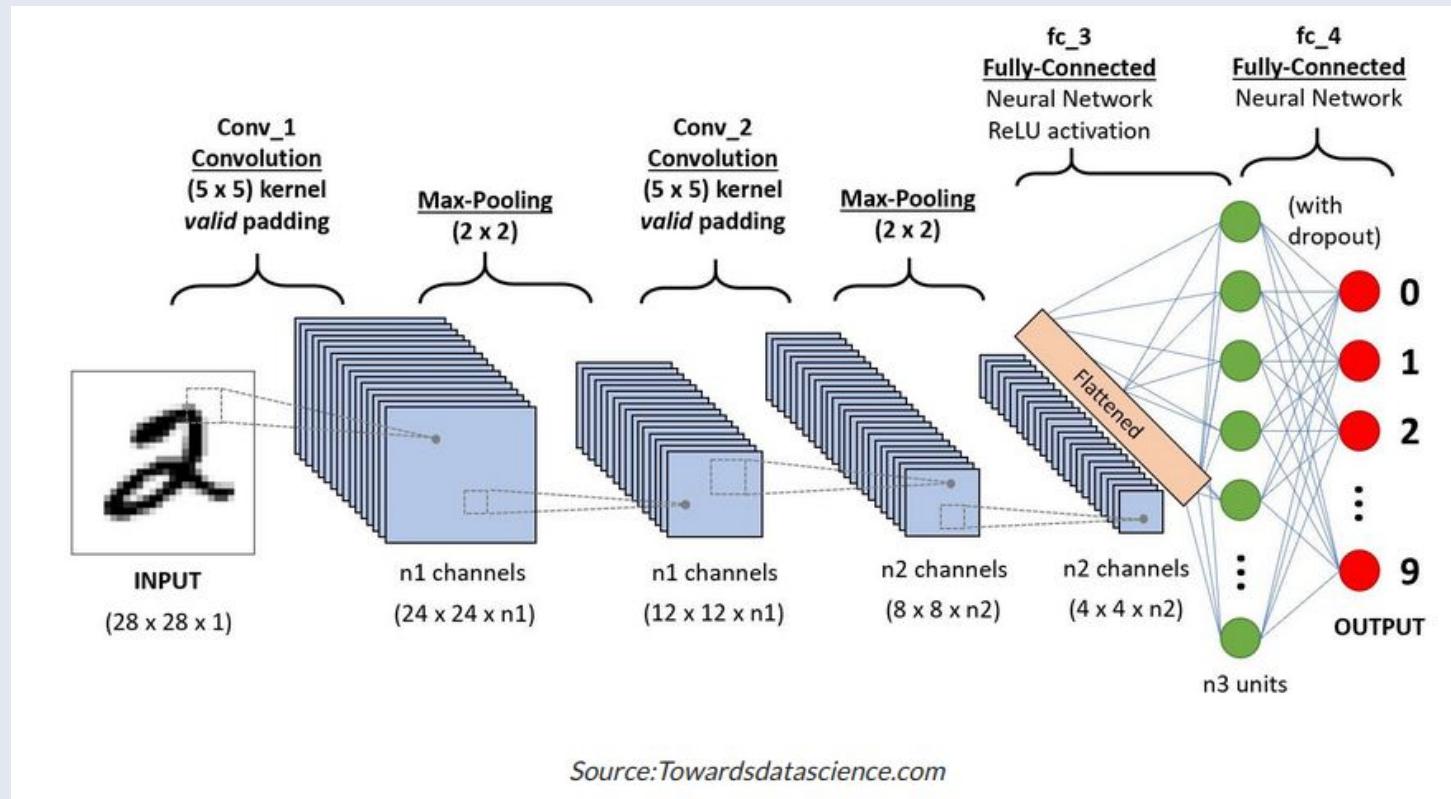
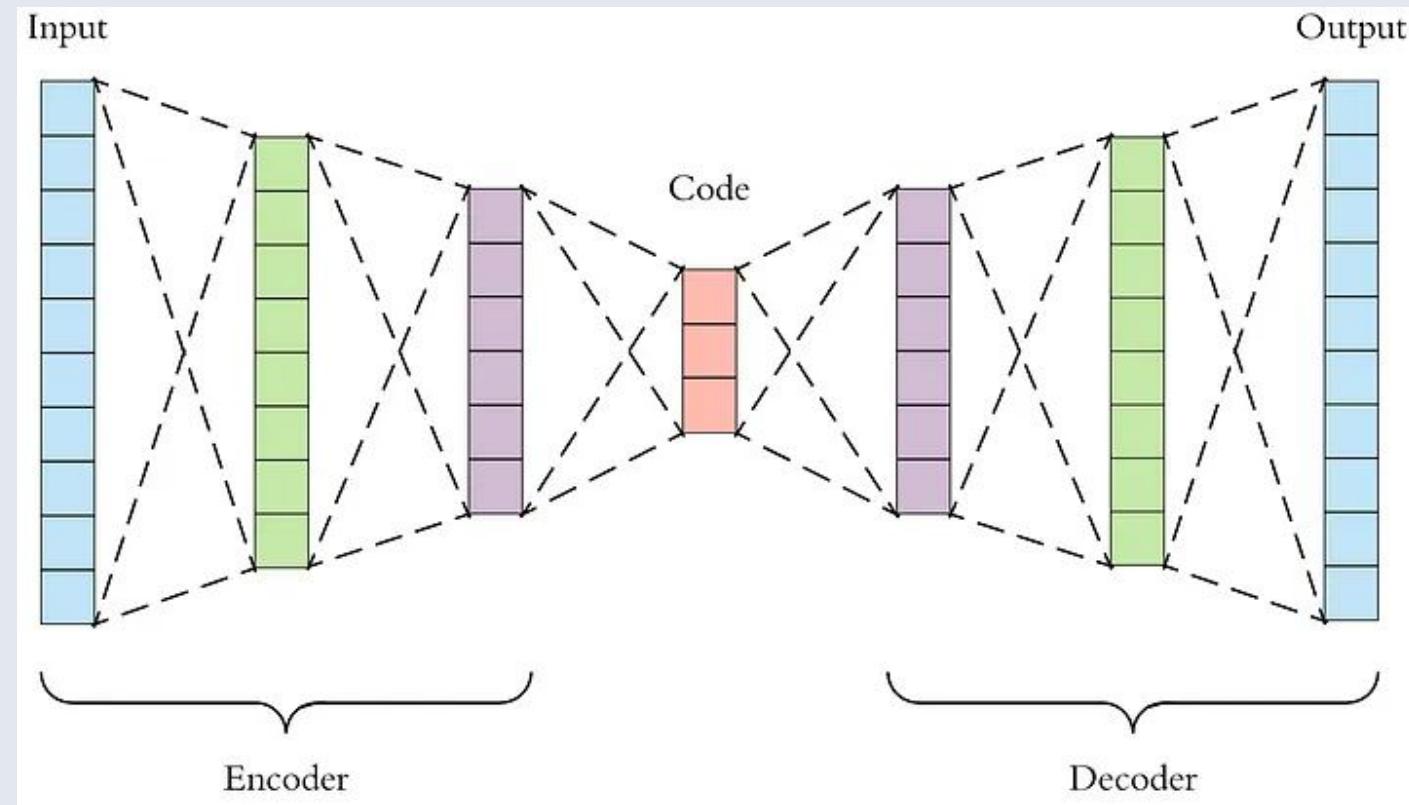


Image classification

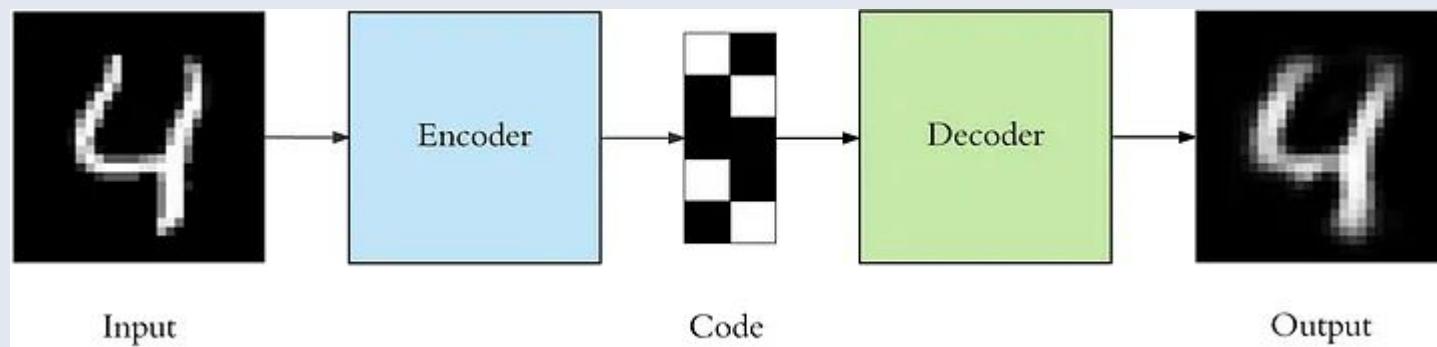


Source: Towardsdatascience.com

Auto-Encoders



Auto-Encoders



Looks like magic ?

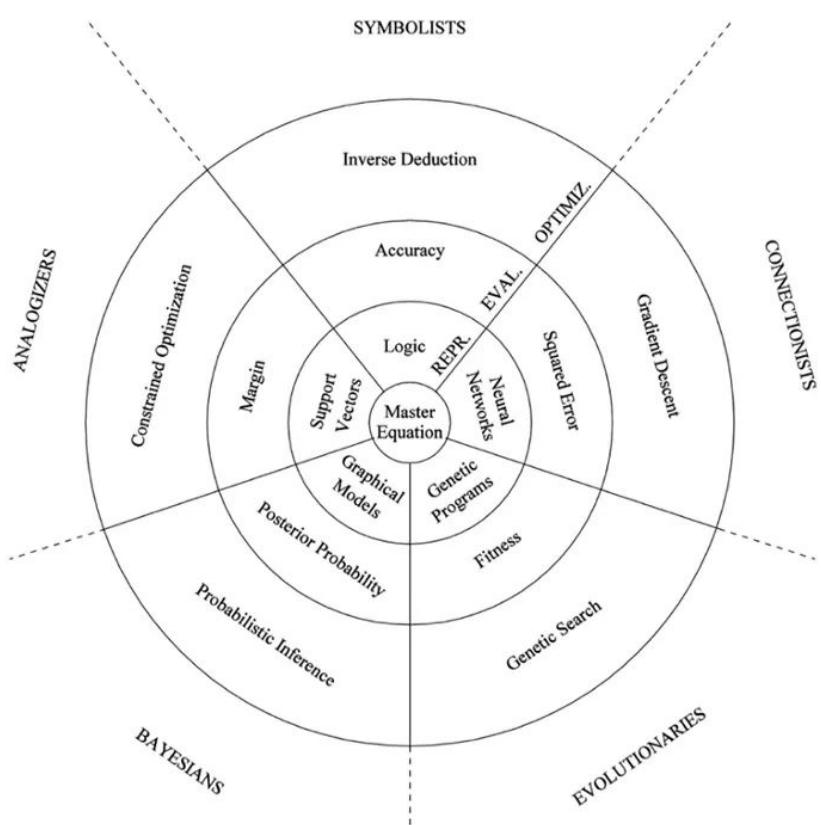
It kind of is... And the major downside of this is **explainability** !

+ other limitations

Need a lot of (clean) data for training

Need a lot of computing power

Deep learning is not the answer to everything...



Symbolists

Filling the gaps in our knowledge

Deduction

From rules to fact

Induction

From facts to rules

Socrate is human
+
Humans are mortal

Socrate is human
+
?

Inductive Logic Programming

Representing data as non-symbolists

	Features		Labels
Name	Job	Enjoys lego	Happy
alice	lego builder	yes	yes
bob	lego builder	no	no
claire	estate agent	yes	no
dave	estate agent	no	no

Table 1: A table representation of a ML task.

For neural networks: finding the appropriate weights (weight the importance of features)

Representing data as symbolists
Background Knowledge

$$B = \left\{ \begin{array}{l} \text{lego_builder(alice).} \\ \text{lego_builder(bob).} \\ \text{estate_agent(claire).} \\ \text{estate_agent(dave).} \\ \text{enjoys_lego(alice).} \\ \text{enjoys_lego(claire).} \end{array} \right\}$$

Positive and negative examples

$$E^+ = \{ \text{happy(alice).} \} \quad E^- = \left\{ \begin{array}{l} \text{happy(bob).} \\ \text{happy(claire).} \\ \text{happy(dave).} \end{array} \right\}$$

For ILP: finding the hypothesis that entails as many positive examples and as few negative examples as possible

Inductive Logic Programming

Background Knowledge

$$B = \left\{ \begin{array}{l} \text{lego_builder(alice).} \\ \text{lego_builder(bob).} \\ \text{estate_agent(claire).} \\ \text{estate_agent(dave).} \\ \text{enjoys_lego(alice).} \\ \text{enjoys_lego(claire).} \end{array} \right\}$$

For ILP: finding the hypothesis that entails as many positive examples and as few negative examples as possible

$$H = \left\{ \forall A. \text{lego_builder}(A) \wedge \text{enjoys_lego}(A) \rightarrow \text{happy}(A) \right\}$$

Induced Hypothesis

Positive and negative examples

$$E^+ = \{ \text{happy(alice).} \} E^- = \left\{ \begin{array}{l} \text{happy(bob).} \\ \text{happy(claire).} \\ \text{happy(dave).} \end{array} \right\}$$

Cropper, Andrew, and Sebastijan Dumančić. "Inductive logic programming at 30: a new introduction." *Journal of Artificial Intelligence Research* 74 (2022): 765-850.

Inductive Logic Programming

Positive examples

Input	Output	
machine	e	$E^+ = \left\{ \begin{array}{l} \text{last}([m,a,c,h,i,n,e], e). \\ \text{last}([l,e,a,r,n,i,n,g], g). \\ \text{last}([a,l,g,o,r,i,t,m], m). \end{array} \right\}$
learning	g	
algorithm	m	

Background Knowledge

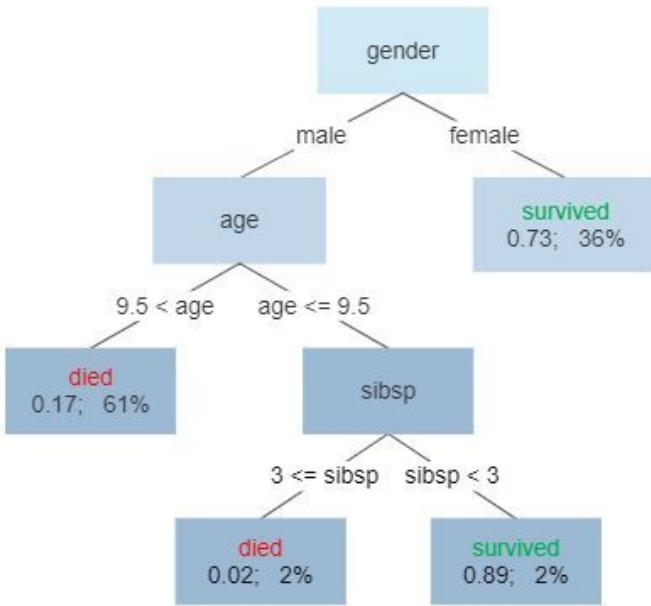
Name	Description	Example
empty(A)	A is an empty list	empty([]).
head(A,B)	B is the head of the list A	head([c,a,t],c).
tail(A,B)	B is the tail of the list A	tail([c,a,t],[a,t]).

$$H = \left\{ \begin{array}{l} \text{last}(A,B) :- \text{head}(A,B), \text{tail}(A,C), \text{empty}(C). \\ \text{last}(A,B) :- \text{tail}(A,C), \text{last}(C,B). \end{array} \right\}$$

Induced Hypothesis

Decision Tree Learning

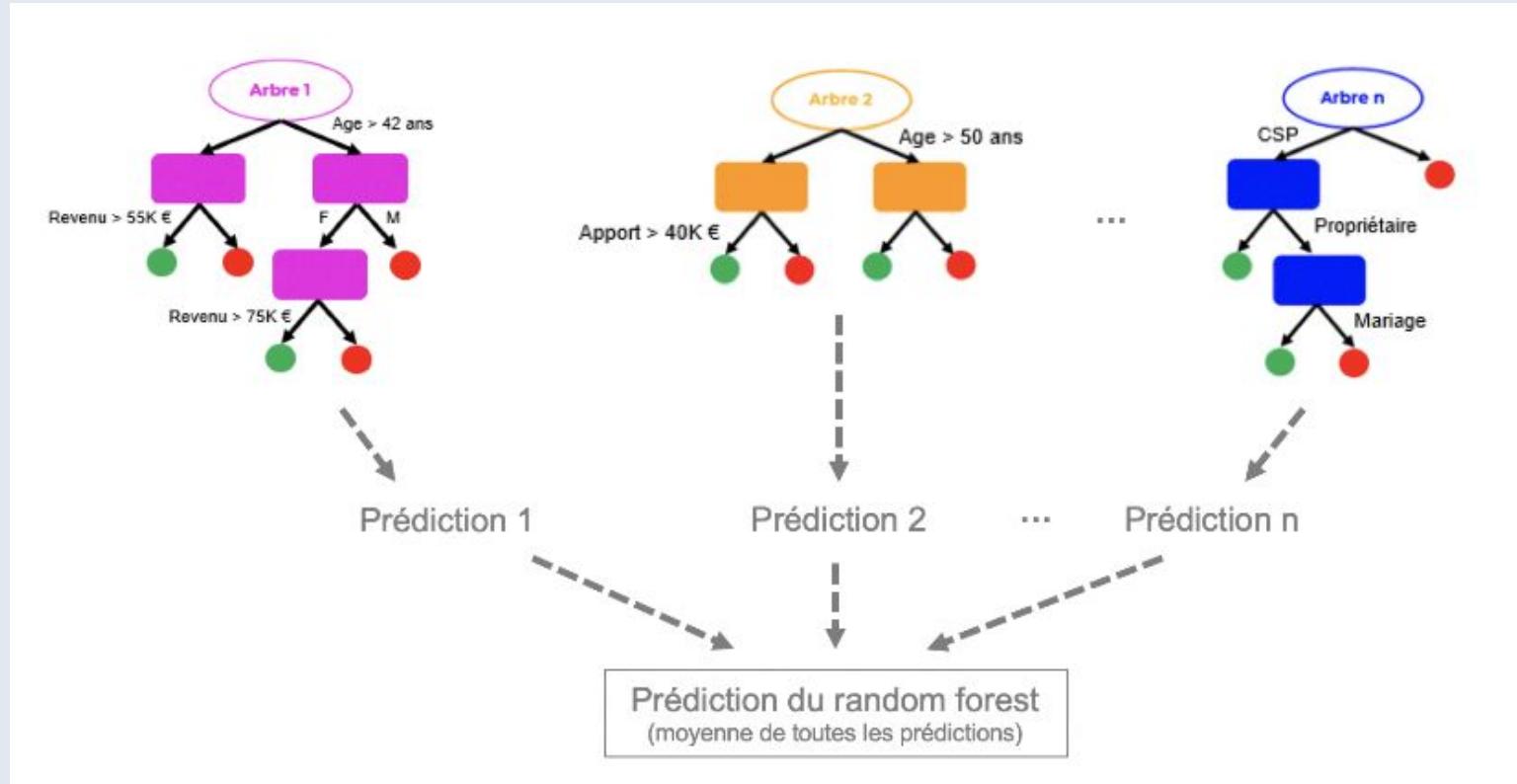
Survival of passengers on the Titanic

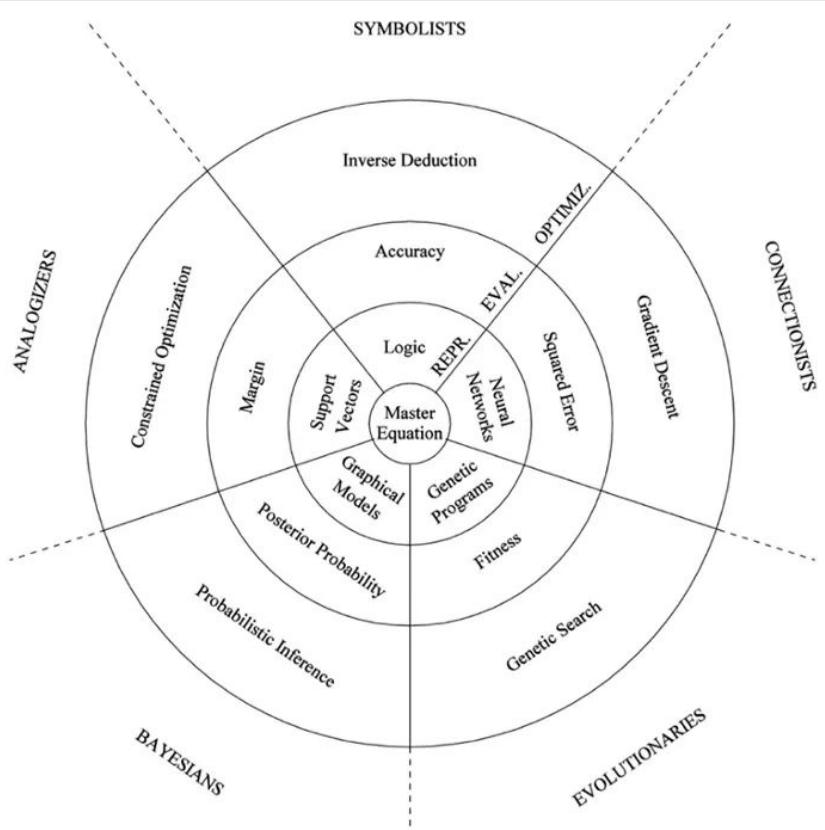


Choosing a variable at each step that best splits the set of items

- Gini index
- Information gain
- ...

