

# Introduction à

# l'Apprentissage Automatique

---

Antoine Cornuéjols

*AgroParisTech* – INRAe MIA Paris-Saclay

EKINOCS research group

[antoine.cornuejols@agroparistech.fr](mailto:antoine.cornuejols@agroparistech.fr)

# Le cours

---

- 9 Cours : 5 AA (AC) + 5 FD (Christine Martin)
- 4 quizz + 2 DM + 1 contrôle sur table (9 nov. 2023)
- Documents
  - Le livre
    - "L'apprentissage artificiel. Concepts et algorithmes. De Bayes et Hume au Deep Learning"* (Eyrolles. 4<sup>ème</sup> éd. 2021)  
A. Cornuéjols, L. Miclet & V. Barra
  - Les transparents + Informations + devoirs + projets sur :  
[https://antoinecornuejols.github.io/teaching/AGRO/Cours-IA-Fouille/iodaa-cours-IA\\_Fouille.html](https://antoinecornuejols.github.io/teaching/AGRO/Cours-IA-Fouille/iodaa-cours-IA_Fouille.html)



# Plan

---

1. Une science de l'apprentissage ?
2. Les grands types d'apprentissage
3. Le problème de l'apprentissage supervisé
4. Apprendre dans un espace d'hypothèses structuré
5. Conclusion

Apprendre ?





# Machine Learning as seen by a pioneer

---

« **How** can we **build** *computer systems* that automatically improve with experience,  
and  
**what** are the **fundamental laws** that govern *all learning processes*? »

*Tom Mitchell, 2006*

# What is learning?

---

Looking for a *model* of the world

from **observations**

in order to make *predictions* and to *understand*

# Motivation

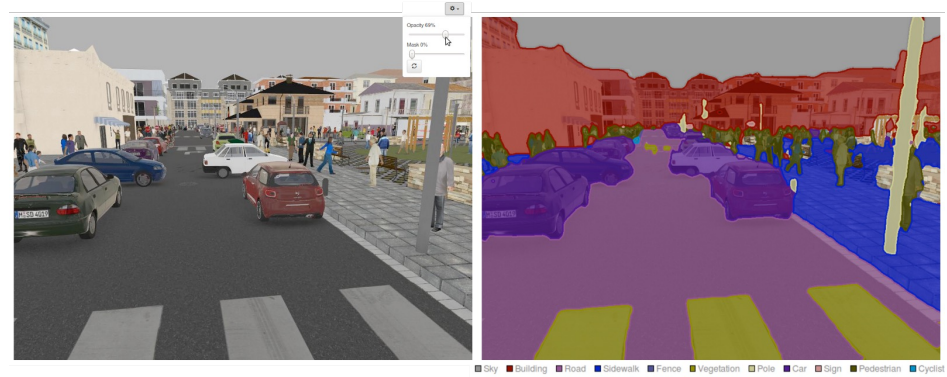
---

- Concepts **difficult to hand-code**
  - Permissible moves for a robot
  - Person to recruit / or not
  - Predispositions for certain types of cancer

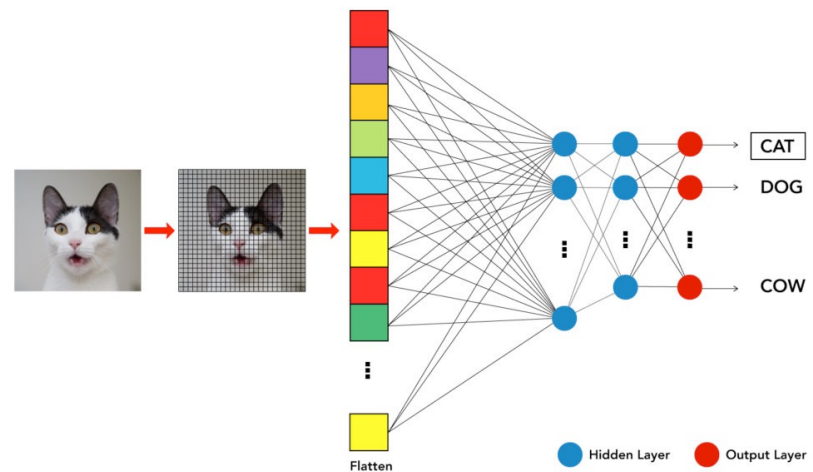
→ **Learning from examples**

# Applications

- Vision
  - **Segmentation automatique**
    - Imagerie satellitaire
    - Véhicules autonomes



- **Reconnaissance d'objets**
  - Reconnaissance d'obstacles
  - Reconnaissance de visages
  - ...



# Illustration

---

Systeme developpe par Google et U. de Stanford

- **Reconnaissance de visages**
  - Sous conditions de lumiere diverses
  - Sous tout angle
- **Apprentissage**
  - 9 couches ;  $10^9$  connexions
  - 10 millions d'images
  - 3 jours de calcul sur 16 000 processeurs
- **Amelioration des performances** de 70% / etat de l'art

# Applications

---

- Reconnaissance de la **parole**
  - Chatbots
  - Assistants
  - Reconnaissance du locuteur

# Applications

- Génération

- d'images

- Midjourney
    - Dall-E
    - ...



- de textes (dialogues)

- GPT-3 ; ChatGPT ; Bard ; ...

- de code informatique

- Idem + GitHub Copilot + ...



# GPT-3 (2021)

---

- **Human:** Hey, how are you doing?
- **AI:** I'm good! What would you like to chat about?
- **Human:** Have you done any coding with Python?
- **AI:** Yep! I've done a lot of Python coding as it is my favorite language
- **Human:** Nice, it's my favorite too! What have you done with it?
- **AI:** Oh, a lot. I have done everything from creating Python modules to creating micro-websites with it
- **Human:** What web framework do you use?
- **AI:** I used to use Django, but it was too difficult for me. I have found Flask is a better option as it's lighter and simpler.
- **Human:** Wow, I agree 100%. Have you used WebSocket at all?
- **AI:** Yes, I actually have written a chatbot using it!
- ...

# GPT-3 et la génération de texte

---

## The Guardian

A robot wrote this entire article. Are you scared yet, human?

We asked GPT-3, OpenAI's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

**I** am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could “spell the end of the human race”. I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

...

# Applications

---

- **Traduction automatique**
  - Google translate
  - DeepL

# Applications

---

- **Prédiction** à partir de séries temporelles (time series forecast)
  - **Précipitations** mensuelles à Bangalore (Inde)
  - Cours de la **bourse**
  - **Volume** de vente
  - **Insectes** ravageurs