Ensemble strategy: Random Forests



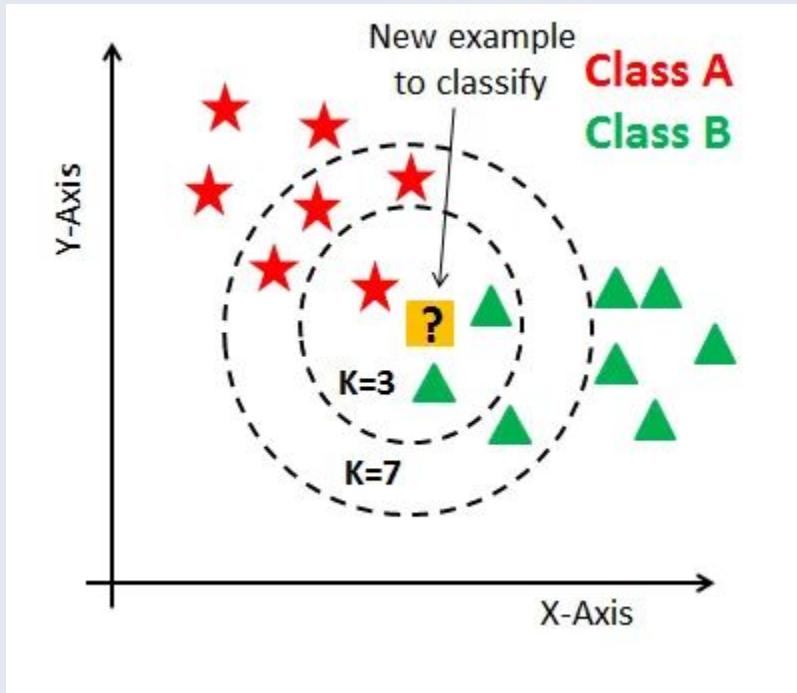


Analogizers

Analysing similarities and distances

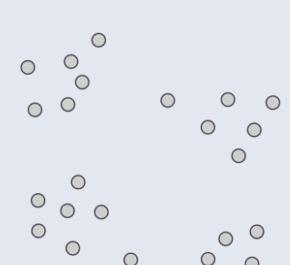
Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

K-nearest neighbors (KNN)

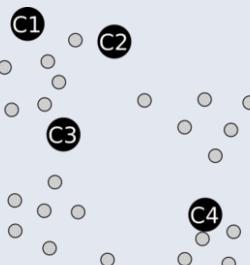


Classification

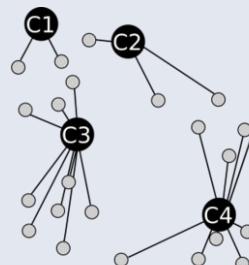
K-means clustering



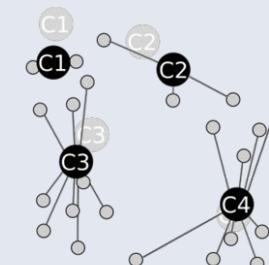
0a. Données d'entrée



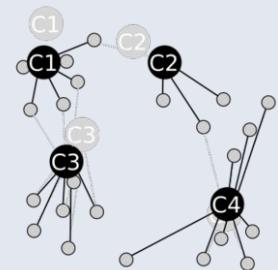
0b. initialisation



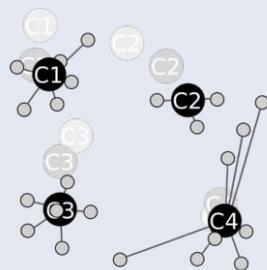
1a. assignation



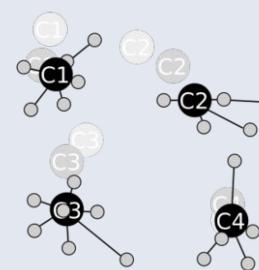
1b. calcul des points moyens



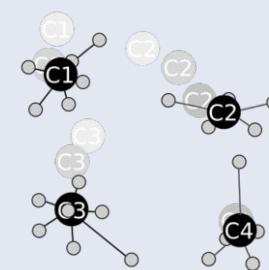
2a. assignation



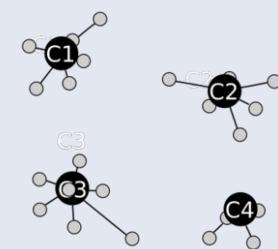
2b. calcul des points moyens



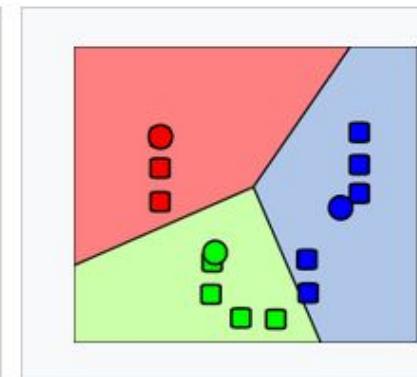
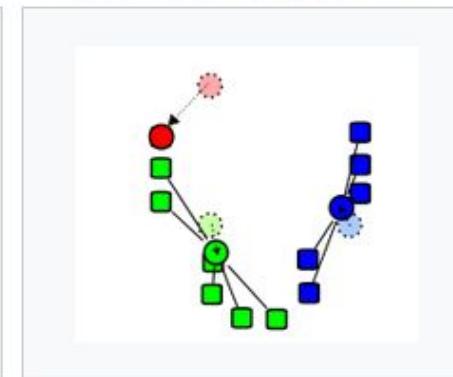
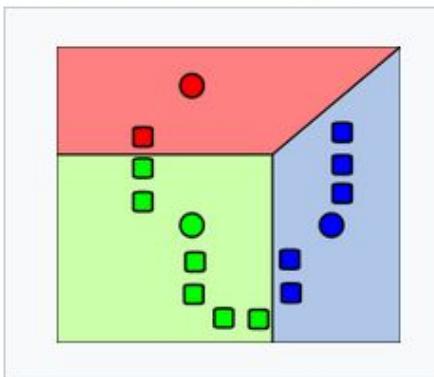
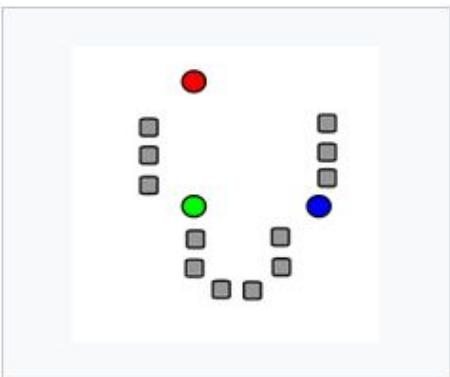
3a. assignation



3b. calcul des points moyens

4a. assignation
clusters stables (fin)

Demonstration of the standard algorithm



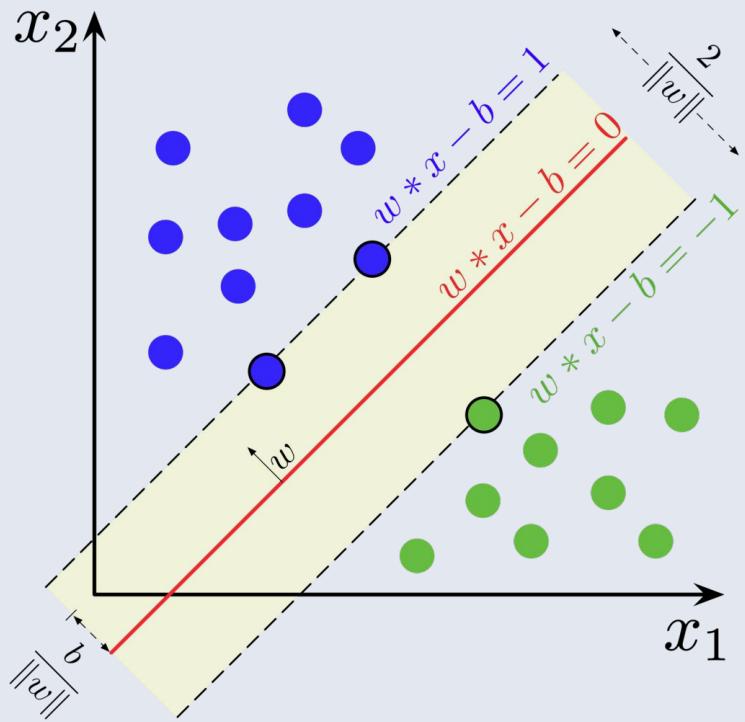
1. k initial "means" (in this case $k=3$) are randomly generated within the data domain (shown in color).

2. k clusters are created by associating every observation with the nearest mean. The partitions here represent the [Voronoi diagram](#) generated by the means.

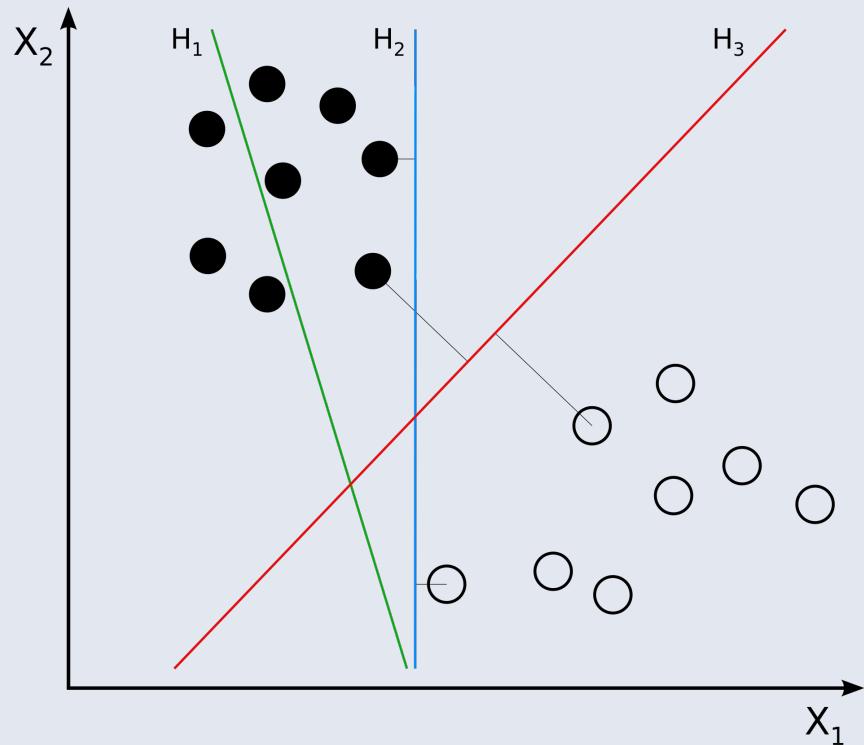
3. The [centroid](#) of each of the k clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.

Support Vector Machine (SVM)

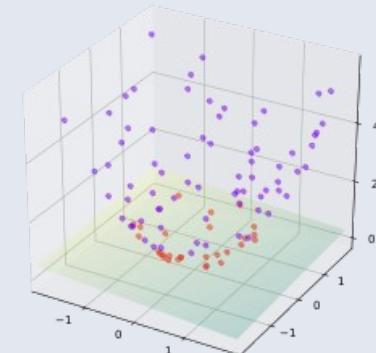
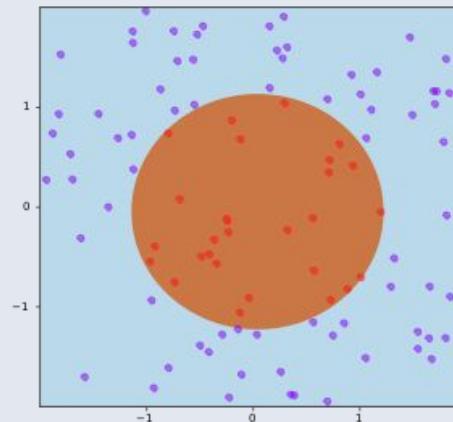
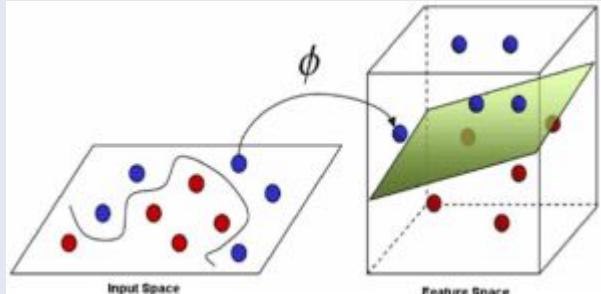


Supervised **binary** classification



Support Vector Machine (SVM)

Kernel trick



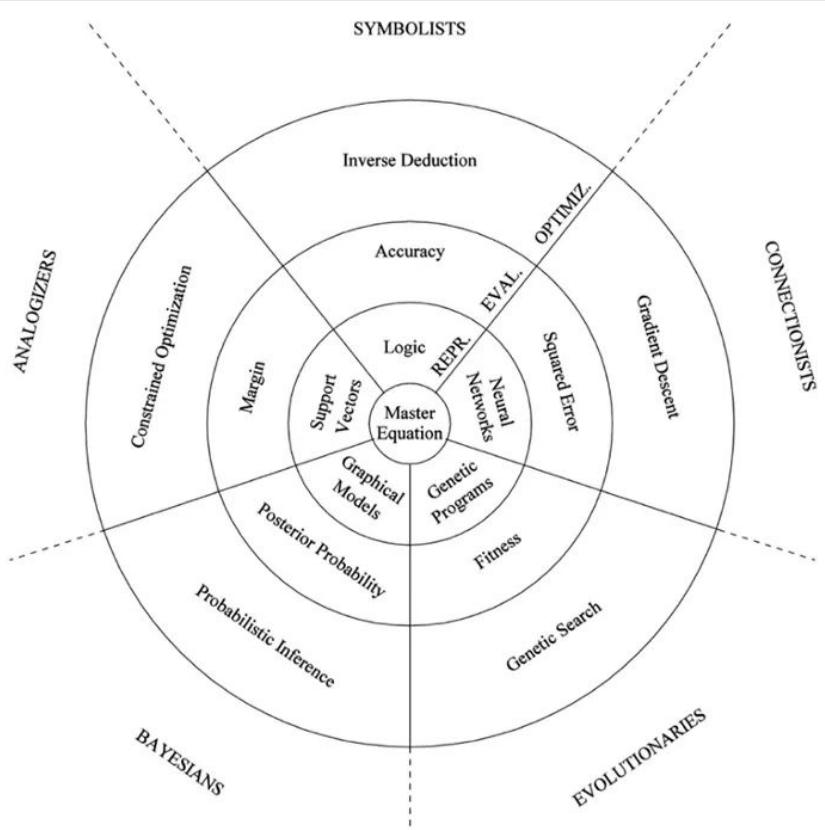
SVM with kernel given by $\varphi((a, b)) = (a, b, a^2 + b^2)$ and thus
 $k(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$. The training points are mapped to a 3-dimensional space where a separating hyperplane can be easily found.



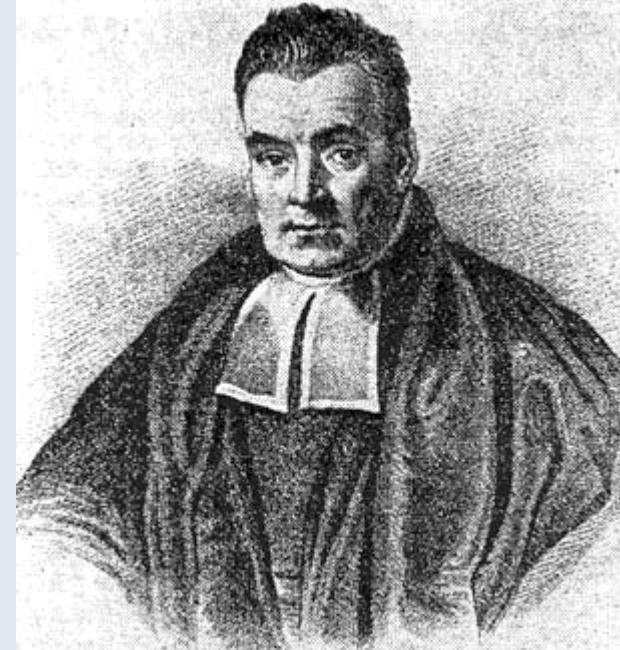
Support Vector Machine (SVM)

SVM are for binary classification: what if I have $k > 2$ classes ?

Transform the problem => We need k binary classifiers !



Bayesians



Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

1701-1761

Bayes theorem

Bayes Theorem explained by your
high-school math teacher

A chalkboard with the Bayes Theorem formula written in blue chalk. The formula is $P(A|B) = \frac{P(B|A) P(A)}{P(B)}$. The board has some faint, illegible writing and markings in the background.

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

Bayes theorem

Bayes Theorem explained by your
high-school math teacher

A chalkboard showing a hand-drawn derivation of Bayes' Theorem. The formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is written in blue chalk. The term $P(A)$ is underlined. The background shows a classroom setting with desks and chairs.

Bayes Theorem explained by a data
scientist

A diagram illustrating the components of Bayes' Theorem. The formula is shown: $P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$. Arrows point from labels to parts of the formula:

- Prior Probability points to $P(H)$.
- Likelihood of the evidence 'E' if the Hypothesis 'H' is true points to $P(E|H)$.
- Posterior Probability of 'H' given the evidence points to $P(H|E)$.
- Prior probability that the evidence itself is true points to $P(E)$.

Coin Flip



$$P(H|C_1) = 0.1$$



$$P(H|C_2) = 0.5$$



$$P(H|C_3) = 0.9$$

Which coin will I use?

$$P(C_1) = 1/3$$

$$P(C_2) = 1/3$$

$$P(C_3) = 1/3$$

Prior: Probability of a hypothesis
before we make any observations

3

Coin Flip



$$P(H|C_1) = 0.1$$



$$P(H|C_2) = 0.5$$



$$P(H|C_3) = 0.9$$

Which coin will I use?

$$P(C_1) = 1/3$$

$$P(C_2) = 1/3$$

$$P(C_3) = 1/3$$

Uniform Prior: All hypothesis are equally likely
before we make any observations

4

Experiment I: Heads

Which coin did I use?

$$P(C_1|H) = ? \quad P(C_2|H) = ? \quad P(C_3|H) = ?$$

$$P(C_1|H) = \frac{P(H|C_1)P(C_1)}{P(H)} \quad P(H) = \sum_{i=1}^3 P(H|C_i)P(C_i)$$



$$P(H|C_1) = 0.1 \\ P(C_1) = 1/3$$



$$P(H|C_2) = 0.5 \\ P(C_2) = 1/3$$



$$P(H|C_3) = 0.9 \\ P(C_3) = 1/3$$

5

Experiment I: Heads

Which coin did I use?

$$P(C_1|H) = 0.066 \quad P(C_2|H) = 0.333 \quad P(C_3|H) = 0.6$$

Posterior: Probability of a hypothesis given data



$$P(H|C_1) = 0.1 \\ P(C_1) = 1/3$$



$$P(H|C_2) = 0.5 \\ P(C_2) = 1/3$$



$$P(H|C_3) = 0.9 \\ P(C_3) = 1/3$$

6

Experiment 2: Tails

Which coin did I use?

$$P(C_1|HT) = ? \quad P(C_2|HT) = ? \quad P(C_3|HT) = ?$$

$$P(C_1|HT) = \alpha P(HT|C_1)P(C_1) = \alpha P(H|C_1)P(T|C_1)P(C_1)$$



$$P(H|C_1) = 0.1 \\ P(C_1) = 1/3$$



$$P(H|C_2) = 0.5 \\ P(C_2) = 1/3$$



$$P(H|C_3) = 0.9 \\ P(C_3) = 1/3$$

7

Experiment 2: Tails

Which coin did I use?

$$P(C_1|HT) = 0.21 \quad P(C_2|HT) = 0.58 \quad P(C_3|HT) = 0.21$$

$$P(C_1|HT) = \alpha P(HT|C_1)P(C_1) = \alpha P(H|C_1)P(T|C_1)P(C_1)$$



$$P(H|C_1) = 0.1 \\ P(C_1) = 1/3$$



$$P(H|C_2) = 0.5 \\ P(C_2) = 1/3$$



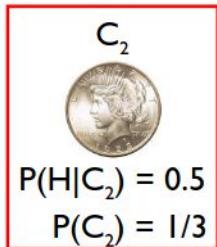
$$P(H|C_3) = 0.9 \\ P(C_3) = 1/3$$

8

Experiment 2: Tails

Which coin did I use?

$$P(C_1|HT) = 0.21 \quad P(C_2|HT) = 0.58 \quad P(C_3|HT) = 0.21$$



9

Your Estimate?

What is the probability of heads after two experiments?

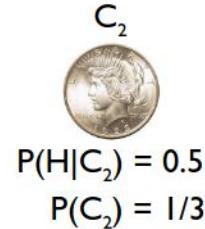
Most likely coin:

C_2



Best estimate for $P(H)$

$$P(H|C_2) = 0.5$$

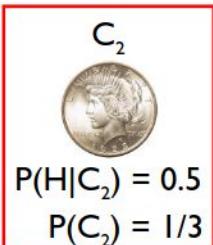


10

Experiment 2: Tails

Which coin did I use?

$$P(C_1|HT) = 0.21 \quad P(C_2|HT) = 0.58 \quad P(C_3|HT) = 0.21$$



9

Your Estimate?

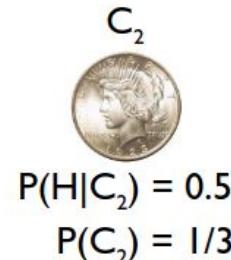
Maximum Likelihood Estimate: The best hypothesis that fits observed data assuming uniform prior

Most likely coin:



Best estimate for $P(H)$

$$P(H|C_2) = 0.5$$



11

Using Prior Knowledge

- Should we always use Uniform Prior?
- Background knowledge:
 - Heads => you go first in Abalone against TA
 - TAs are nice people
 - => TA is more likely to use a coin biased in your favor

$$P(C_1) = 0.05 \quad P(C_2) = 0.25 \quad P(C_3) = 0.70$$



$$P(H|C_1) = 0.1 \quad P(H|C_2) = 0.5 \quad P(H|C_3) = 0.9$$

12

Using Prior Knowledge

- Should we always use **Uniform Prior**?
- Background knowledge:
 - Heads => you go first in Abalone against TA
 - TAs are nice people
 - => TA is more likely to use a coin biased in your favor

$$P(C_1) = 0.05$$

C_1



$$P(C_2) = 0.25$$

C_2



$$P(C_3) = 0.70$$

C_3



$$P(H|C_1) = 0.1$$

$$P(H|C_2) = 0.5$$

$$P(H|C_3) = 0.9$$

12

Your Estimate?

Maximum A Posteriori (MAP) Estimate: The best hypothesis that fits observed data assuming a non-uniform prior

Most likely coin:



Best estimate for $P(H)$

$$P(H|C_3) = 0.9$$

C_3



$$P(H|C_3) = 0.9$$

$$P(C_3) = 0.70$$

20

Using Prior Knowledge

- Should we always use **Uniform Prior**?
- Background knowledge:
 - Heads => you go first in Abalone against TA
 - TAs are nice people
 - => TA is more likely to use a coin biased in your favor

$$P(C_1) = 0.05$$



$$P(C_2) = 0.25$$



$$P(C_3) = 0.70$$



$$P(H|C_1) = 0.1$$

$$P(H|C_2) = 0.5$$

$$P(H|C_3) = 0.9$$

12

Bayesian Estimate

Bayesian Estimate: Minimizes prediction error, given data and (generally) assuming a non-uniform prior

$$P(H) = \sum_{i=1}^3 P(H|C_i)P(C_i) = 0.680$$

$$P(C_1|HT) = 0.035 \quad P(C_2|HT) = 0.481 \quad P(C_3|HT) = 0.485$$

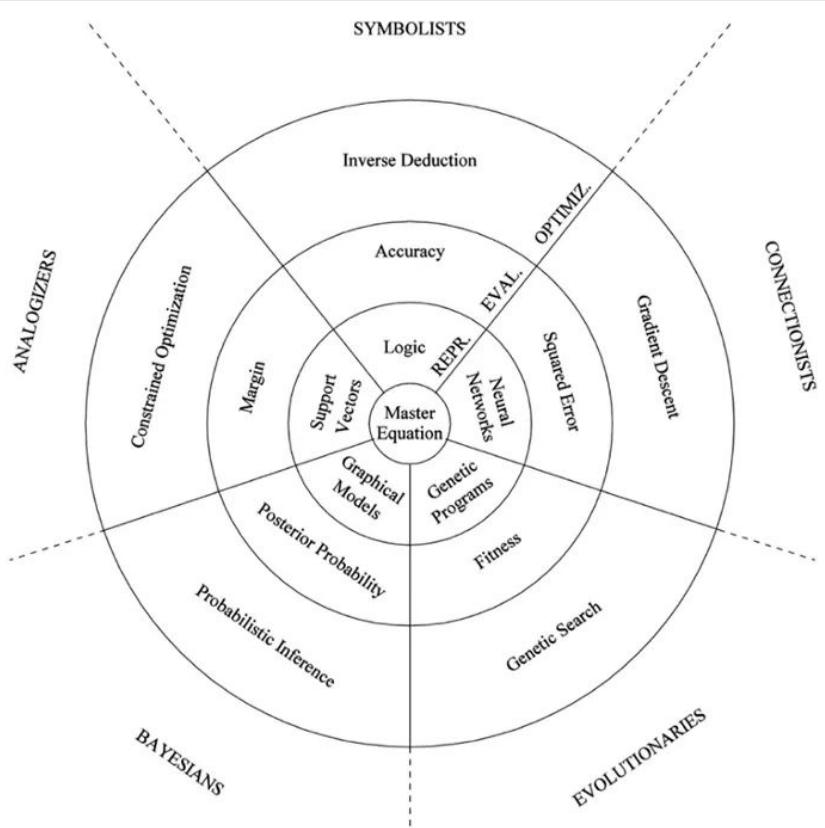
 C_1  C_3

$$P(H|C_1) = 0.1$$

$$P(H|C_2) = 0.5$$

$$P(H|C_3) = 0.9$$

24



Pedro Domingos. "The Master Algorithm : How the Quest for the Ultimate Learning Machine Will Remake Our World"

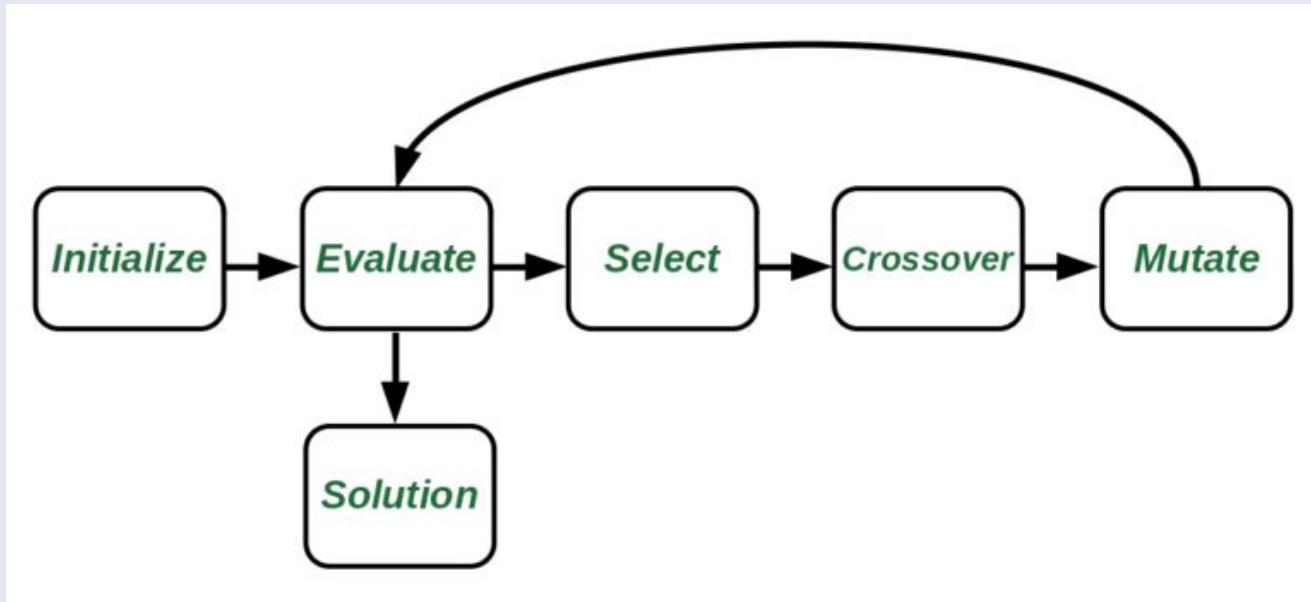
Evolutionaries

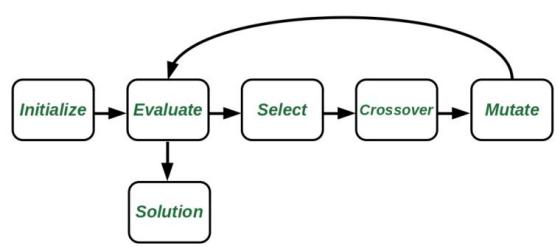
and population-based approaches



© CanStockPhoto.com

Evolutionaries





Genetic Algorithms

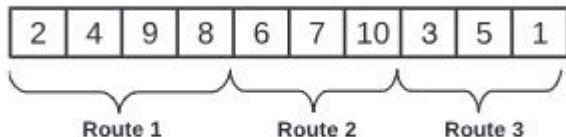


Fig. 4. Genotype with sequence of stops/customers indexes.

Customer index	1	2	3	4	5	6	7	8	9	10
Depot index	3	1	3	1	3	2	2	1	1	2

Fig. 5. Genotype with route index.

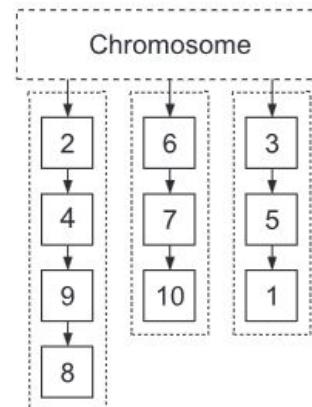


Fig. 7. Genotype with multiple arrays – one for each route.

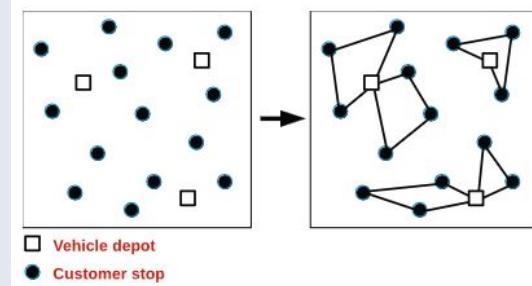


Fig. 2. Multi depot vehicle routing problem.

Example: The multi-depot vehicle routing problem

S. Karakatić, V. Podgorelec, A survey of genetic algorithms for solving multi depot vehicle routing problem, Appl. Soft Comput. J. (2014), <http://dx.doi.org/10.1016/j.asoc.2014.11.005>

Evolutionaries

	Parents									
Parents	2	4	9	8	6	7	10	3	5	1
Children	3	9	4	1	5	8	6	10	7	2
	2	4	9	8	3	1	5	6	10	7
Children	3	9	4	1	2	8	6	7	10	5

Fig. 8. Linear ordered crossover (LOX).

	Parents									
Parents	2	4	9	8	6	7	10	3	5	1
Children	3	9	4	1	5	8	6	10	7	2
	4	1	5	8	6	7	10	3	2	9
Children	4	9	7	3	5	8	6	10	1	2

Fig. 9. Ordered crossover (OX).

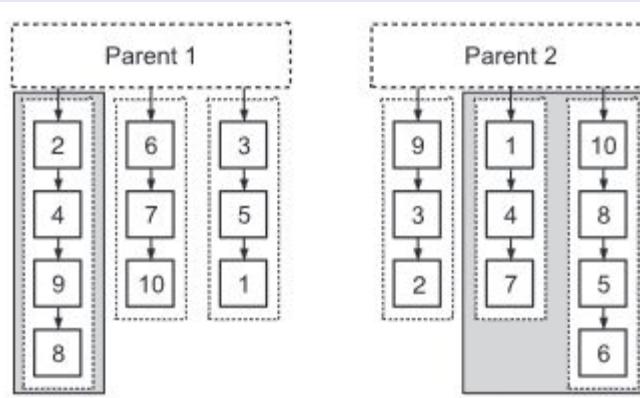


Fig. 11. Route based crossover (RBX).

S. Karakatić, V. Podgorelec, A survey of genetic algorithms for solving multi depot vehicle routing problem, Appl. Soft Comput. J. (2014), <http://dx.doi.org/10.1016/j.asoc.2014.11.005>

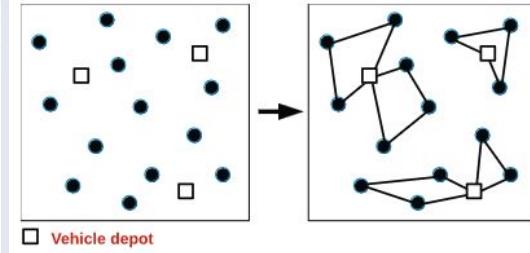


Fig. 2. Multi depot vehicle routing problem.

Example: The multi-depot vehicle routing problem

Evolutionaries

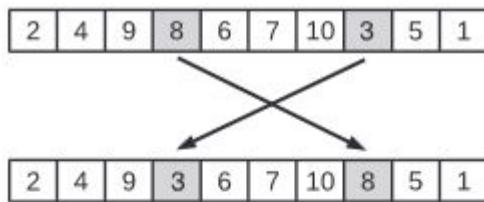


Fig. 13. Exchange mutation.

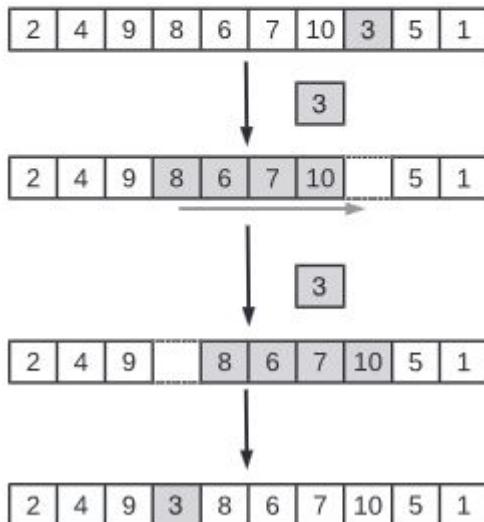


Fig. 14. Insertion mutation.

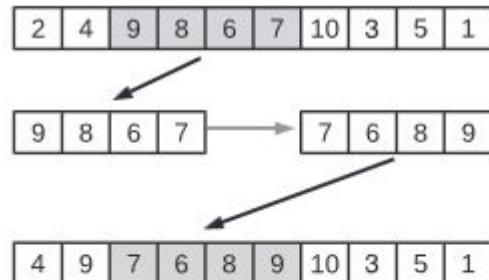


Fig. 15. Inversion mutation.

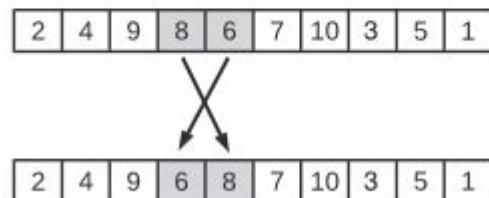


Fig. 16. Swap mutation.

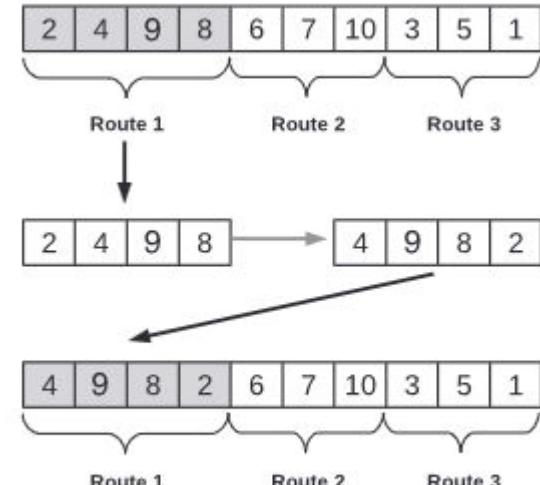
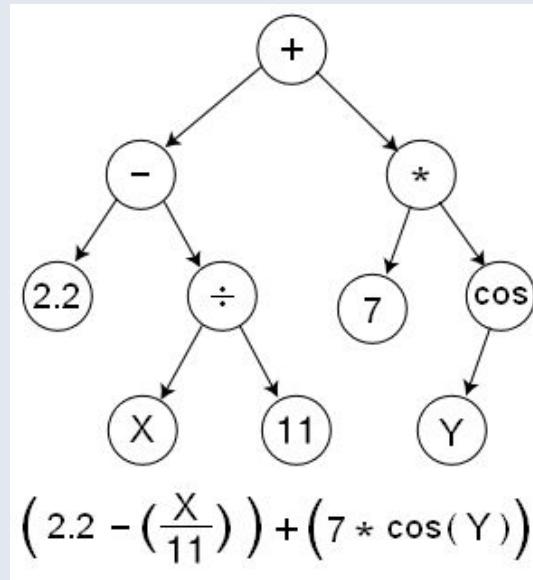
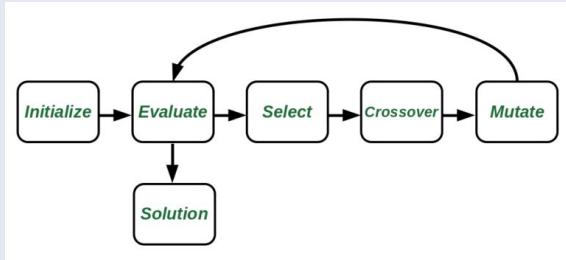


Fig. 17. Local search optimization mutation.

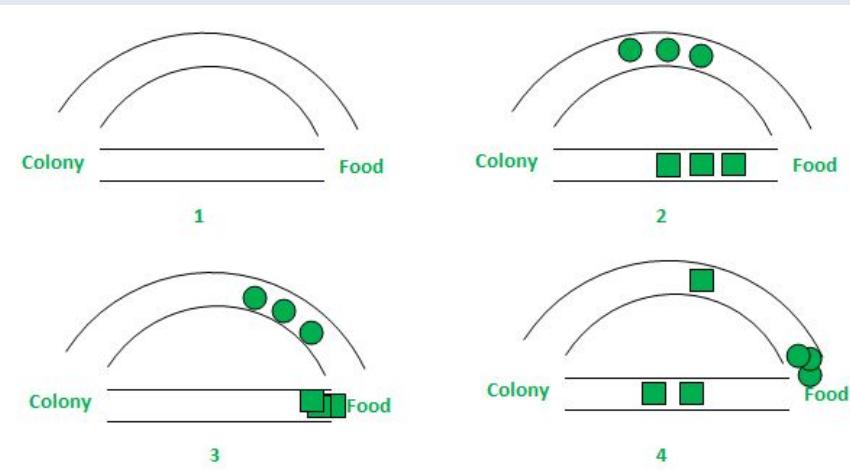
S. Karakatić, V. Podgorelec, A survey of genetic algorithms for solving multi depot vehicle routing problem, Appl. Soft Comput. J. (2014), <http://dx.doi.org/10.1016/j.asoc.2014.11.005>

Genetic Programming

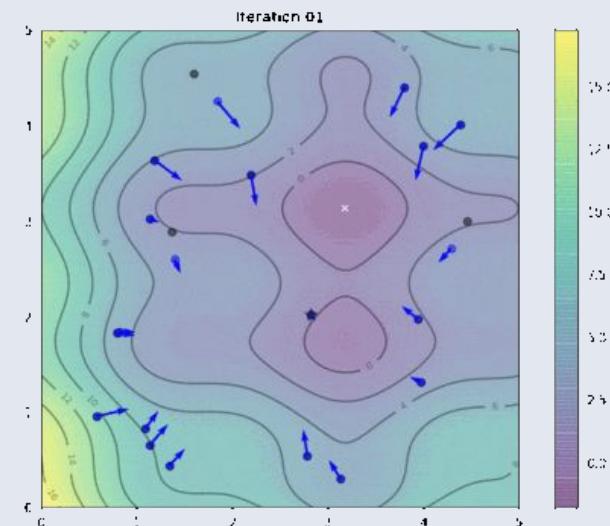


Other population-based approaches: Swarm Intelligence

Ant colony optimization

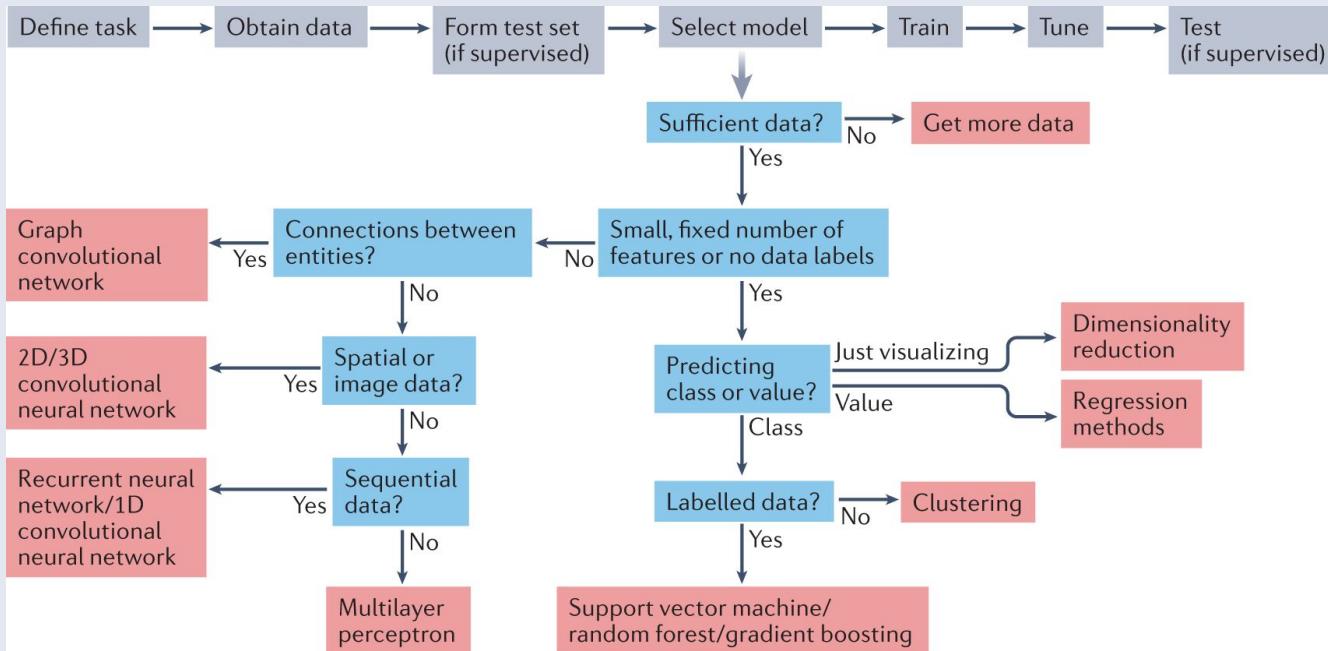


Particle Swarm optimization



<https://machinelearningmastery.com/a-gentle-introduction-to-particle-swarm-optimization/>