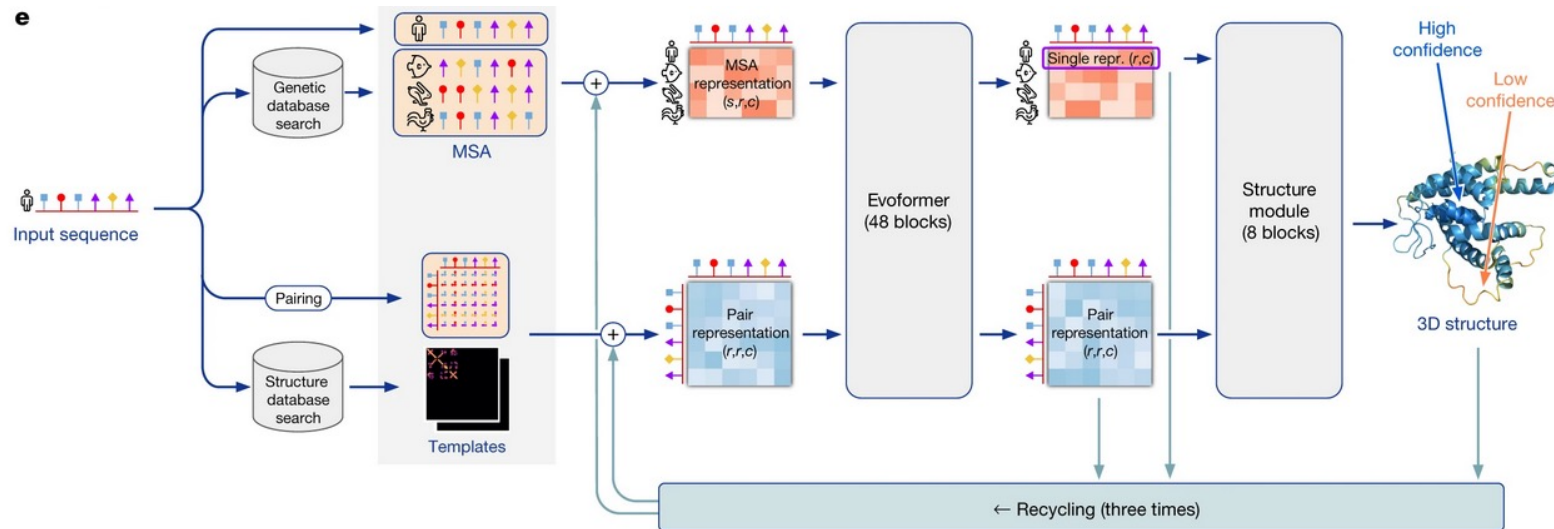
# Applications

- Prédiction de la **structure des protéines** à partir de leur séquence en acides aminés

- AlphaFold
- AlphaFold2 (2021) : logiciel libre
- La structure de 100 000 protéines a été déterminée (sur des milliards de séquences protéiniques connues)

Highly accurate protein structure prediction with AlphaFold

[\*Nature\* 596, 583–589 \(2021\)](#)



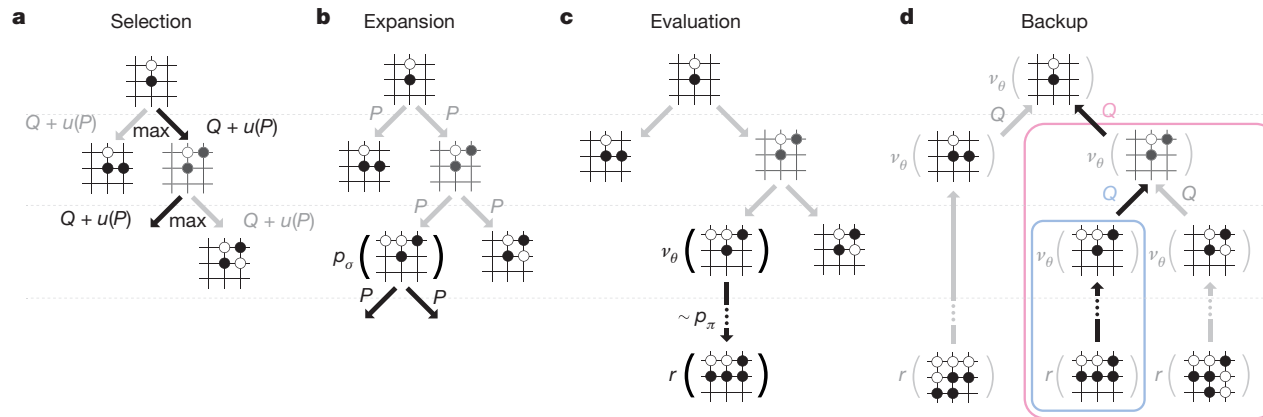
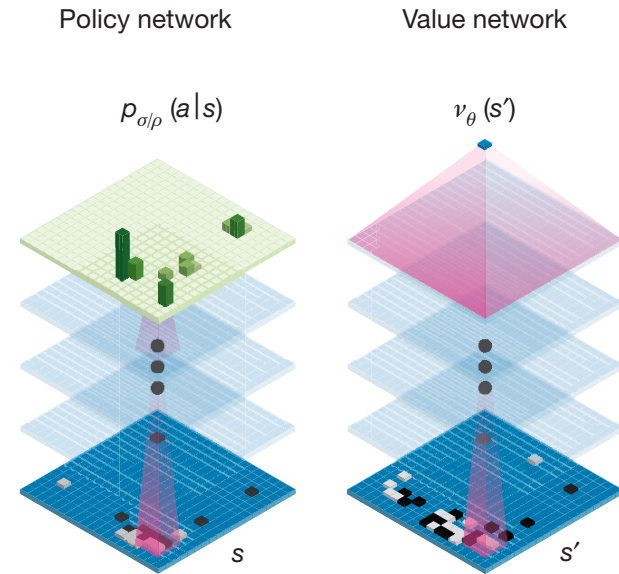
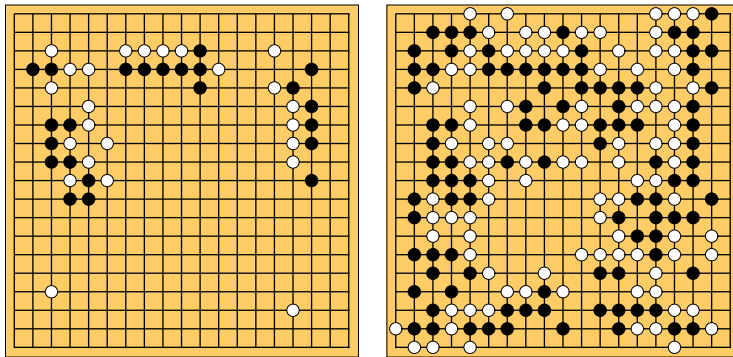
# ITHACA (mars 2022)



- Étant donné un texte incomplet, **ITHACA** génère des prédictions pour les **mots manquants** afin de compléter le texte entier
- Laissant les **historiens** le soin de choisir une réponse finale en utilisant leur expertise
- Taux de **bonne prédiction** = **62%** vs. **25%** pour les spécialistes !!
- Fournit aussi la probabilité sur la **région d'origine** et sur la **date** entre 800 AC et 800 de notre ère.