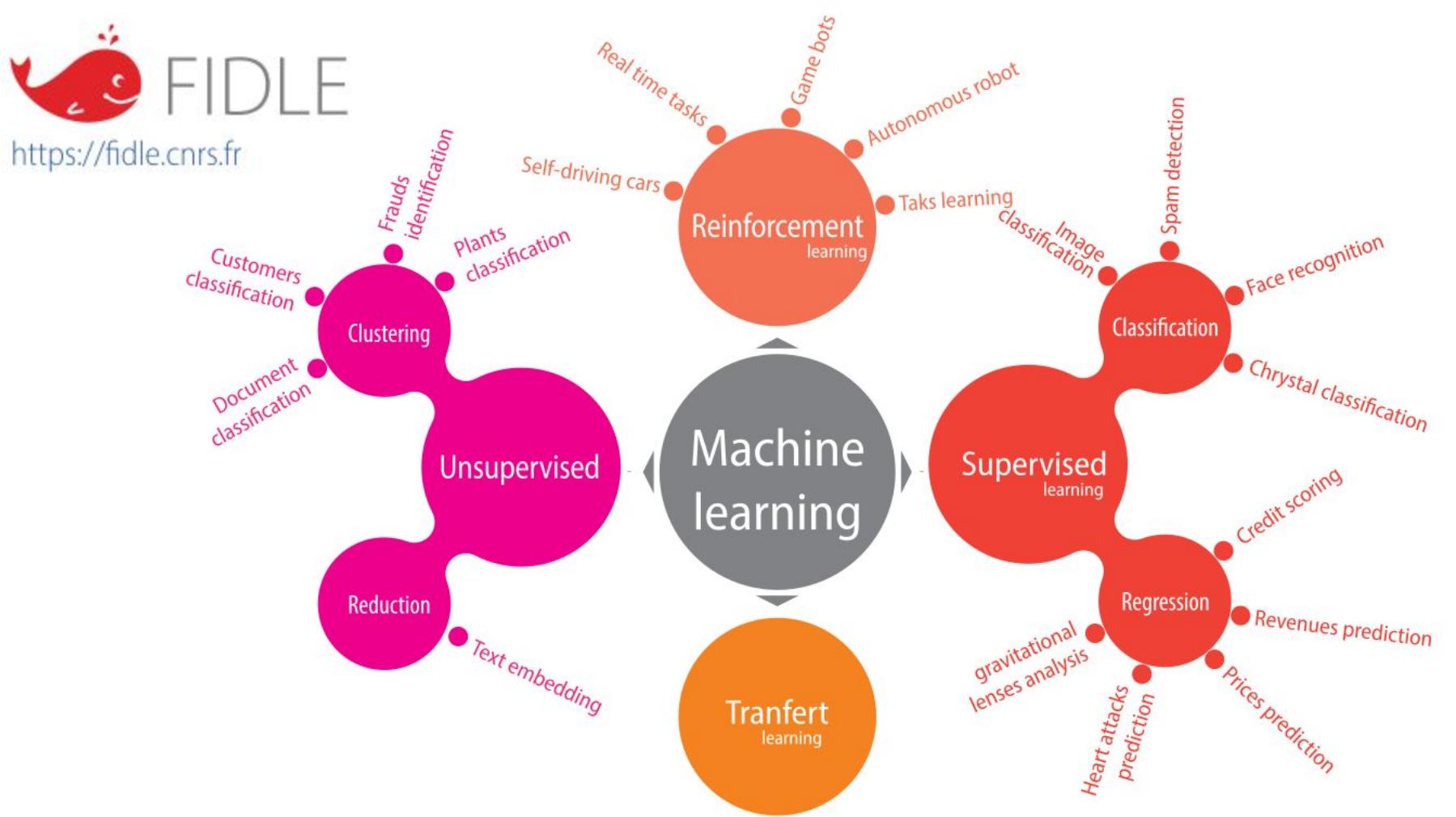
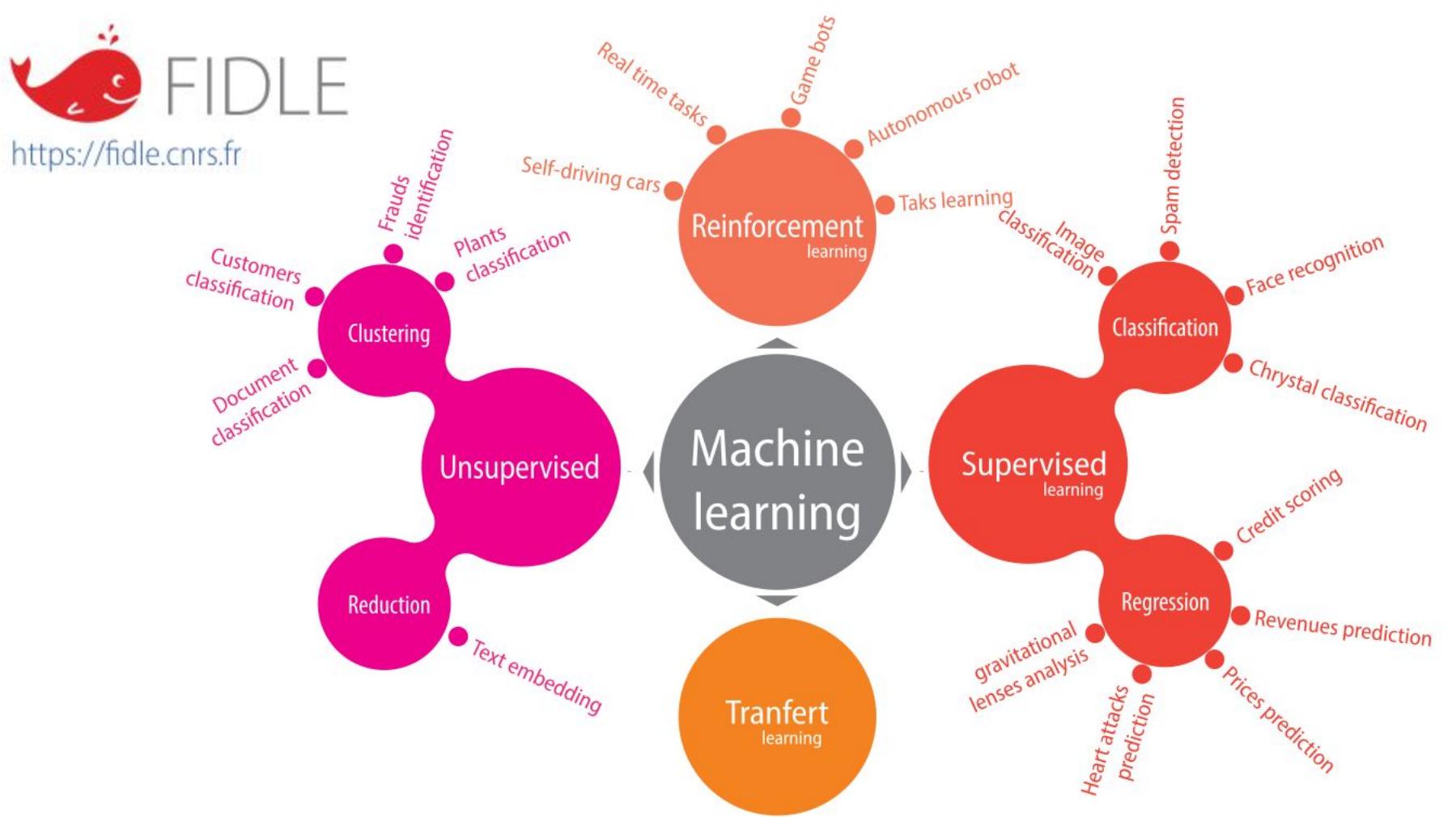
In a nutshell...

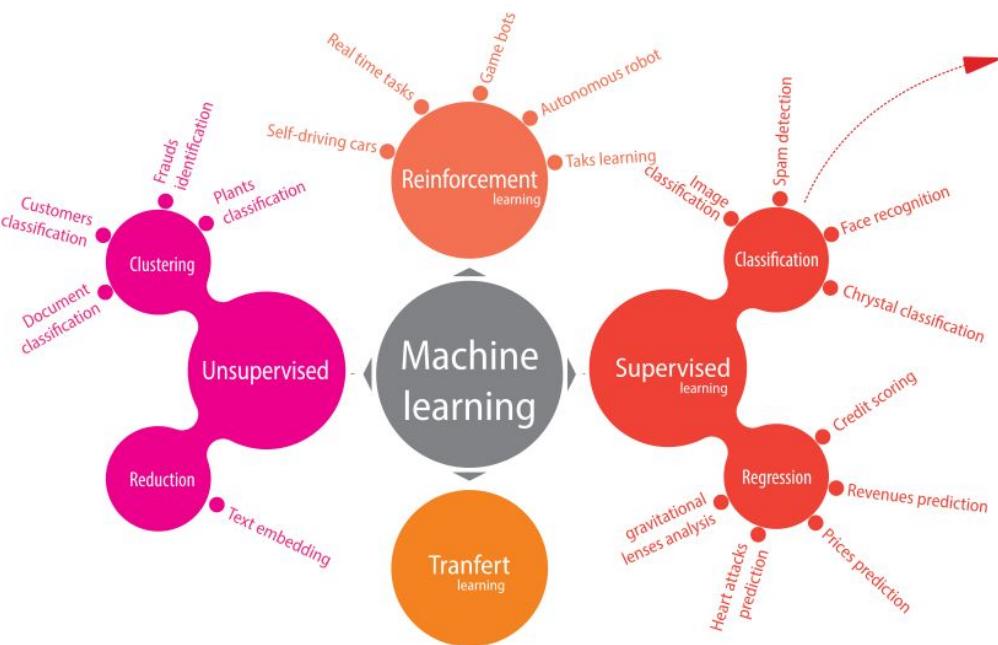


Greener, J.G., Kandathil, S.M., Moffat, L. et al. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 23, 40–55 (2022). <https://doi.org/10.1038/s41580-021-00407-0>

Applications

Focus on bioinformatics





Classification :

Predict qualitative informations



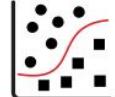
This is a cat

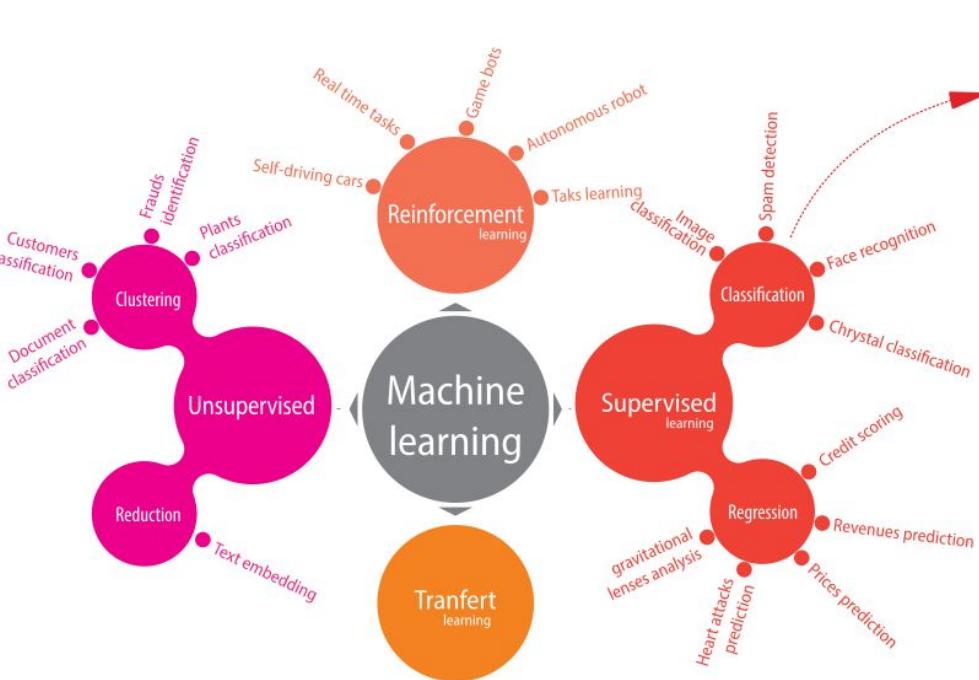


This is a rabbit



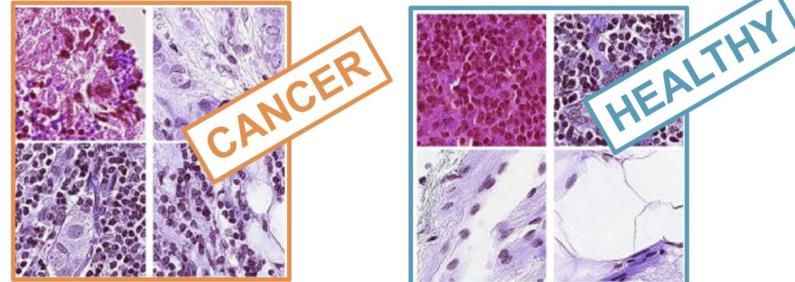
Tell me,
what is it ?





Classification :

Example: Identification of metastases in lymph nodes biopsies





Classification :

Customers classification
Document classification
Cluster

Red

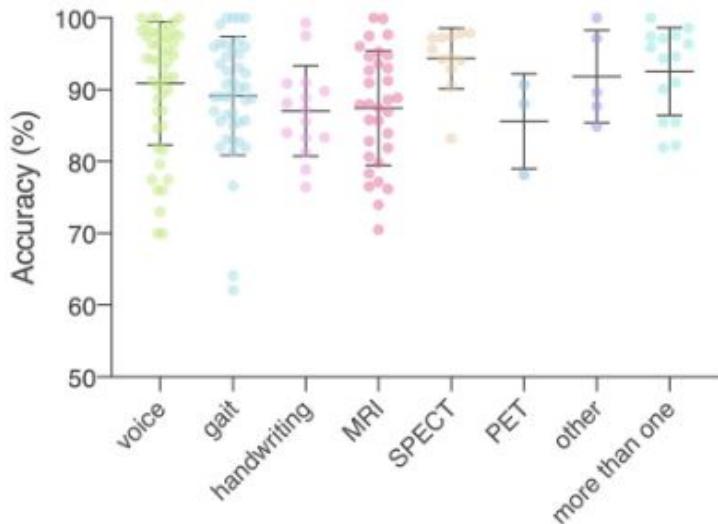
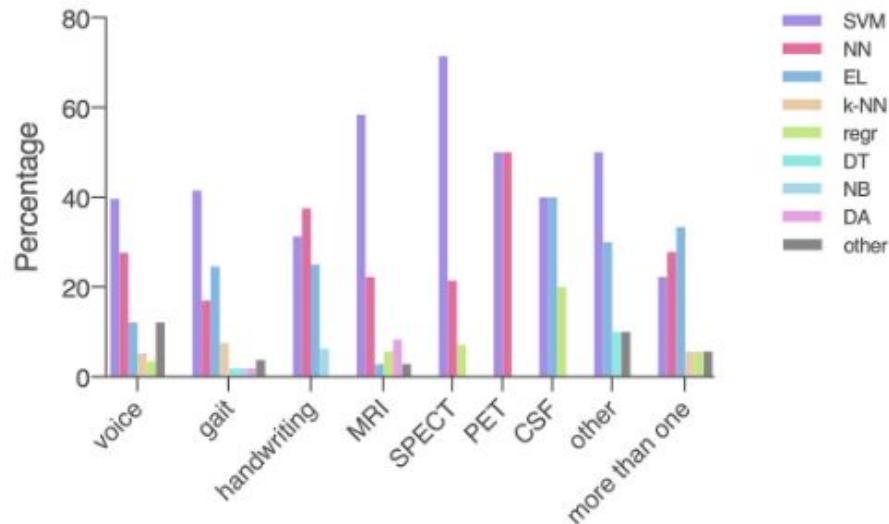
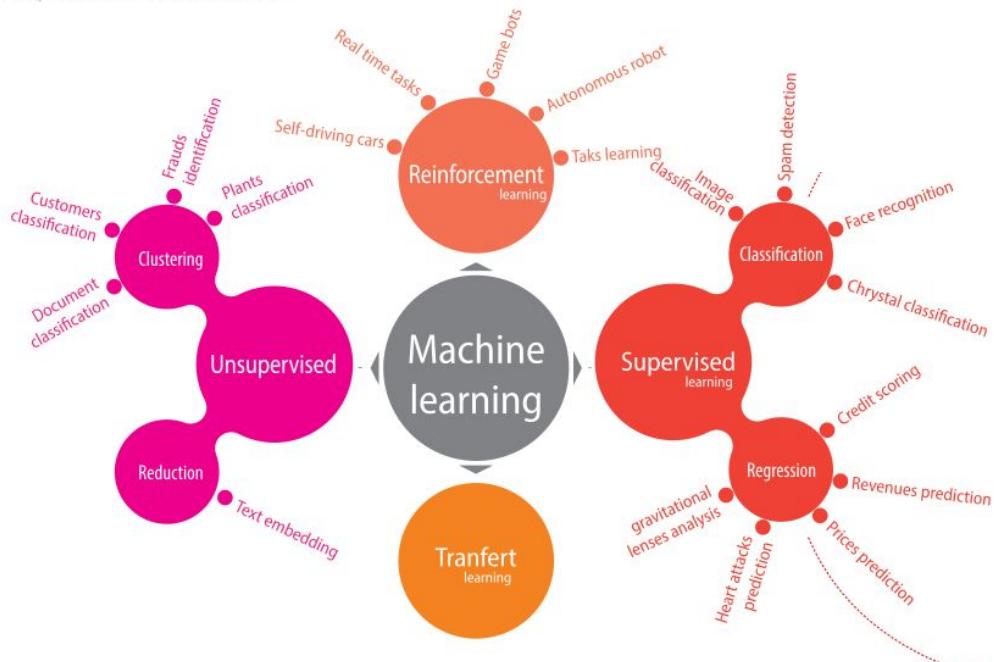
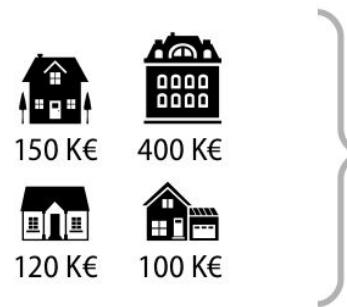
A**B**

FIGURE 4 | Data type, machine learning models applied, and accuracy. **(A)** Accuracy achieved in individual studies and average accuracy for each data type. Error bar: standard deviation. **(B)** Distribution of machine learning models applied per data type. MRI, magnetic resonance imaging; SPECT, single-photon emission computed tomography; PET, positron emission tomography; CSF, cerebrospinal fluid; SVM, support vector machine; NN, neural network; EL, ensemble learning; k-NN, nearest neighbor; regr, regression; DT, decision tree; NB, naïve Bayes; DA, discriminant analysis; other: data/models that do not belong to any of the given categories.

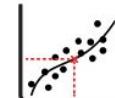


Régression :

Predict quantitative informations

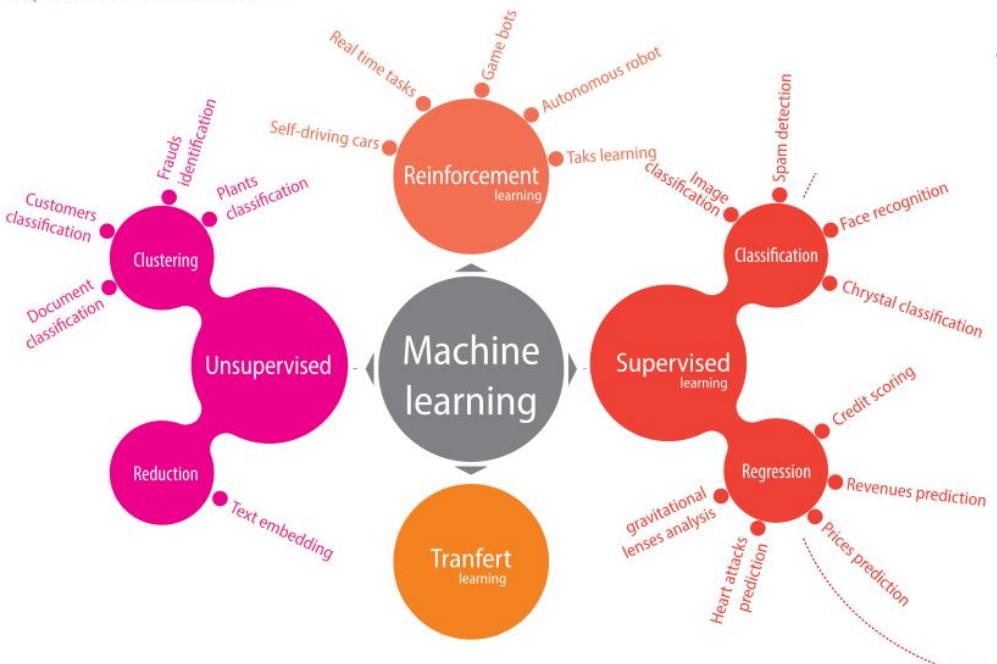



Tell me,
what's the
price ?

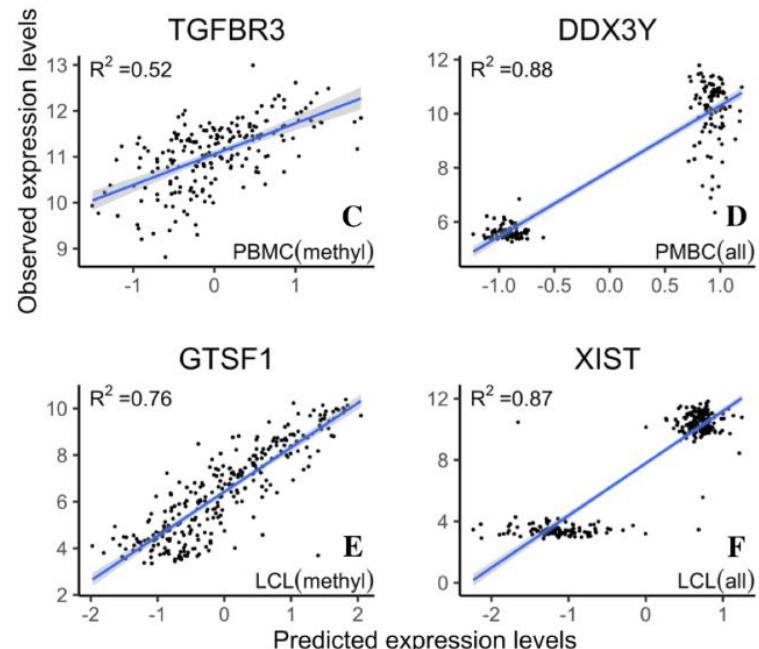


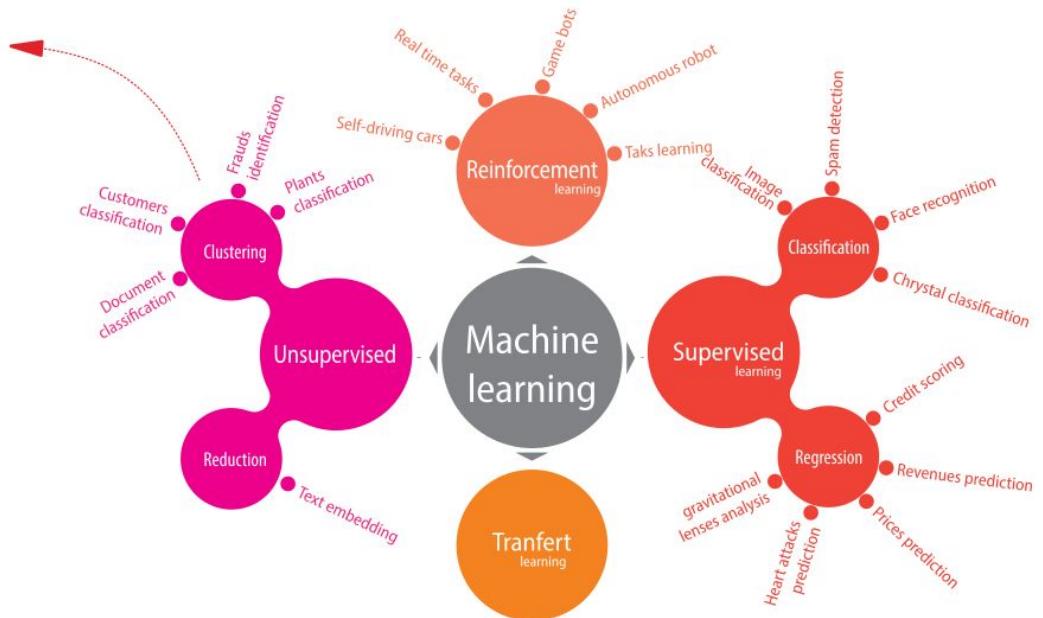
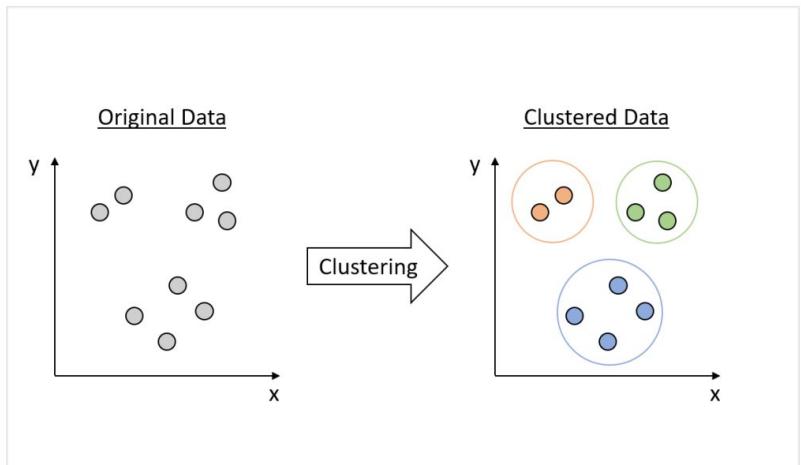


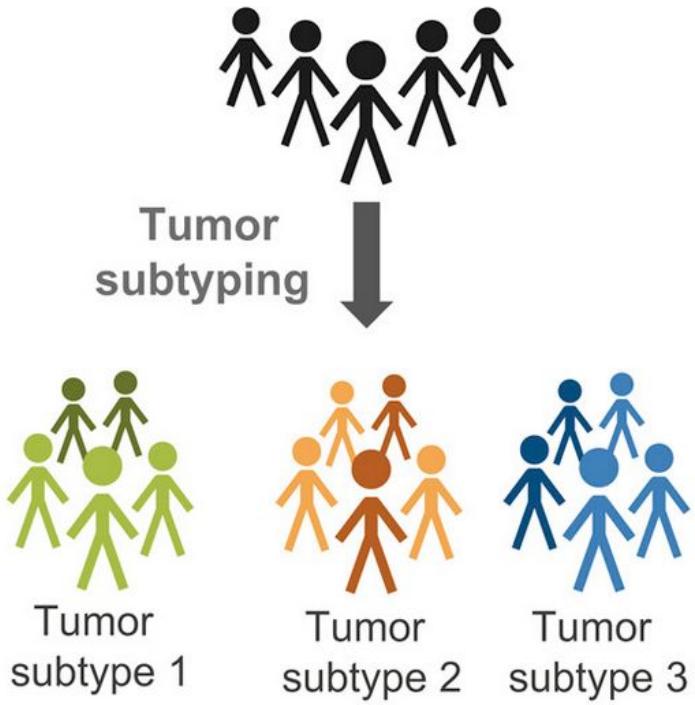
<https://fidle.cnrs.fr>



Zhong, Huan, et al. "Predicting gene expression using DNA methylation in three human populations." *PeerJ* 7 (2019): e6757.

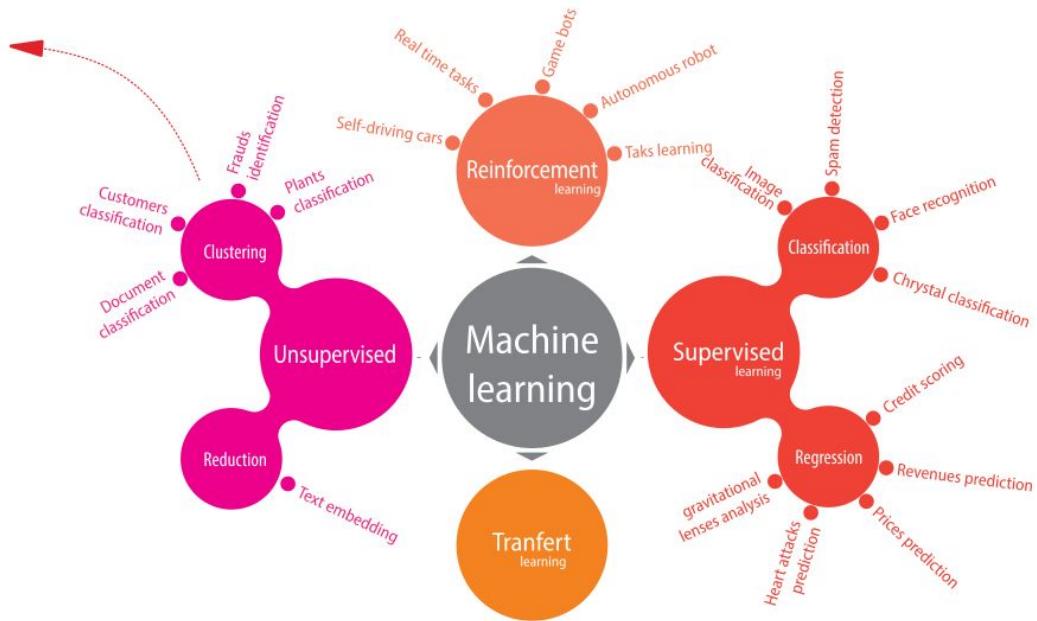




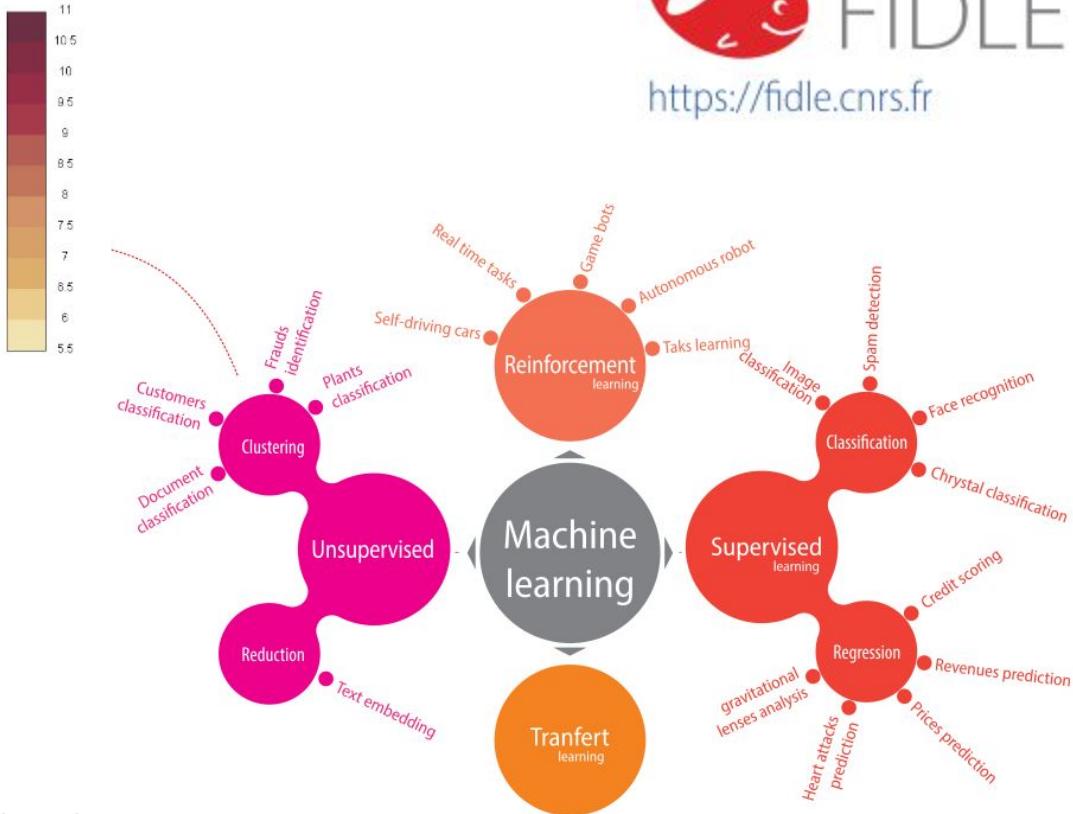
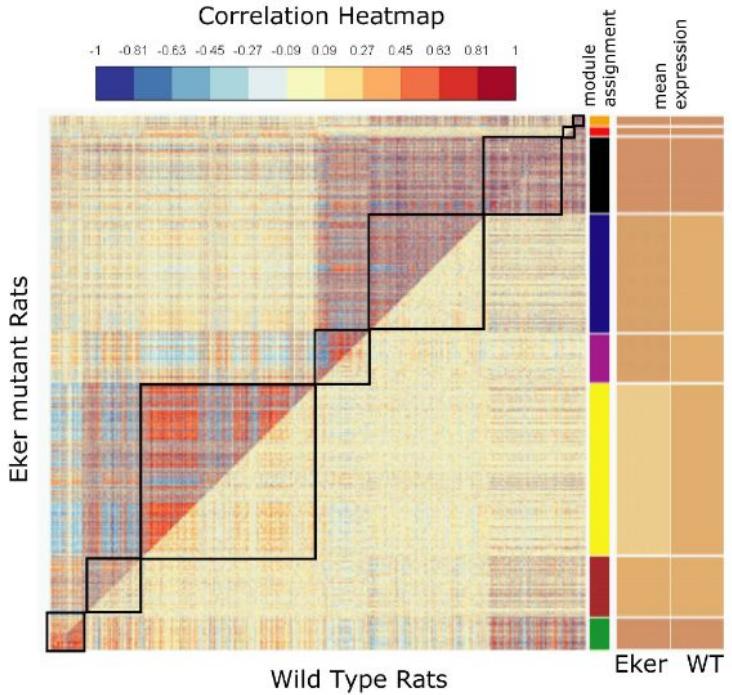


Disease subtyping

Bramsen, Jesper Bertram, et al. "Molecular-subtype-specific biomarkers improve prediction of prognosis in colorectal cancer." *Cell reports* 19.6 (2017): 1268-1280.



A



Co-expression analysis

Tesson, B.M., Breitling, R. & Jansen, R.C. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* 11, 497 (2010).
<https://doi.org/10.1186/1471-2105-11-497>

a

Global unsupervised overview of the data

Methods should focus on:
Ease of interpretation
Ease of visualization

Modules to unravel function and disease

Methods should model:
Local co-expression
Overlap

Inferring regulatory module networks

Methods should focus on:
Accuracy of inferred network

b

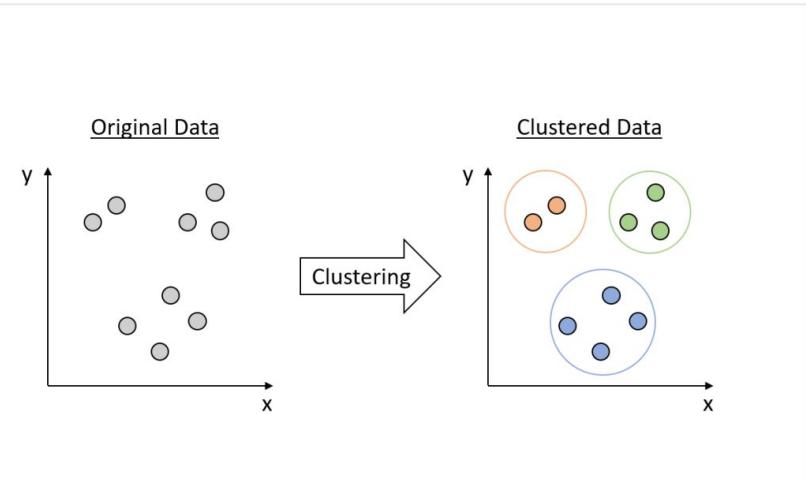
Supplementary Table 1

Module detection



	Module detection	Parameter estimation	Module visualization	Functional interpretation
Global unsupervised overview of the data		Balancing tightness and separateness <code>nbclust, clValid, clusterCrit</code>	Additional annotations: <i>pheatmap</i> , <i>gitools</i> , ... Interactivity: <i>d3heatmap</i> , <i>plotly</i> , <i>gitools</i> Alternatives visualizations: Parallel coordinates, Co-expression network	Functional enrichment: What biological functions are overrepresented in the module? Pathway analysis: What pathways are overrepresented in the module, and what downstream pathways do these induce? Disease associations: Which modules are associated with a particular disease?
Modules to unravel function and disease		Balancing functional coverage and module enrichment Modules Functional terms Module 1 DNA metabolism ✓ Module 1 Chemotaxis ✓ Module 2 Development ✗ Module 2 Cell cycle ✓ Module 3 Cell cycle ✓ 2/3 3/4	Relations between modules: <i>Furby</i> , <i>BicOverlapper</i> Alternative visualizations: Parallel coordinates Interactivity: <i>Furby</i> , <i>BicOverlapper</i> , <i>expressionView</i> Additional annotations: <i>Furby</i>	Functional enrichment, Pathway analysis, Disease associations
Inferring regulatory module networks	+ network inference 	Connections between modules: <i>Cytoscape</i> Regulatory programs: <i>Genomica</i> , <i>Lemon-tree</i> Enrichment map, <i>ReViGO</i> , <i>FGNet</i>	Extracting the main functions/pathways/diseases reducing redundancy DNA metabolism, Development, Cell cycle	

Saelens, Wouter, Robrecht Cannoodt, and Yvan Saeys. "A comprehensive evaluation of module detection methods for gene expression data." *Nature communications* 9.1 (2018): 1090.

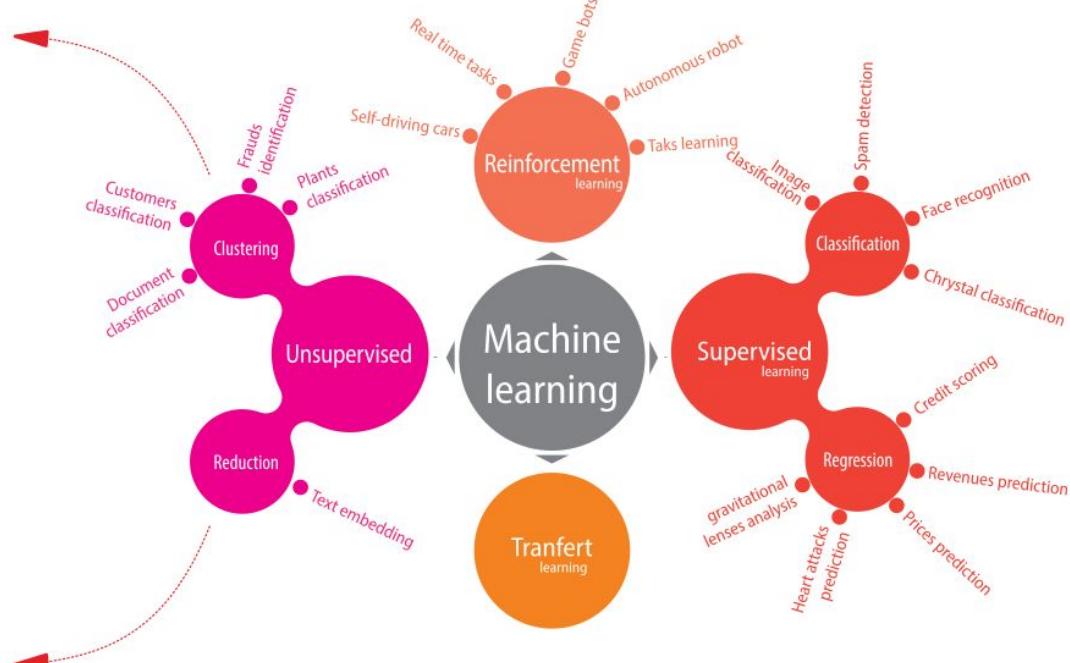


Reduction :

Reduce the number of dimensions



Simplify
while
keeping
meaning



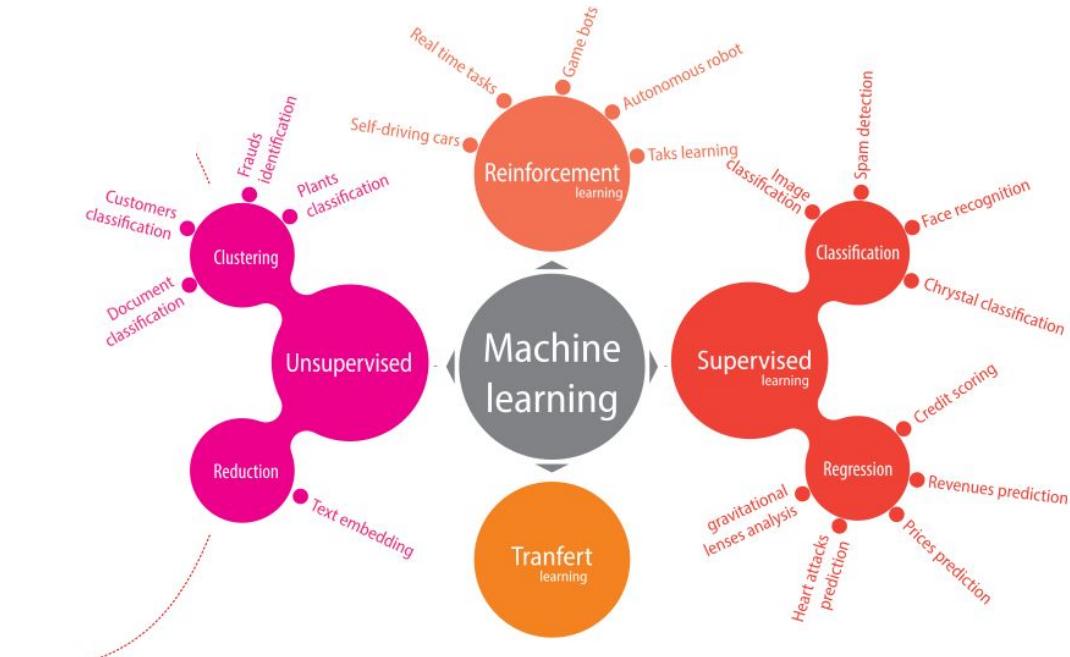
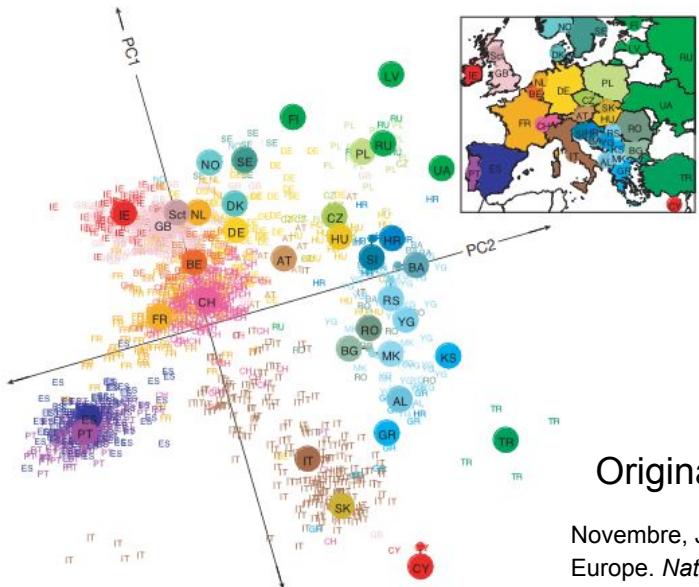
Simplification (for further downstream analysis)

Visualisation

Embedding (networks, words, ...)