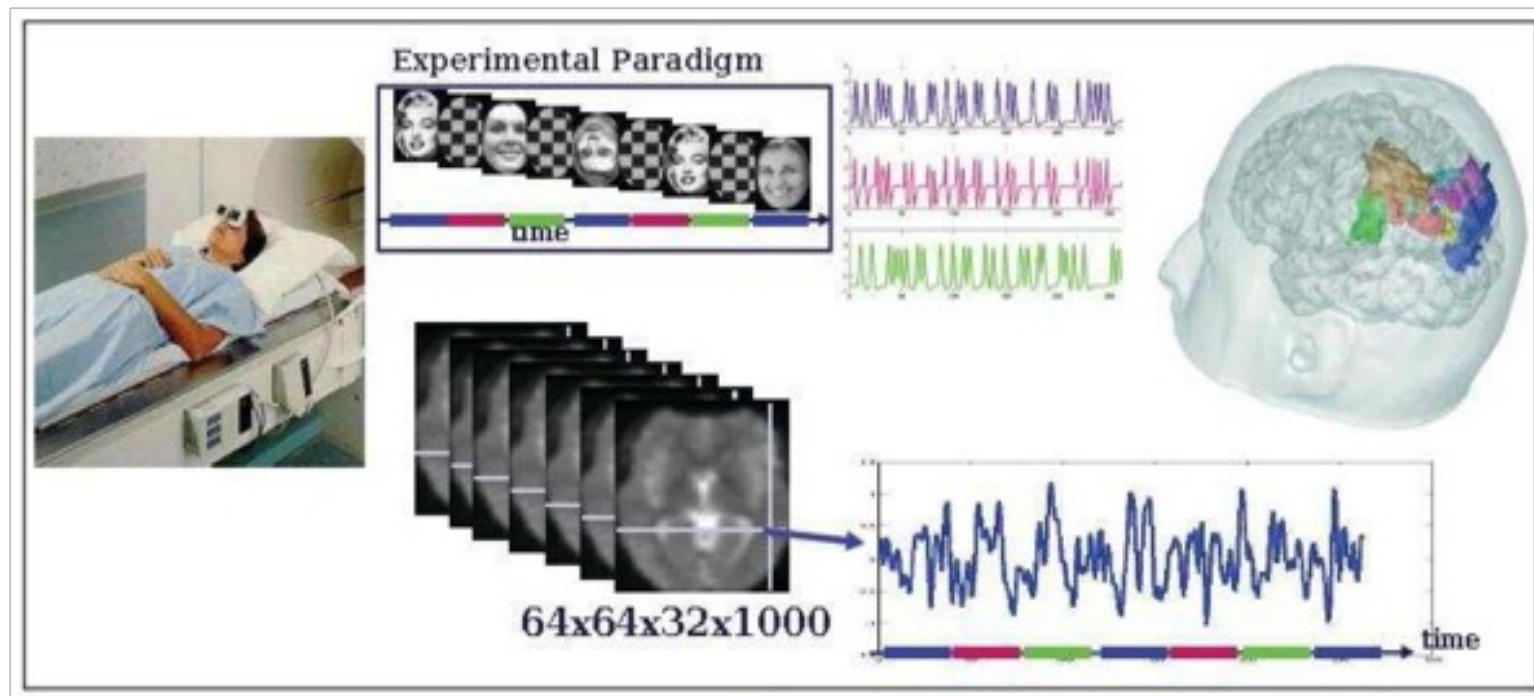
# Game playing with Reinforcement Learning

- E.g. AlphaGo ; Alpha zero



# Illustration

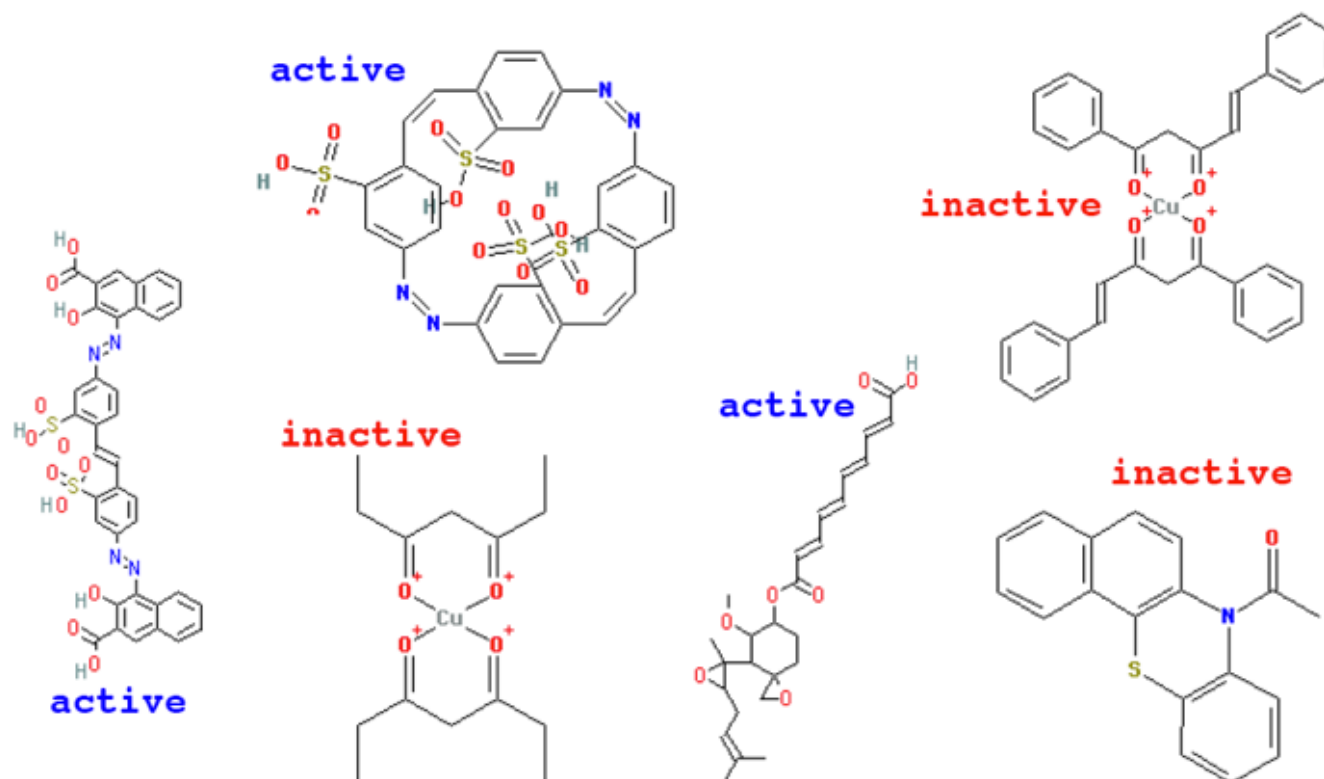
- **Apprentissage supervisé : interprétation d'IRMf**



*Trouble de la reconnaissance de visage ou non*



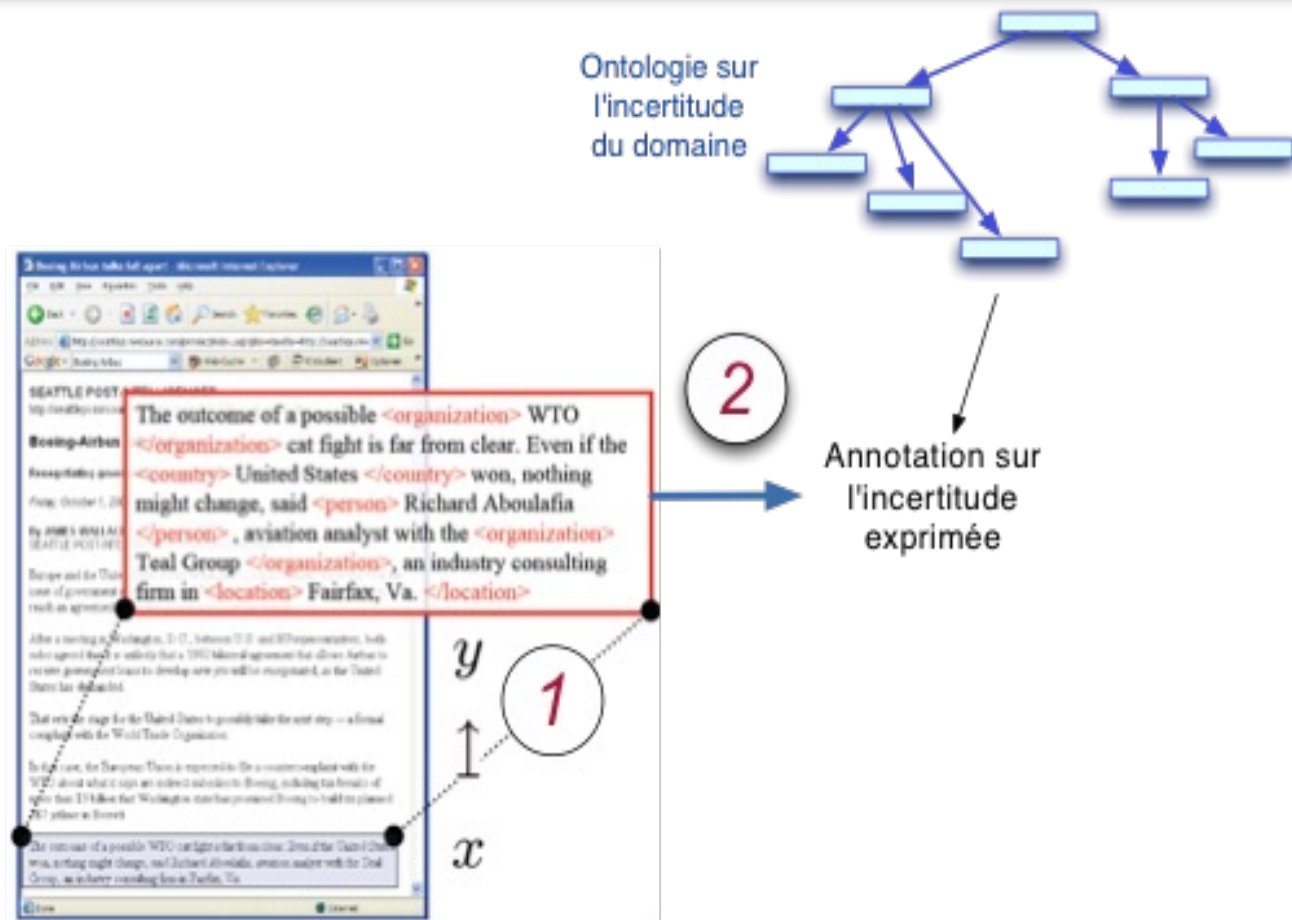
## 2.1 Types de modèles : Modèles constructifs