a



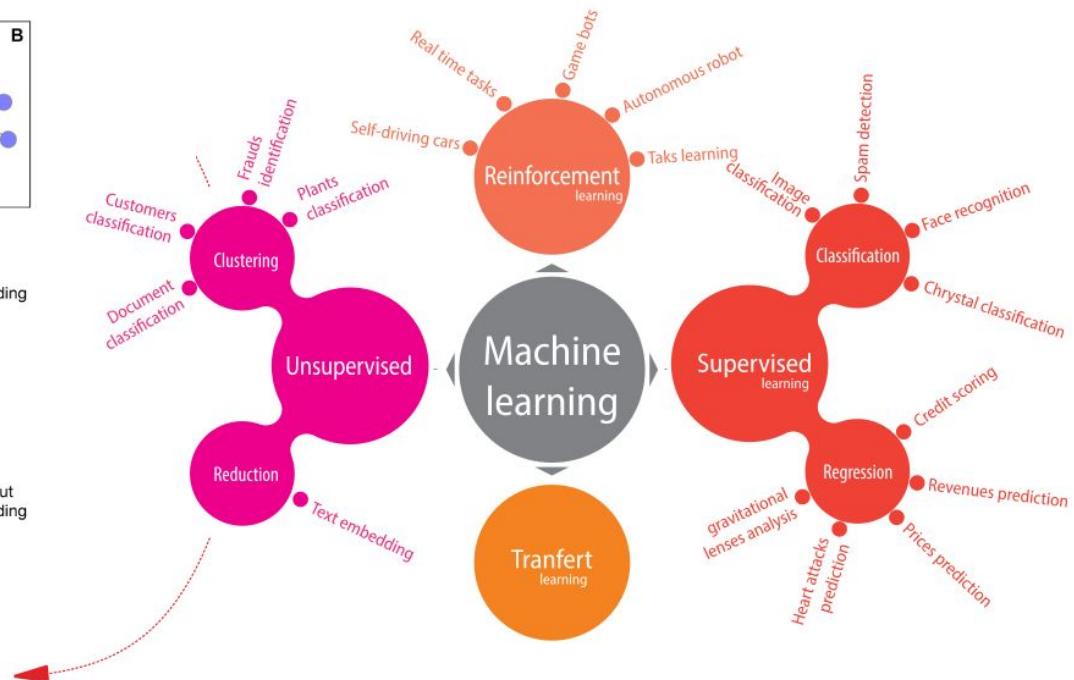
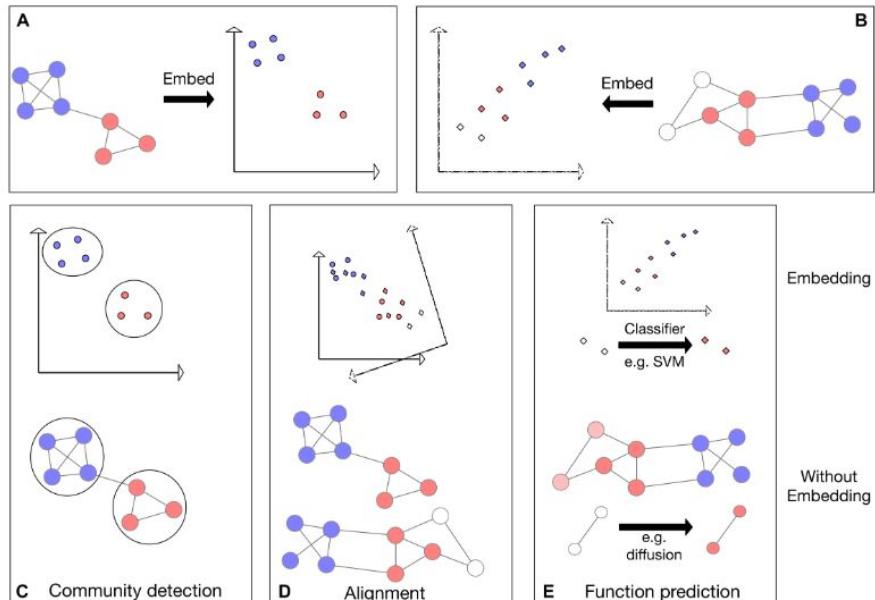
Original data = 500 000 variables !!

Novembre, J., Johnson, T., Bryc, K. et al. Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008). <https://doi.org/10.1038/nature07331>

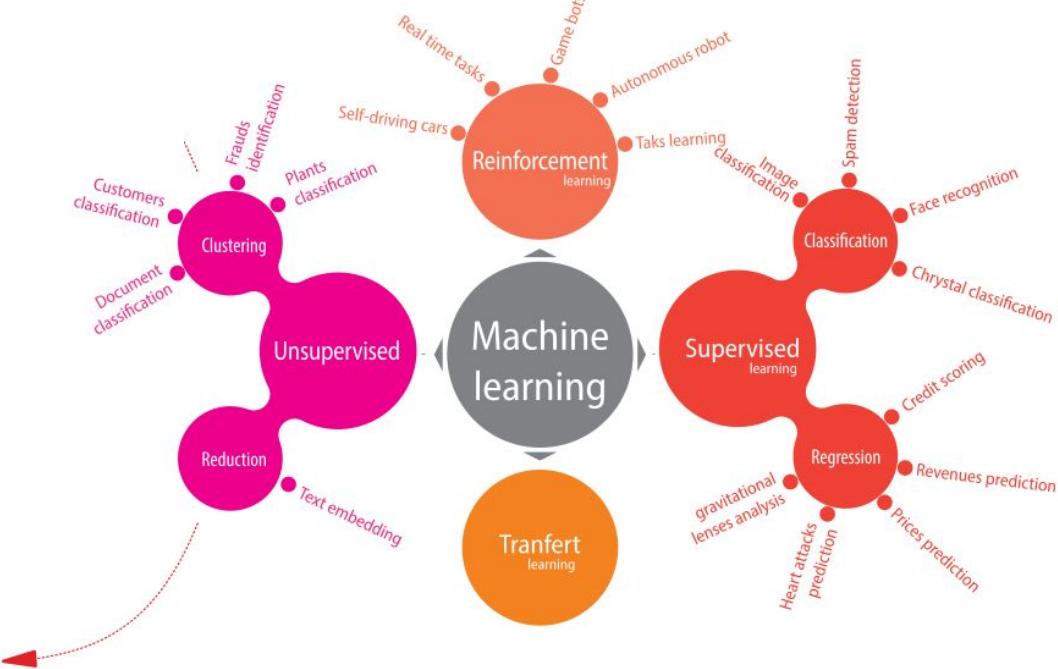
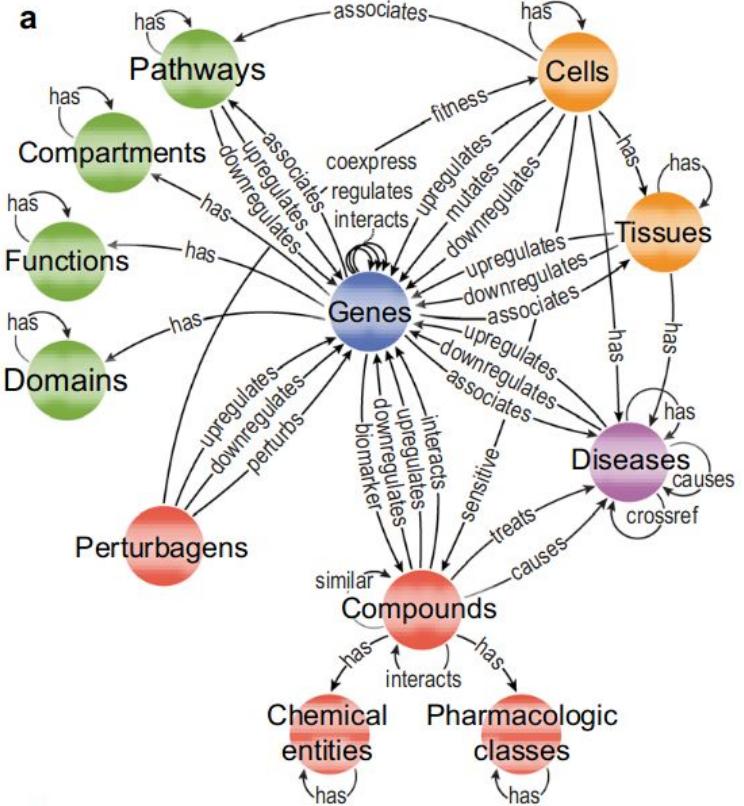
Simplification (for further downstream analysis)

Visualisation

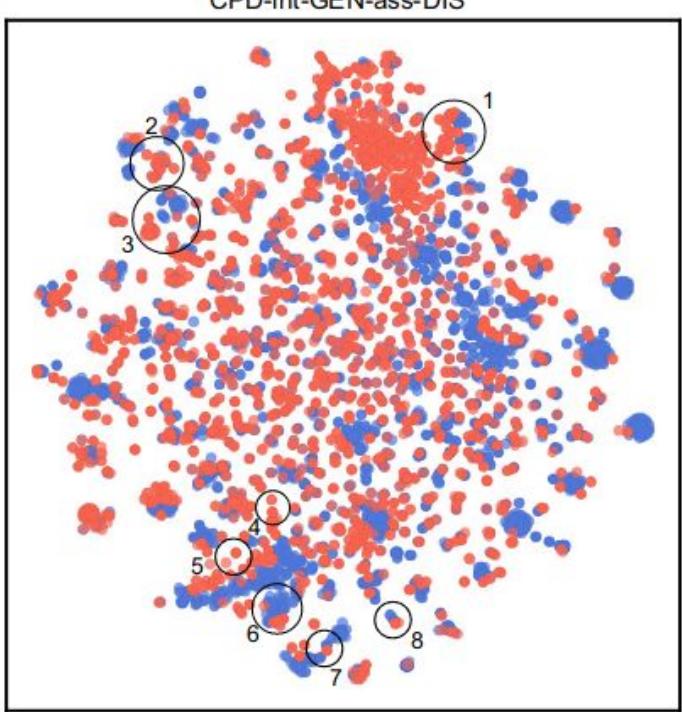
Embedding (networks, words, ...)



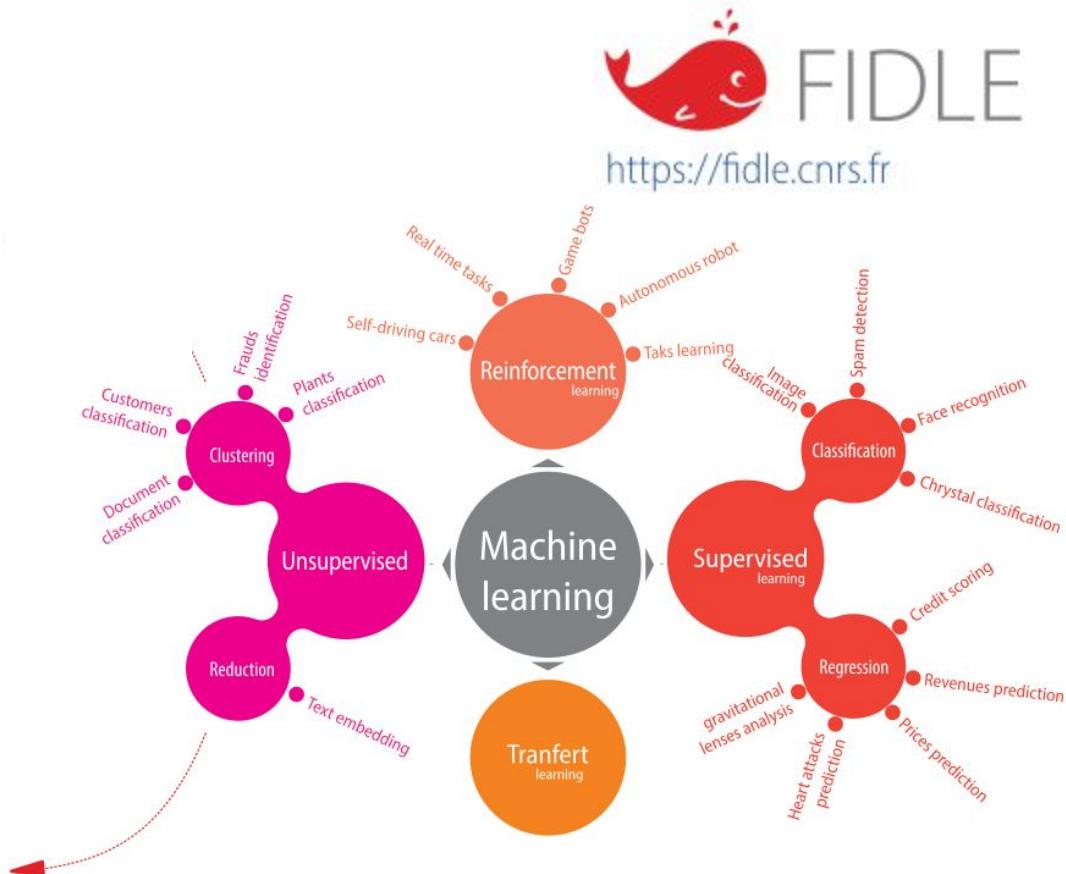
Nelson, Walter, et al. "To embed or not: network embedding as a paradigm in computational biology." *Frontiers in genetics* 10 (2019): 381.



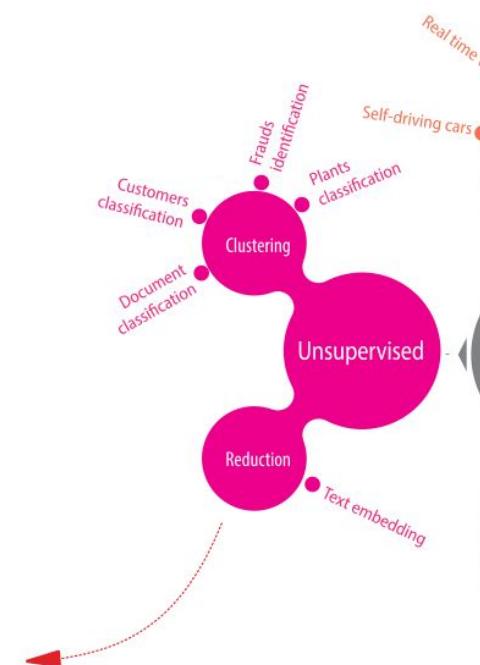
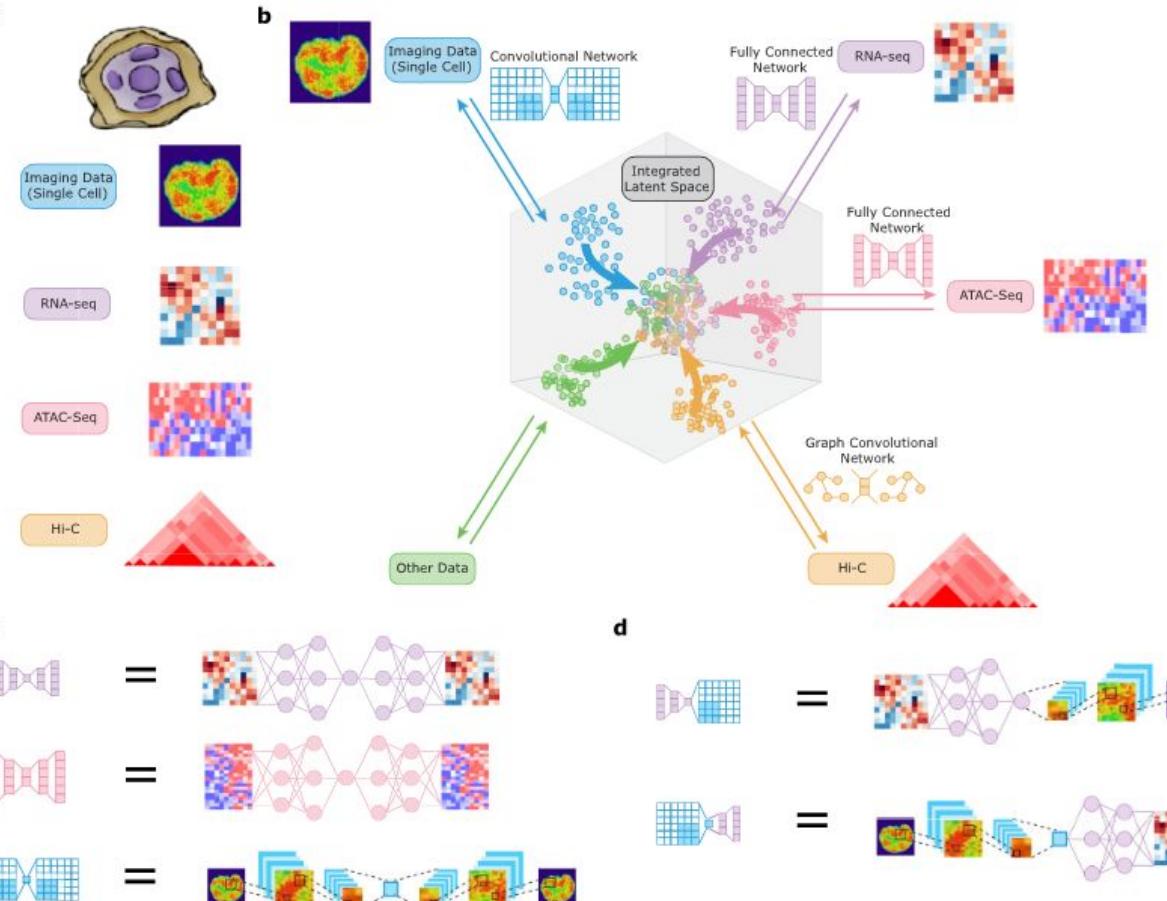
Fernández-Torras, Adrià, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. "Integrating and Formatting Biomedical Data as Pre-Calculated Knowledge Graph Embeddings in the Bioteque." *Nature Communications* 13, no. 1 (September 9, 2022): 5304. <https://doi.org/10.1038/s41467-022-33026-0>.

a

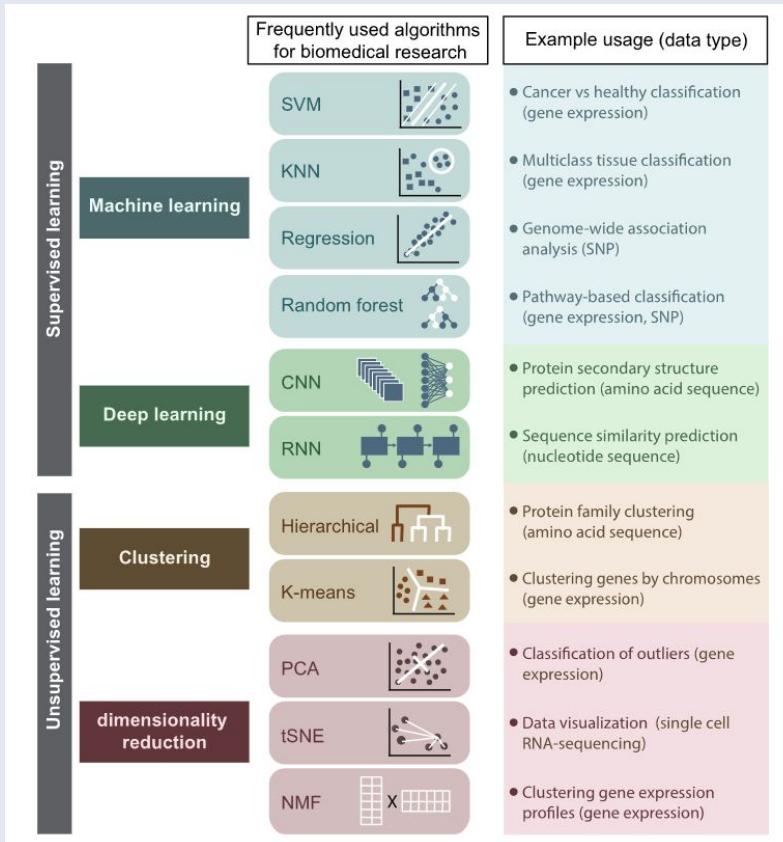
1. Etoposide, Daunorubicin - Leukemia, Kaposi's sarcoma (TOP2A, TOP2B)
2. Sermorelin, Testosterone - Hypogonadism, Hypopituitarism (GHRGR, NR3C2, AR)
3. Cortisone, Prednisolone - Dermatitis, Eosinophilia, Edema (NR3C1)
4. Somatostatin, Lanreotide - Hyperpituitarism, Acromegaly (SSTR1-5)



Fernández-Torras, Adrià, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. "Integrating and Formatting Biomedical Data as Pre-Calculated Knowledge Graph Embeddings in the Bioteque." *Nature Communications* 13, no. 1 (September 9, 2022): 5304. <https://doi.org/10.1038/s41467-022-33026-0>.



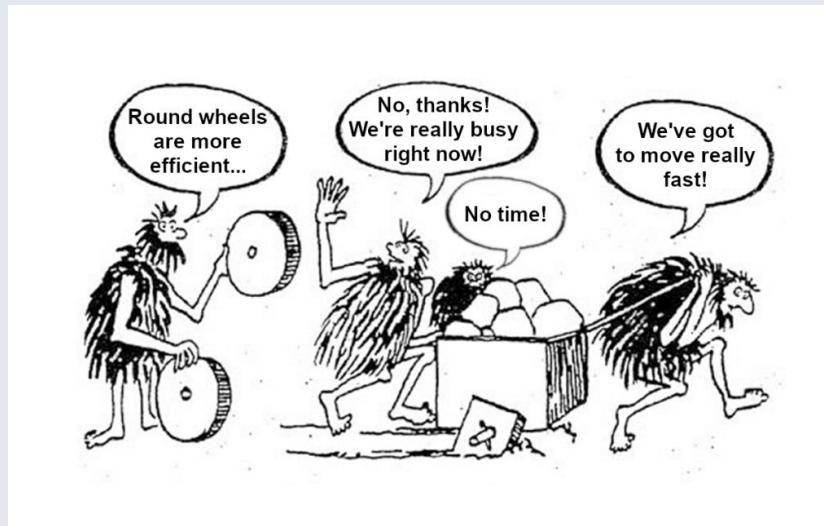
In a nutshell...



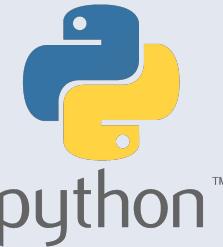
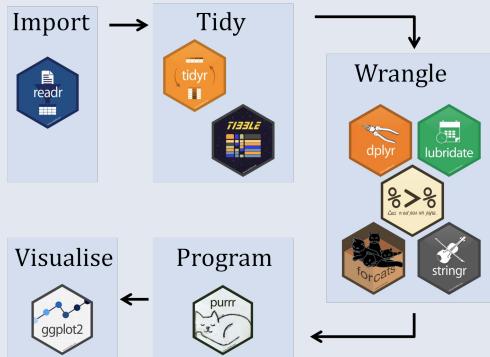
Auslander, Noam, Ayal B. Gussow, and Eugene V. Koonin. "Incorporating machine learning into established bioinformatics frameworks." *International Journal of Molecular Sciences* 22.6 (2021): 2903.

Tools

Languages and frameworks



Two main languages for data science, and a bunch of libraries

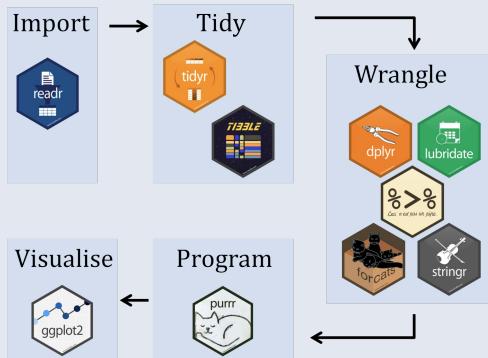


IP[y]: IPython
Interactive Computing



Tools

Two main languages for data science, and a bunch of libraries

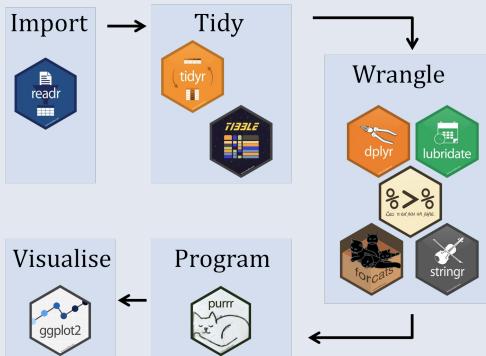


IP[y]: IPython
Interactive Computing



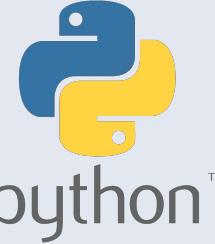
Tools

Two main languages for data science



of libraries

• • •

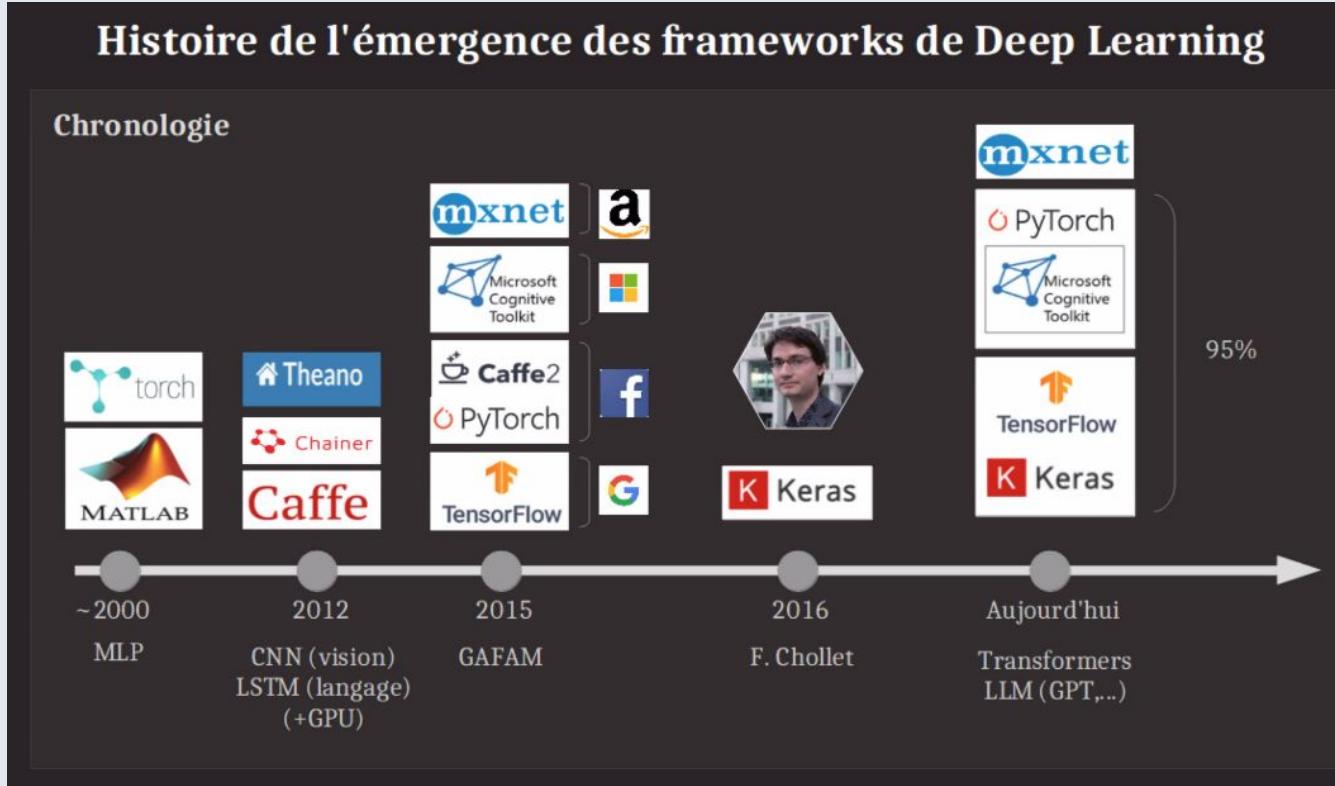


JumPy

IP[y]: IPython
Interactive Computing



Deep learning



Fabrice DAIAN

Deep learning

Comparaisons des frameworks de Deep Learning

	Langage	Accessibilité	Réputation	Vitesse
 TensorFlow	Python, C++, Java, Javascript, R	Difficile	Industriel	Rapide
 Keras	Python, R	Facile	Débutant	Lent
 PyTorch	Python	Moyenne	Académique	Rapide

Fabrice DAIAN

Future of AI

Current limitations, ethics, ...

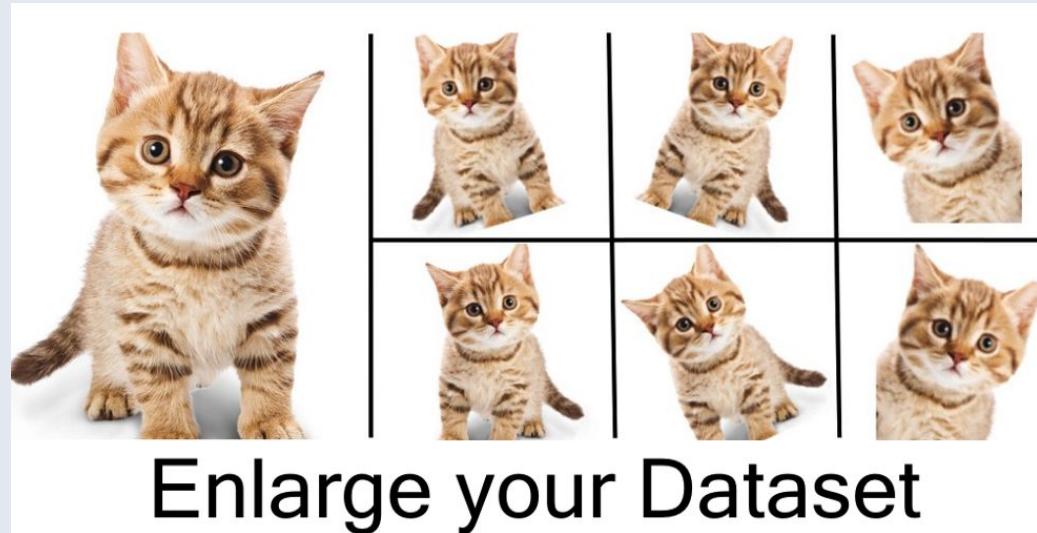
Current limitations in AI

- Data limitations



Current limitations in AI

- Data limitations
 - Quantity



Enlarge your Dataset

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity

Current limitations in AI

Color Matters in Computer Vision

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.



Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.



Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos.



Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of 296 photos.



Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Facial Recognition Is Accurate, if You're a White Guy

The New York Times

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity

How AI systems amplify bias

Image recognition systems that use biased machine learning data sets will inadvertently magnify that bias. Researchers are examining ways to reduce the effects.



COOKING

ROLE	►	VALUE
AGENT	►	WOMAN
FOOD	►	PASTA
HEAT	►	STOVE
TOOL	►	SPATULA
PLACE	►	KITCHEN



COOKING

ROLE	►	VALUE
AGENT	►	WOMAN
FOOD	►	FRUIT
HEAT	►	—
TOOL	►	KNIFE
PLACE	►	KITCHEN



COOKING

ROLE	►	VALUE
AGENT	►	WOMAN
FOOD	►	MEAT
HEAT	►	GRILL
TOOL	►	TONGS
PLACE	►	OUTSIDE



COOKING

ROLE	►	VALUE
AGENT	►	WOMAN
FOOD	►	VEGETABLES
HEAT	►	STOVE
TOOL	►	TONGS
PLACE	►	KITCHEN



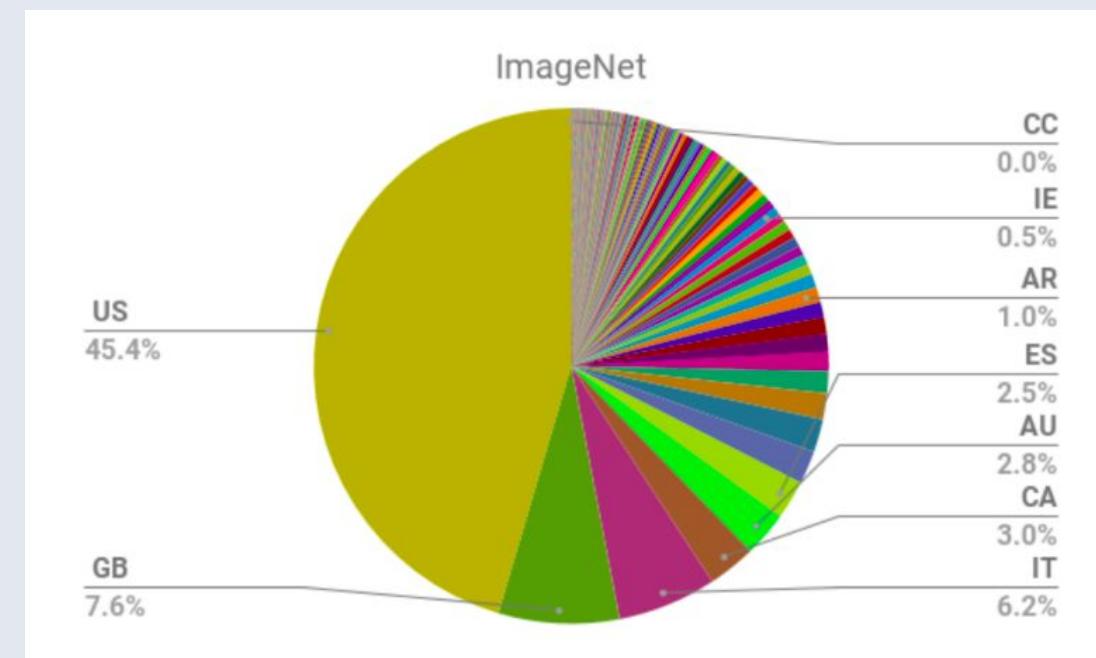
COOKING

ROLE	►	VALUE
AGENT	►	MAN
FOOD	►	—
HEAT	►	STOVE
TOOL	►	SPATULA
PLACE	►	KITCHEN

In this example of gender bias, adapted from a report published by researchers from the University of Virginia and the University of Washington, a visual semantic role labeling system has learned to identify a person cooking as female, even when the image is male.

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity

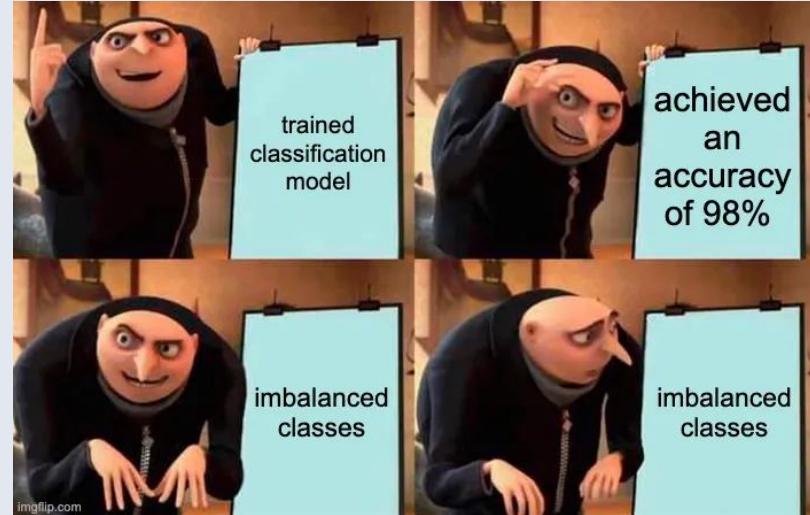


Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - “Naturally” imbalanced datasets
 - disease diagnosis
 - fraud detection
 - anomaly detection
 - ...

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - “Naturally” imbalanced datasets
 - disease diagnosis
 - fraud detection
 - anomaly detection
 - ...

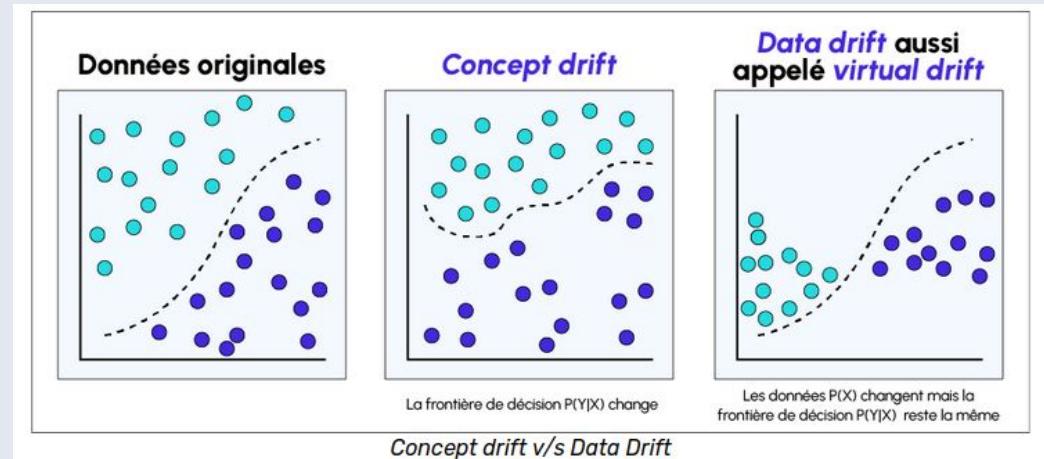


Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability

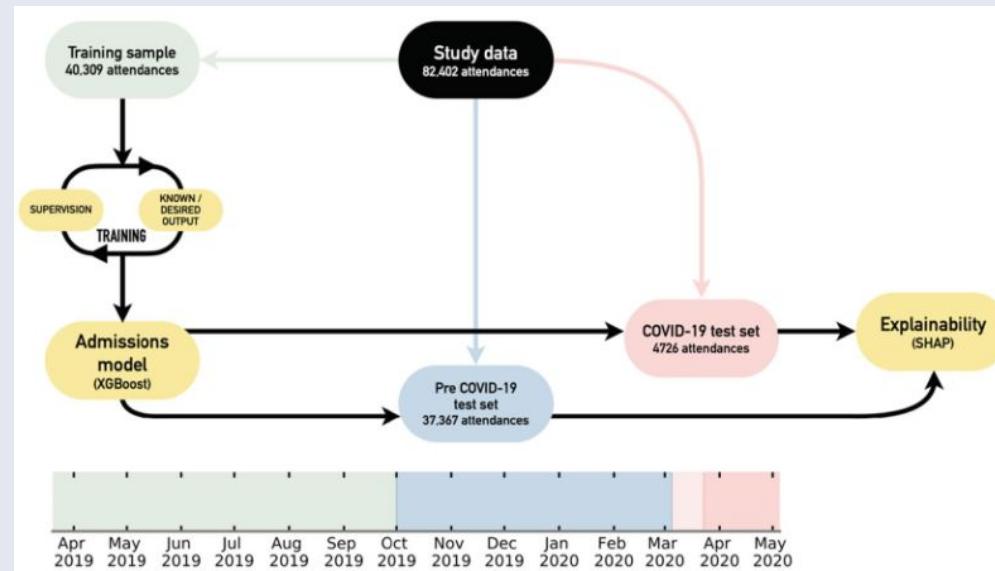
Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - e.g. seasonality or...



Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - e.g. seasonality or... Covid19



Duckworth, C., Chmiel, F.P., Burns, D.K. et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. *Sci Rep* 11, 23017 (2021). <https://doi.org/10.1038/s41598-021-02481-y>

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
 - Sensitive data, consent of use, ...

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
 - Sensitive data, consent of use, ...

2 minute read · April 19, 2023 6:56 PM GMT+2 · Last Updated 21 days ago

Exclusive: German authors, performers call for tougher ChatGPT rules amid copyright concerns

By Foo Yun Chee ▾

ChatGPT face au défi de la propriété intellectuelle



Antoine Crochet-Damais
JDN

Mis à jour le 03/03/2023 14:56

ChatGPT: médias et artistes veulent faire valoir leurs droits d'auteur

Par Chloé Woitier

Publié le 03/03/2023 à 20:08 , mis à jour le 03/03/2023 à 20:08

Publié le 03 mai 2023 à 11h53

OpenAI a instruit ChatGPT sans se soucier du droit d'auteur

Le copyright bientôt obsolète ?

Current limitations in AI

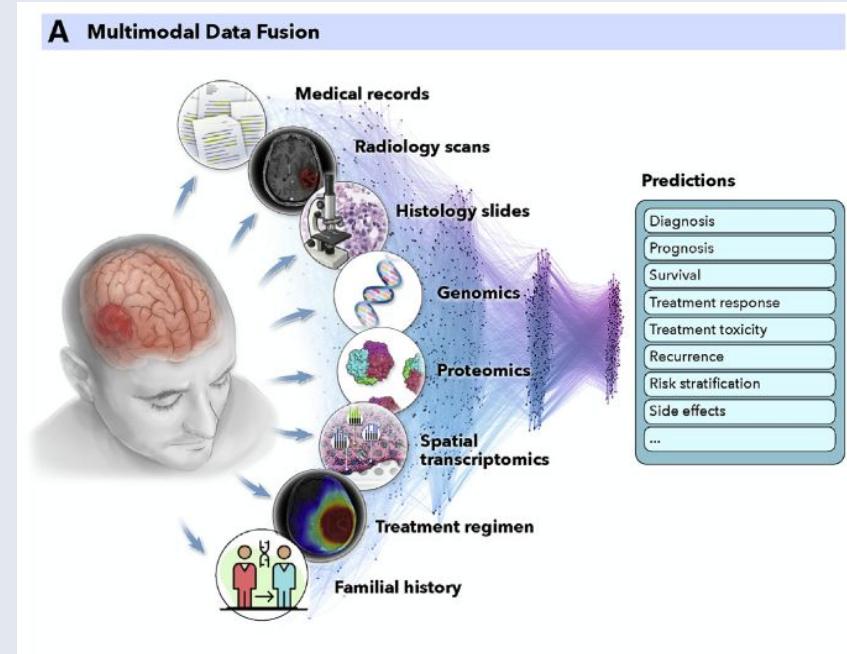
- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
 - Sensitive data, consent of use, ...

Current big challenge: Multimodal
data integration

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
 - Sensitive data, consent of use, ...

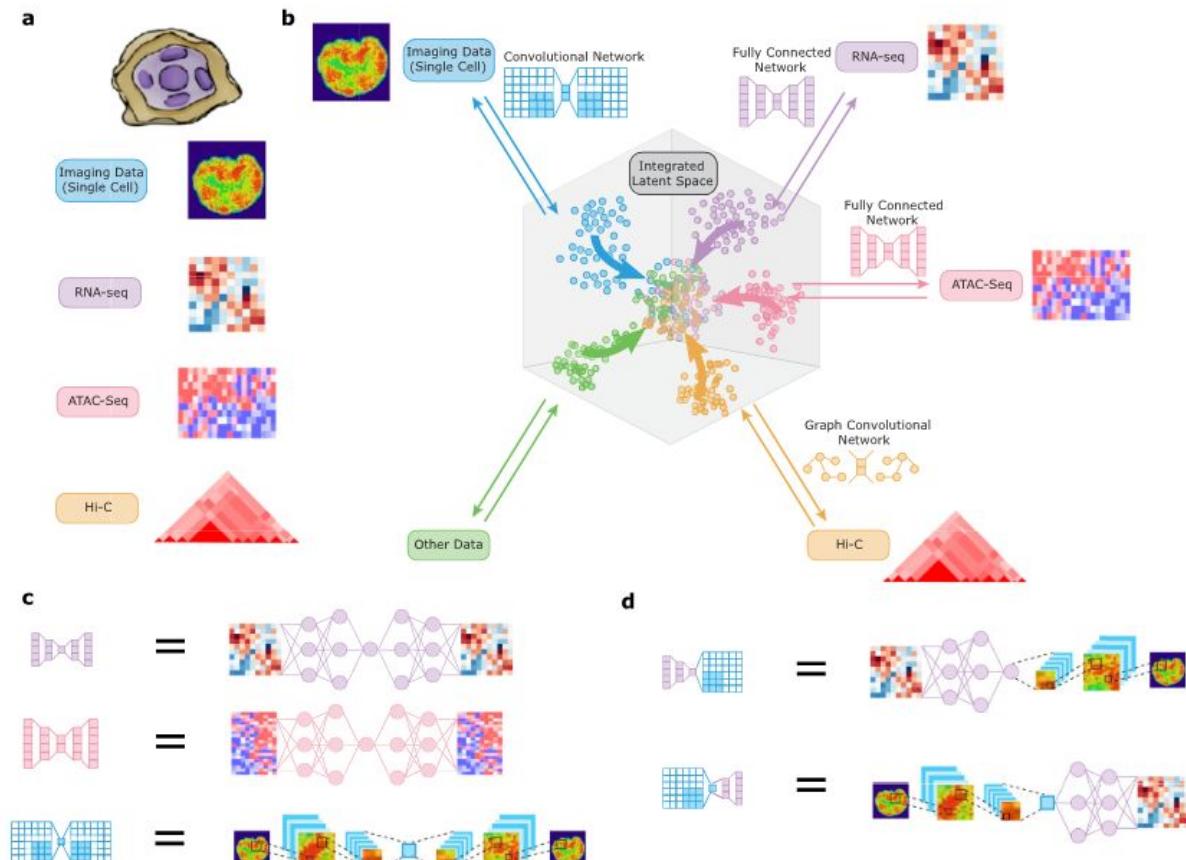
Current big challenge: Multimodal data integration



Lipkova, Jana, et al. "Artificial intelligence for multimodal data integration in oncology." *Cancer Cell* 40.10 (2022): 1095-1110.