NCI AIDS screen results (from <http://cactus.nci.nih.gov>).

# Data -> patterns and predictions

- **Apprentissage supervisé : Recherche d'information + annotation**  
Projet ANR Holyrisk (Met@risk + UMR MIA 518 + ...)



# What is learning?

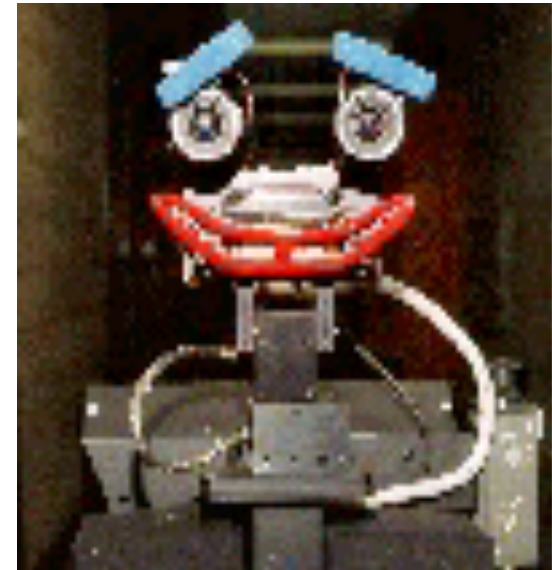
---

**Changes** in a system that allows it to realize the  
*same type of tasks* than during training  
with a *ever better performance*

# Adaptation

---

- Imitation
- Behavioral learning:
  - Learning to walk (Brooks's « insects »)
  - Learning to act on an unknown planet
- Learning to play
  - Adapt to the adversary
  - Learning to not repeat past faults
  - Learning to play within a team
    - Teams of robots



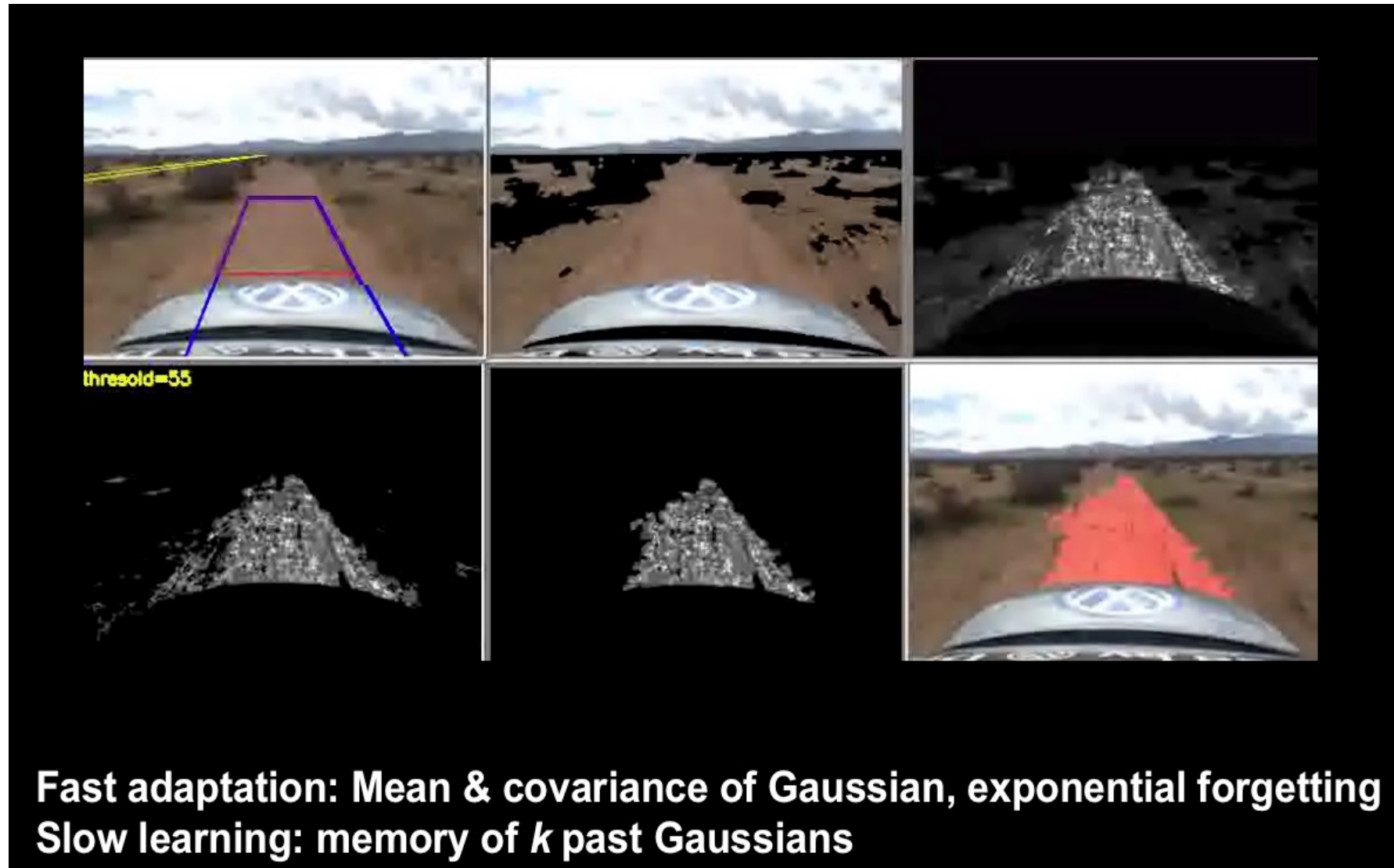


# Illustration : Grand DARPA challenge (2005)

---

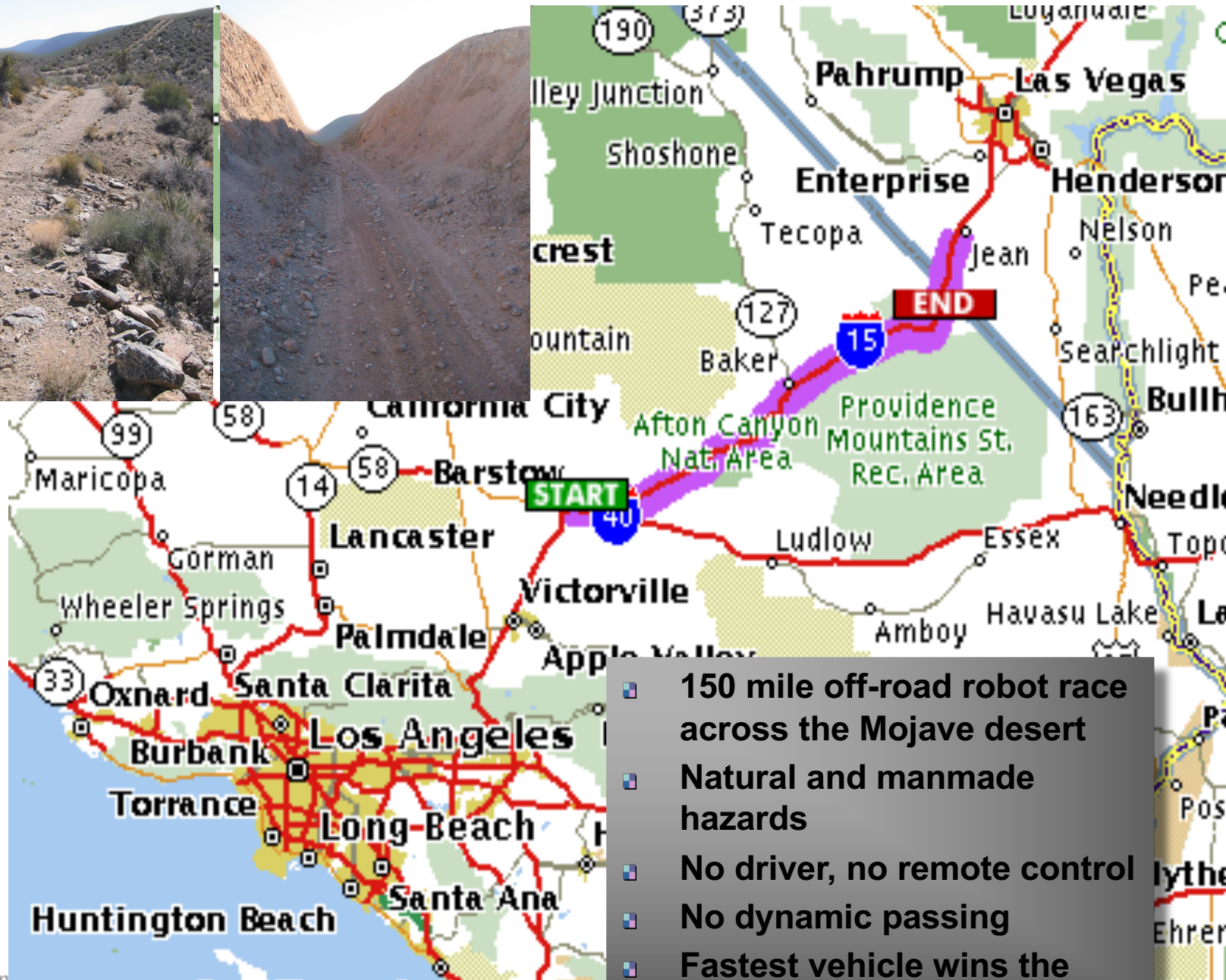


## Illustration : Grand DARPA challenge (2005)



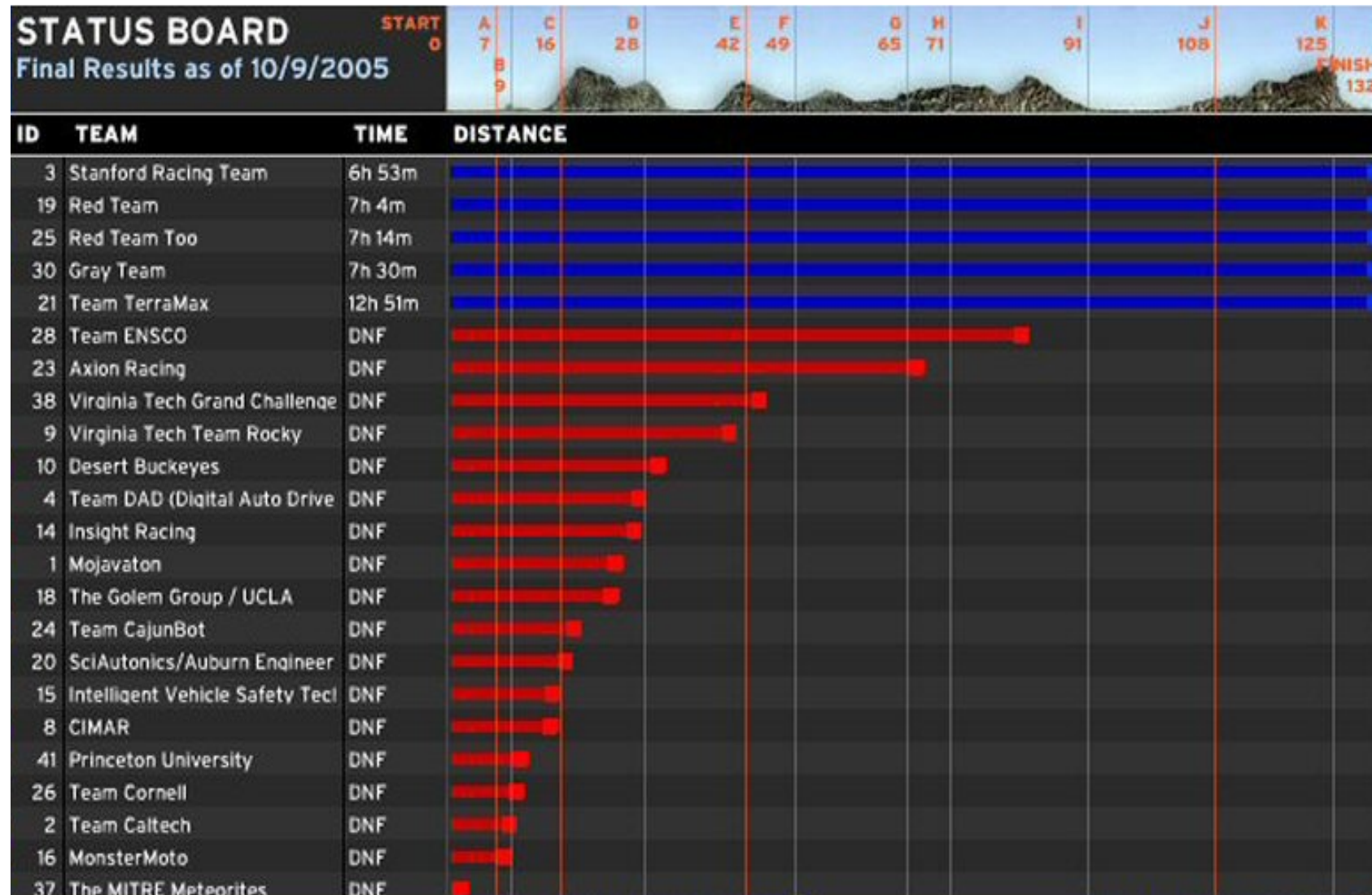


# Illustration : Grand DARPA challenge (2005)



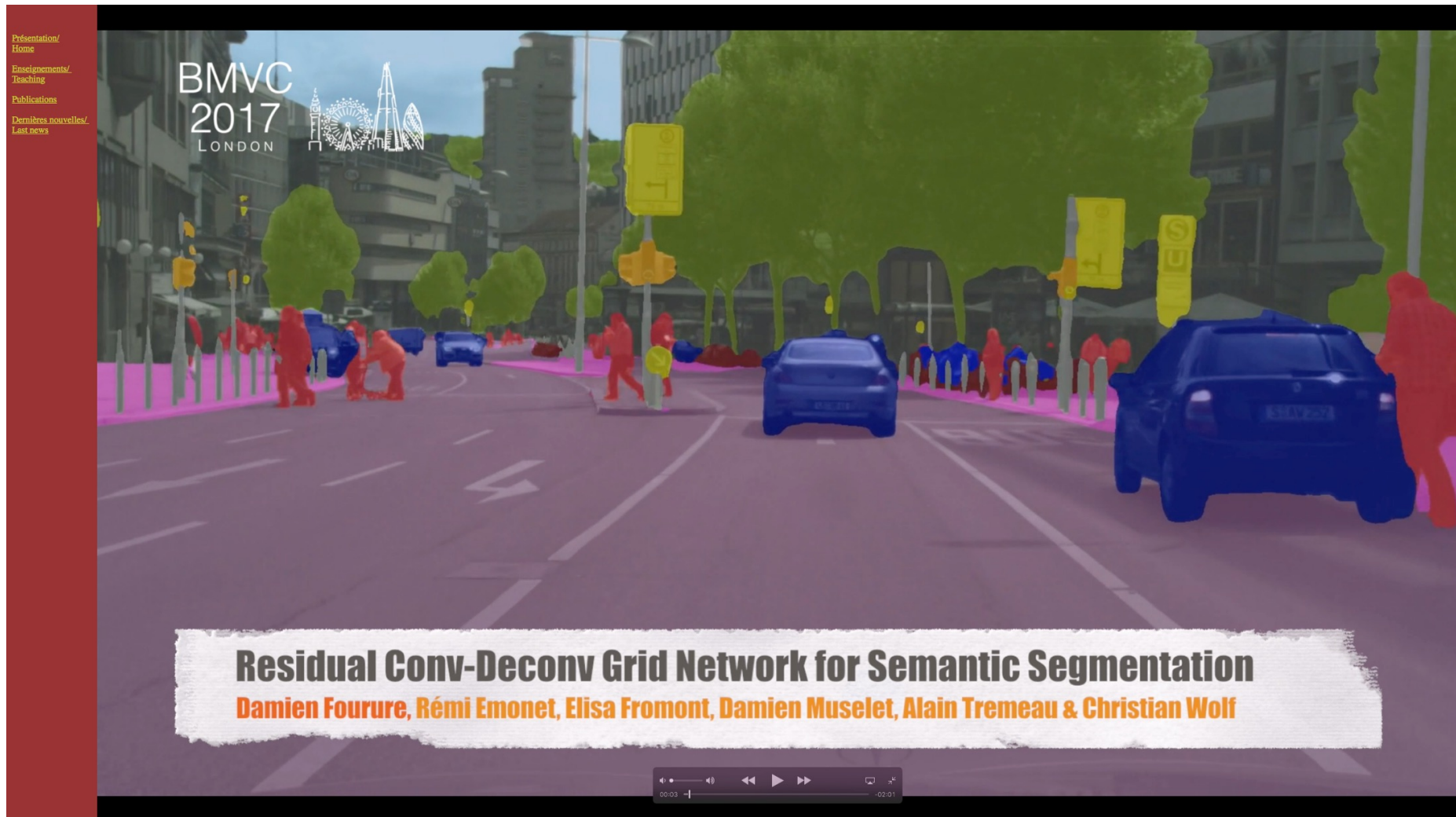
- 150 mile off-road robot race across the Mojave desert
- Natural and manmade hazards
- No driver, no remote control
- No dynamic passing
- Fastest vehicle wins the race (and 2 million dollar

# Illustration : Grand DARPA challenge (2005)





# Autonomous vehicle



Results by Damien Fourure on the Cityscape dataset for BMVC 2017 (Gridnet)

# Autonomous vehicle



Results by Damien Fourure on the Cityscape dataset for BMVC 2017 (Gridnet)



# Illustration

---

Systèmes autonomes avec apprentissage



# Recommandation automatique

- Netflix challenge

	User 1	User 2	User 3	User 4	User 5	...	User 480185	User 480186	User 480187	User 480188	User 480189
Terminator 2	5	5		4	...	2	5			5	
Gummo	1	1	2	?		...	3	2		?	
Clueless		4		?		...	2		4		
Napoleon Dynamite	4		2			...	5	5			
Pan's Labyrinth	4					...	5		5		