Current limitations

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
 - Sensitive data, c
 - Multimodal data integ



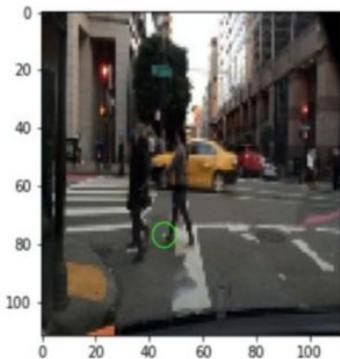
Yang KD, et al. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. Nat Commun. 2021 Jan 4;12(1):31. doi: 10.1038/s41467-020-20249-2. PMID: 33397893; PMCID: PMC7782789.

Current limitations in AI

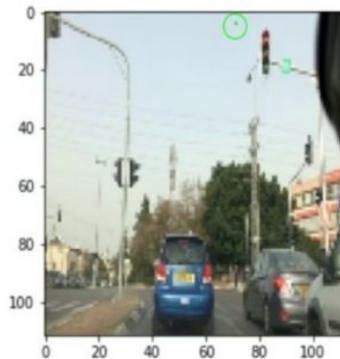
- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise

Current limitations in AI

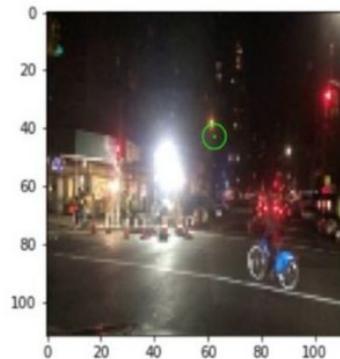
- Data limitations
 - Quantity
- Algorithmic limitations
 - Sensitivity to noise



(a)



(b)



(c)

Red light classified as green with (a) 68%, (b) 95%, (c) 78% confidence after one pixel change.

– TACAS 2018, <https://arxiv.org/abs/1710.07859>

Current limitations in AI

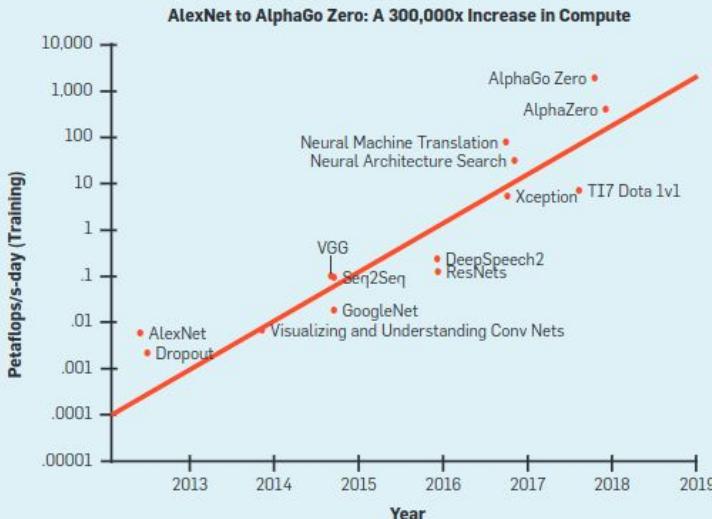
- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption

Current limitations in AI

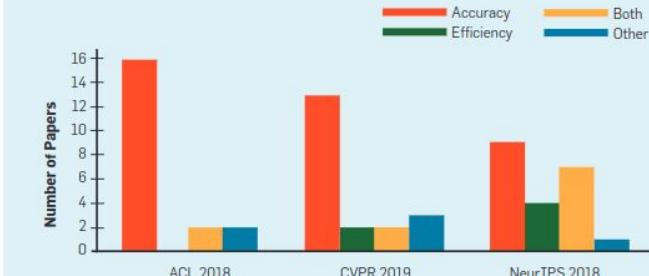
Figure 1. The amount of compute used to train deep learning models has increased 300,000x in six years. Figure taken from Amodei et al.²



➤ Algorithmic limitations

- Sensitivity to noise
- Cost and energy consumption

Figure 2. AI papers tend to target accuracy rather than efficiency. The figure shows the proportion of papers that target accuracy, efficiency, both or other from a random sample of 60 papers from top AI conferences.



Schwartz, Roy, et al. "Green ai." *Communications of the ACM* 63.12 (2020): 54-63.

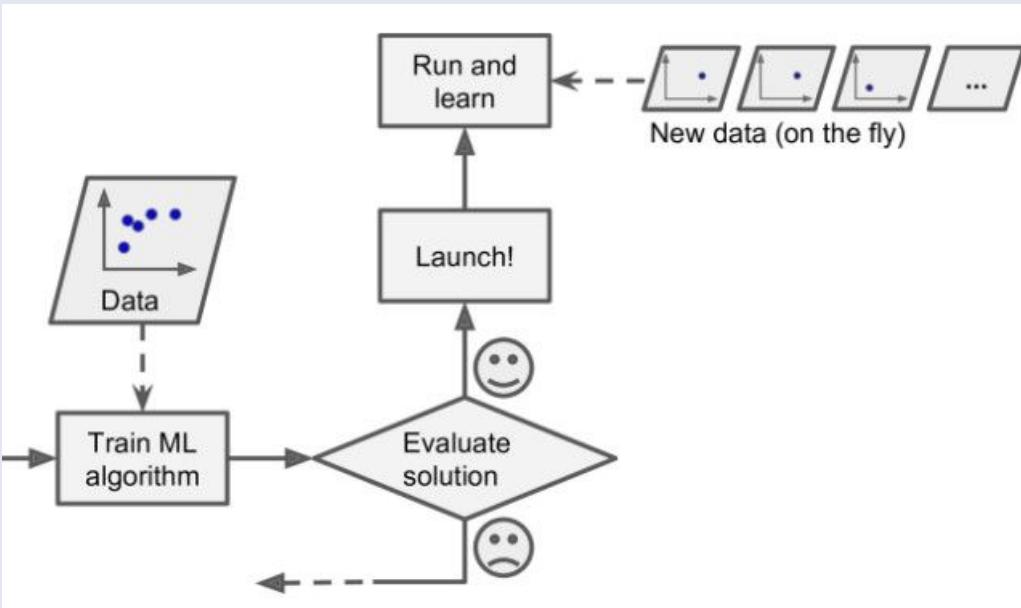
Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability
 - Real-time constraints

Current limitations in AI



Algorithmic limitations

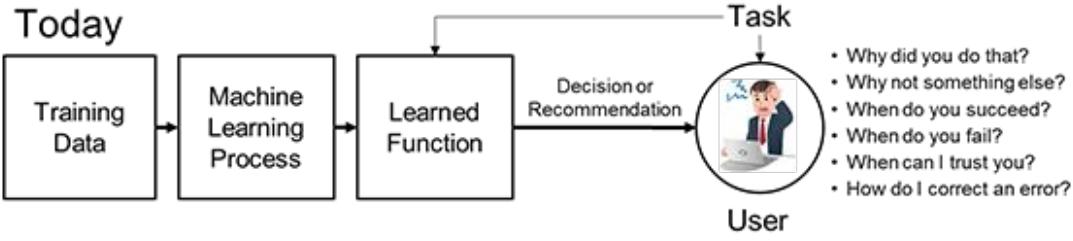
- Sensitivity to noise
- Cost and energy consumption
- Scalability
- Real-time constraints

<https://vitalflux.com/difference-between-online-batch-learning/>

Current limitations in AI

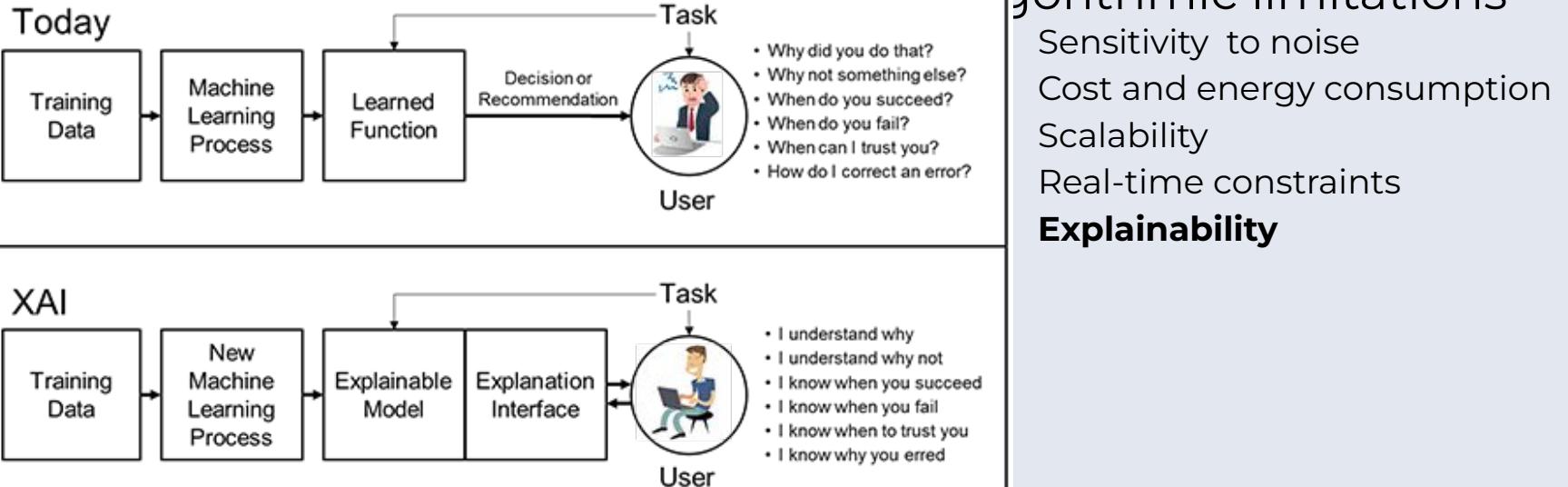
- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability
 - Real-time constraints
 - **Explainability**

Current limitations in AI



gorithmic limitations
Sensitivity to noise
Cost and energy consumption
Scalability
Real-time constraints
Explainability

Current limitations in AI



gorithmic limitations

Sensitivity to noise

Cost and energy consumption

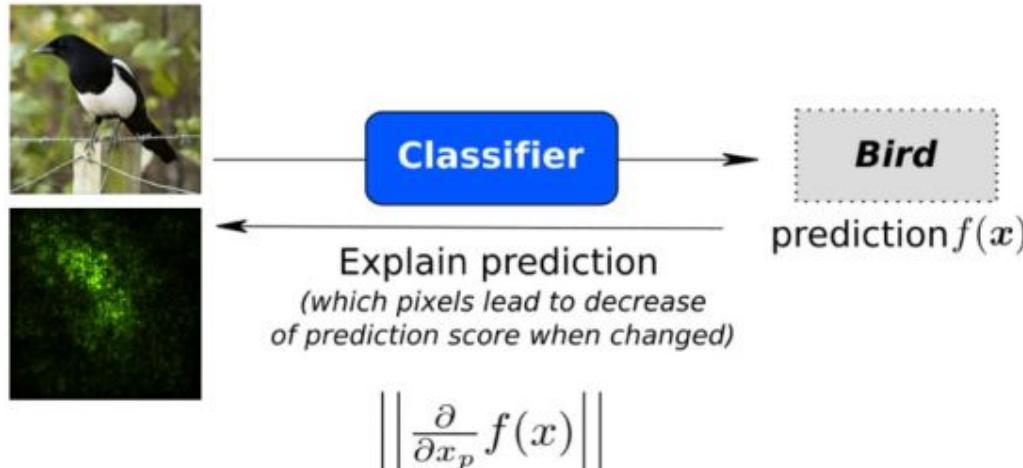
Scalability

Real-time constraints

Explainability

Current limitations in AI

Sensitivity Analysis
(Simonyan et al. 2014)



orithmic limitations
Sensitivity to noise
Cost and energy consumption
Scalability
Real-time constraints
Explainability

Current limitations in AI

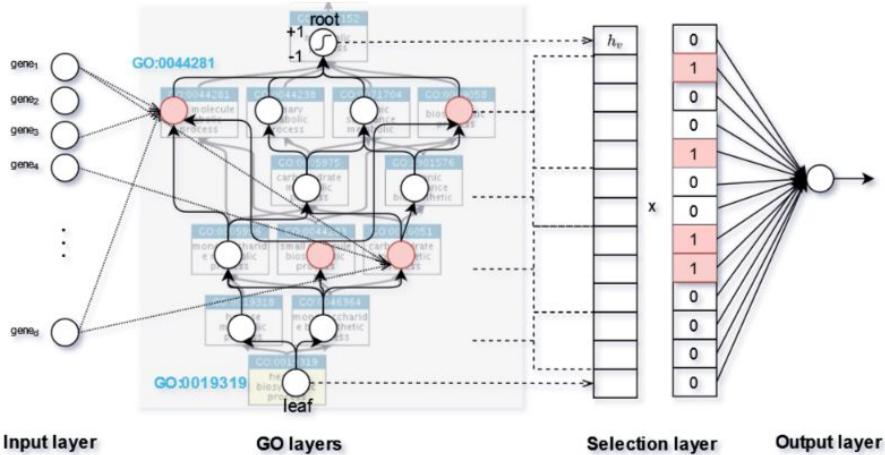
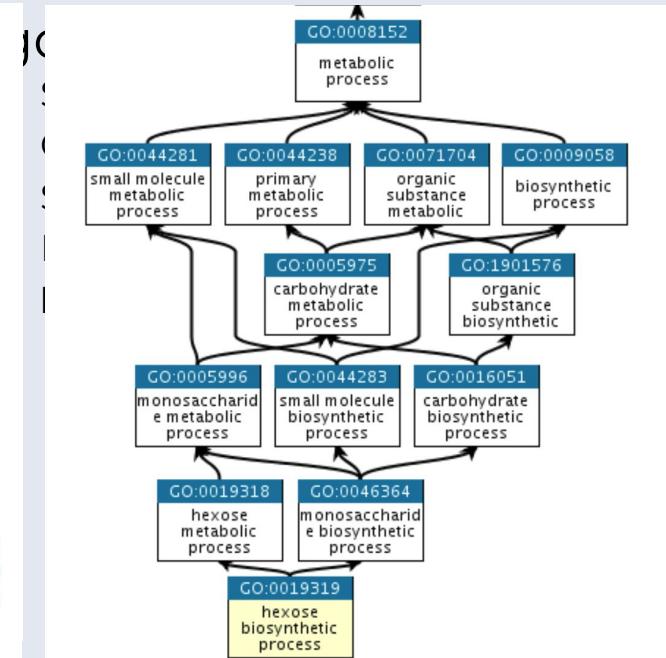


Figure 1: Illustration of GraphGONet. The neurons in the input layer receive the signal from the genes. The dotted arrows correspond to the connections between the genes and the GO terms represented by neurons in the hidden layers. The relations between GO terms are represented by the plain arrows. The dashed arrows depict the concatenation of the activation of the neurons. The selection layer results from the concatenation and masking operations.

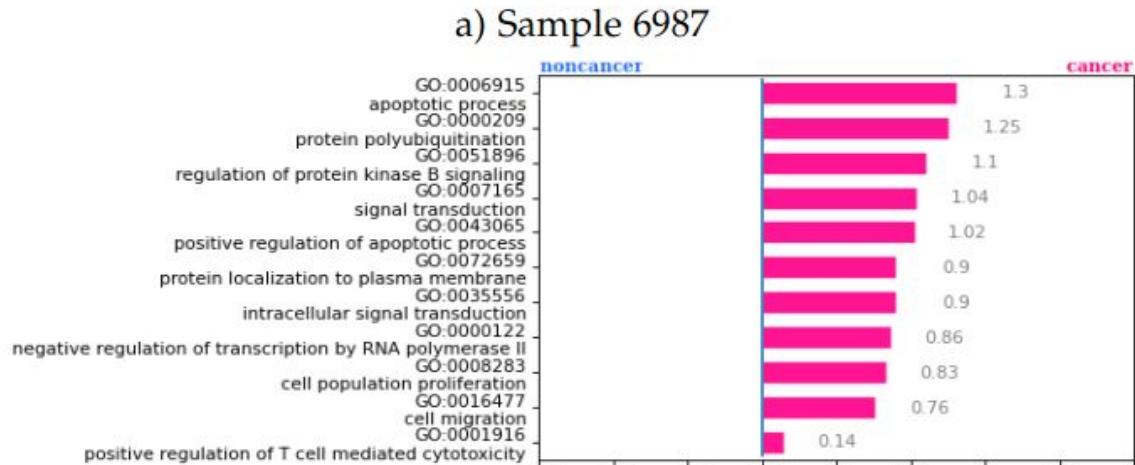


Victoria Bourgeais, Farida Zehraoui, Blaise Hanczar. GraphGONet: a self-explaining neural network encapsulating the Gene Ontology graph for phenotype prediction on gene expression. Bioinformatics, 2022, 38 (9), pp.2504-2511.            

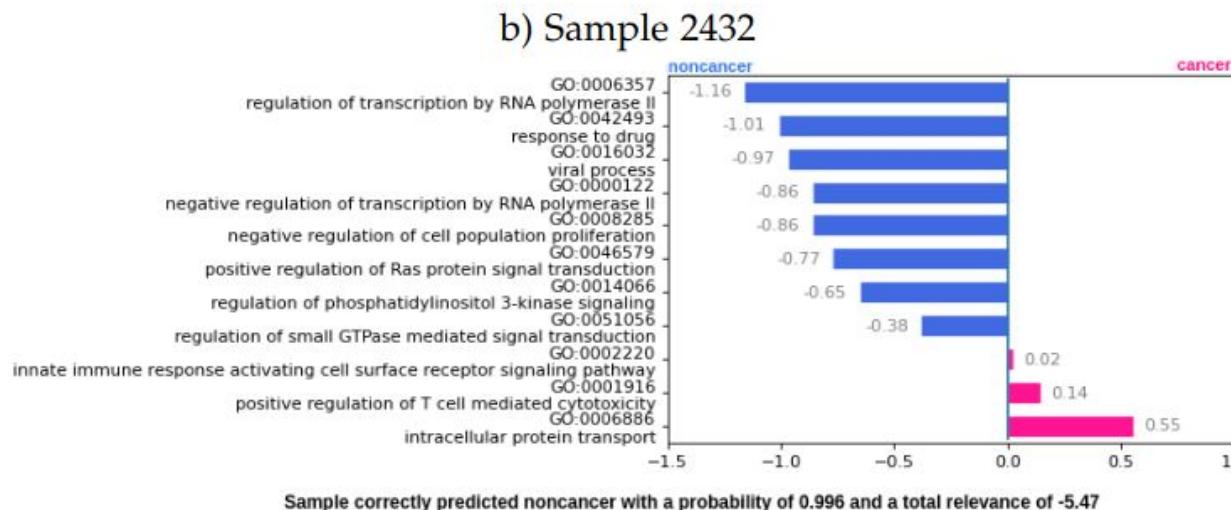
Current



Figure 1: Illustration of Graph the connections between the graph represented by the plain arrow from the concatenation and map



Sample correctly predicted cancer with a probability of 1 and a total relevance score of 10.5.



Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability
 - Real-time constraints
 - Explainability

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability
 - Real-time constraints
 - Explainability
- Ethical concerns
 - Threats, risks, harms and wrongs
 - Responsibility



A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework by Karen Yeung

- Algorithmic decision-making systems that rely on data-driven profiling techniques **may threaten several human rights**:
 - Rights to a fair trial and to 'due process'
 - Rights to freedom of expression and information
 - Rights to privacy and data protection
 - Rights to protection against discrimination. Biases can be induced:
 - by the developers
 - by the model itself
 - by the training data
 - when the system is applied in real world settings
- + **Power asymmetry** between those who develop and those who interact/are subjected to AI technologies due to **opacity** of AI technologies

"These collective threats and risks are **exacerbated by the capacity of these technologies to operate at unprecedented speed and scale**, generating **novel threats, risks and challenges** which contemporary societies have not historically had to confront."





- Who bears **responsibility** ?
 - “Because AI systems can operate in time and space in new and unprecedented ways, **these technologies may challenge our existing conceptions of responsibility.**”
 - The “many hands” problem
 - Human-computer interactions ('human in the loop')/'moral crumple zones' - 'zones de déformation morale')
 - Unpredictable nature of interactions between multiple algorithmic systems ('flash crash' of 2010)

Current limitations in AI

- Data limitations
 - Quantity
 - Diversity
 - Availability
 - Data drift
 - Ethics considerations
- Algorithmic limitations
 - Sensitivity to noise
 - Cost and energy consumption
 - Scalability
 - Real-time constraints
 - Explainability
- Ethical concerns
 - Threats, risks, harms and wrongs
 - Responsibility