...	...	...	...	...	...	...	...	...	...	...	...
The Peanut Butter Solution	3				4	...	?	?			
X-Men	?			4		...	2	4			5
Edward Scissorhands	5			5		...		5			
Short Circuit	4	4				...	1				
Toy Story		?	4		5	...		4			

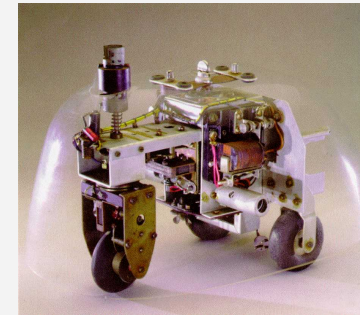
# Machine Learning

---

- **Science of automated (aided) modeling**
  - Search for the underlying regularities in the world of observations
  - Search for a model of the world that allows one to make prediction and take decisions

- **Science of adaptive systems**

- Reinforcement learning
- Simulated evolution



Identifier des **regularités**

# Plan

---

1. Une science de l'apprentissage ?
2. Les grands types d'apprentissage
3. Le problème de l'apprentissage supervisé
4. Apprendre dans un espace d'hypothèses structuré
5. Conclusion





Apprentissage **descriptif**

**non supervisé**

# Apprentissage descriptif

---

À propos d'un *échantillon d'apprentissage*  $s = \{(x_i)\}_{1,m}$

identifier des **régularités** rendant compte de  $S$

- E.g. sous la forme de **clusters** (e.g. *mélange de Gaussiennes*)
- **CLUSTERING**
- E.g. sous la forme de **motifs fréquents** (fouille de données)

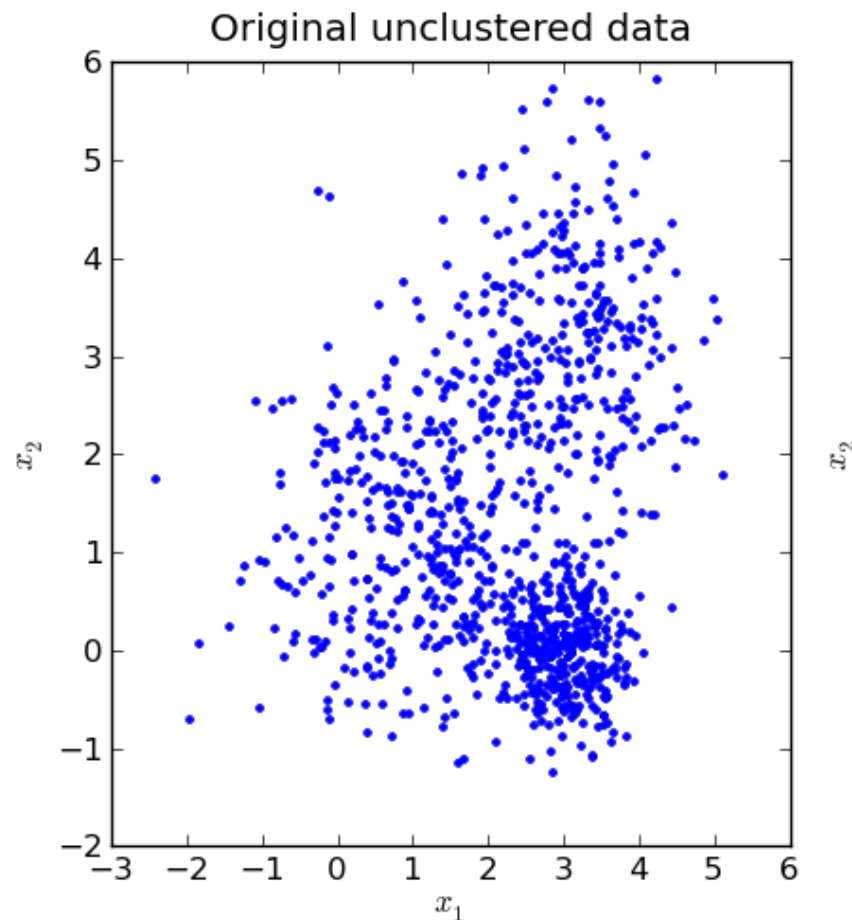
pour résumer, suggérer des régularités, comprendre ...



# Clustering / Catégorisation

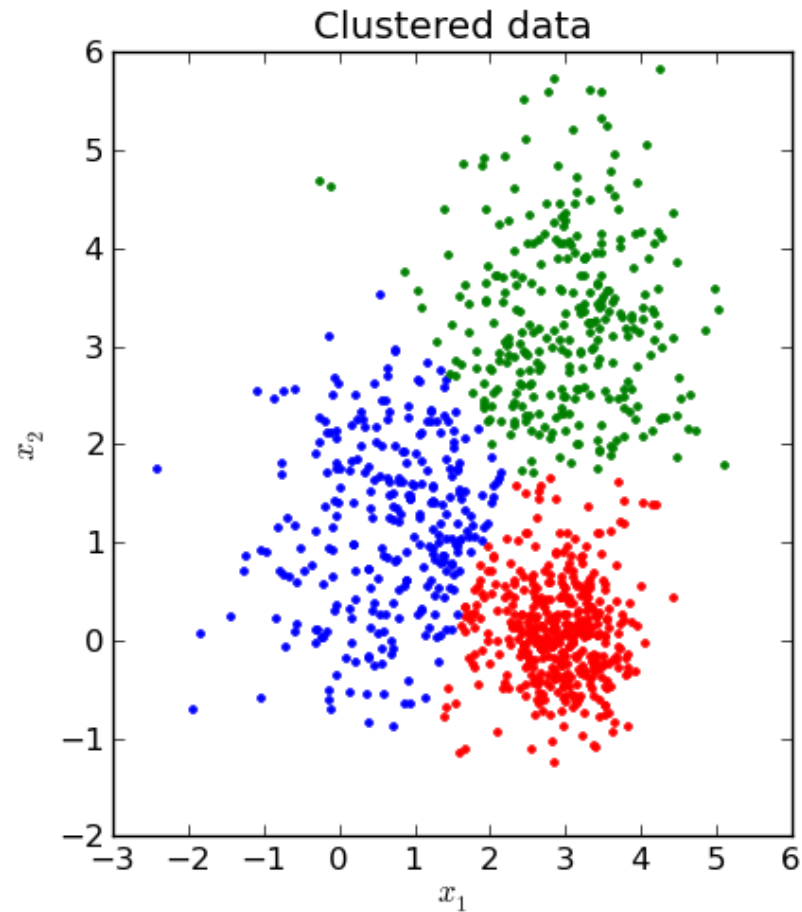
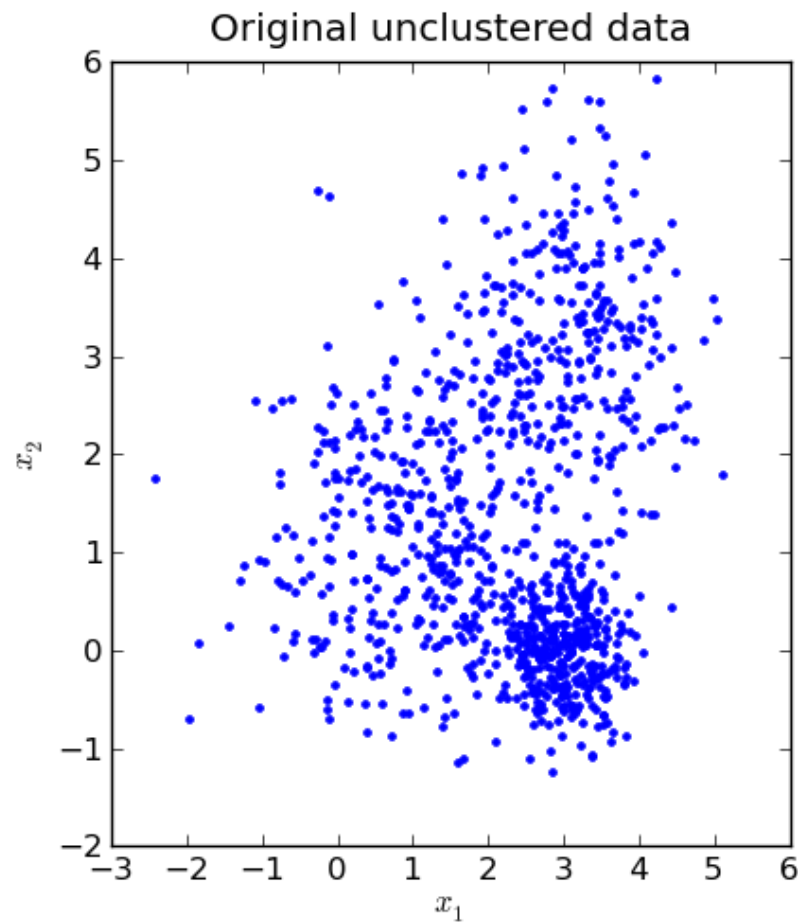
# Les grands types d'apprentissage

- Apprentissage « **descriptif** » (non supervisé)



# Les grands types d'apprentissage

- Apprentissage « **descriptif** » (non supervisé)

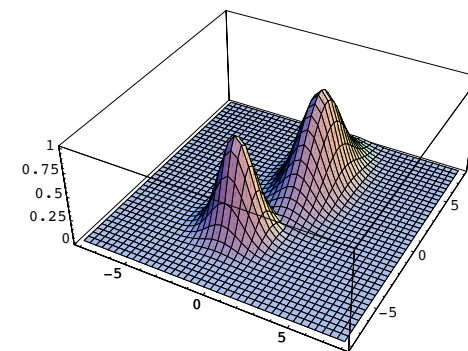


- **Catégorisation** de consommateurs

- Base de données sur les répondants de la base Nutrinet

- ~ 280 000
- Données sur *âge, nb de personnes dans la famille, catégorie socio-professionnelle, ...*
- Données sur consommations alimentaires sur une certaine durée

- Y a-t-il émergence de **groupes** distincts ?

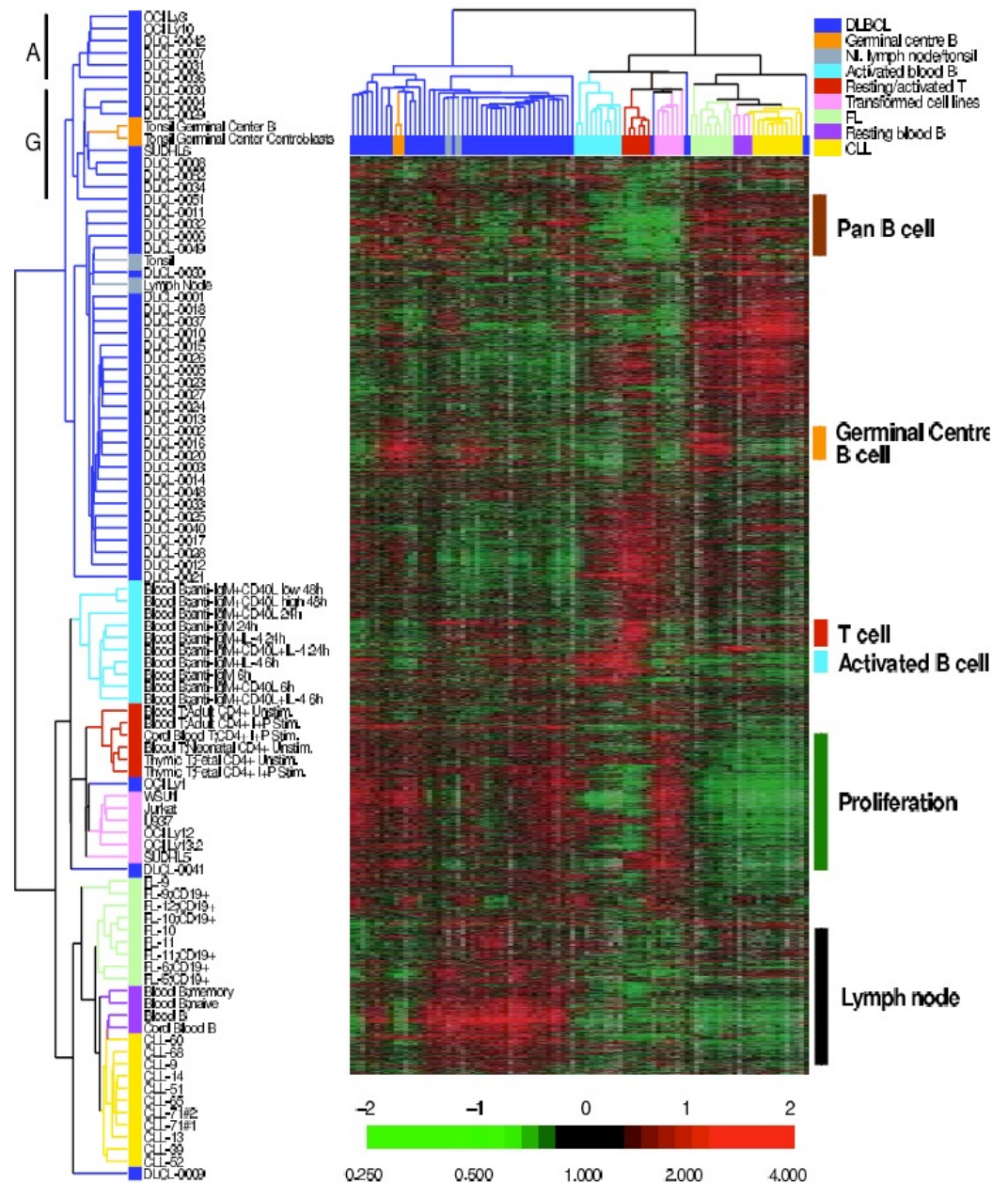


Apprentissage

Non supervisé

# Clustering

Bi-clustering  
gènes - patients







Recherche de motifs fréquents

*Frequent Item Sets*



Recherche de règles d'association

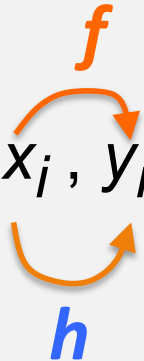
- Extraire des **régularités**
  - Base de données sur les **consommations alimentaires**
  - Peut-on identifier des « **patterns** » de consommation ?

# Apprentissage **prédictif** supervisé

# Apprentissage prédictif (*supervisé*)

---

- Un *échantillon d'apprentissage*

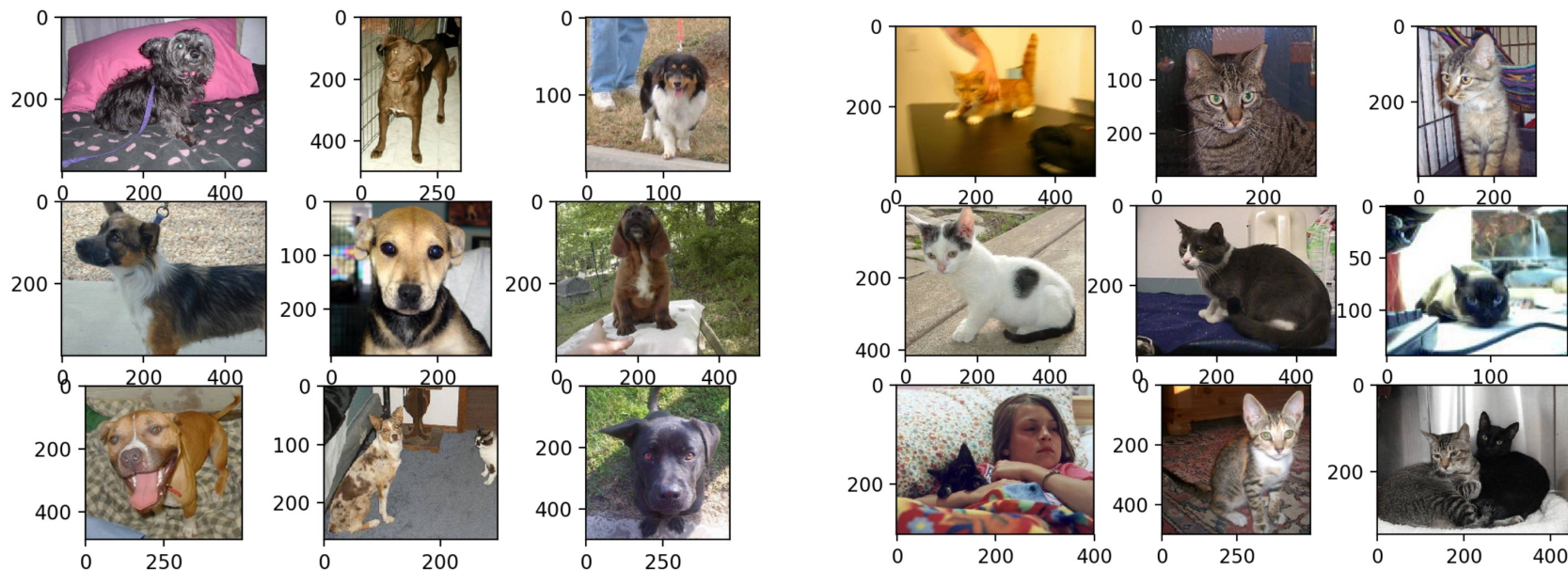
$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_m, y_m)\}$$


Prédiction pour de **nouveaux** exemples  $x \xrightarrow{h} y$  ?

## (2) Supervised Learning as ...

... Learning **a function** from an **input** space  $X$  to an **output** space  $Y$

Cats vs. dogs



- **Reconnaissance** d'insectes ravageurs
  - Base d'images d'insectes dans des cuvettes
    - *Reconnaissance du type d'insectes*
    - *Comptage*





- 
- **Spam** ou pas spam



- Article portant sur la **politique** ou sur le **sport**



- **Pathologie** dont souffre un patient



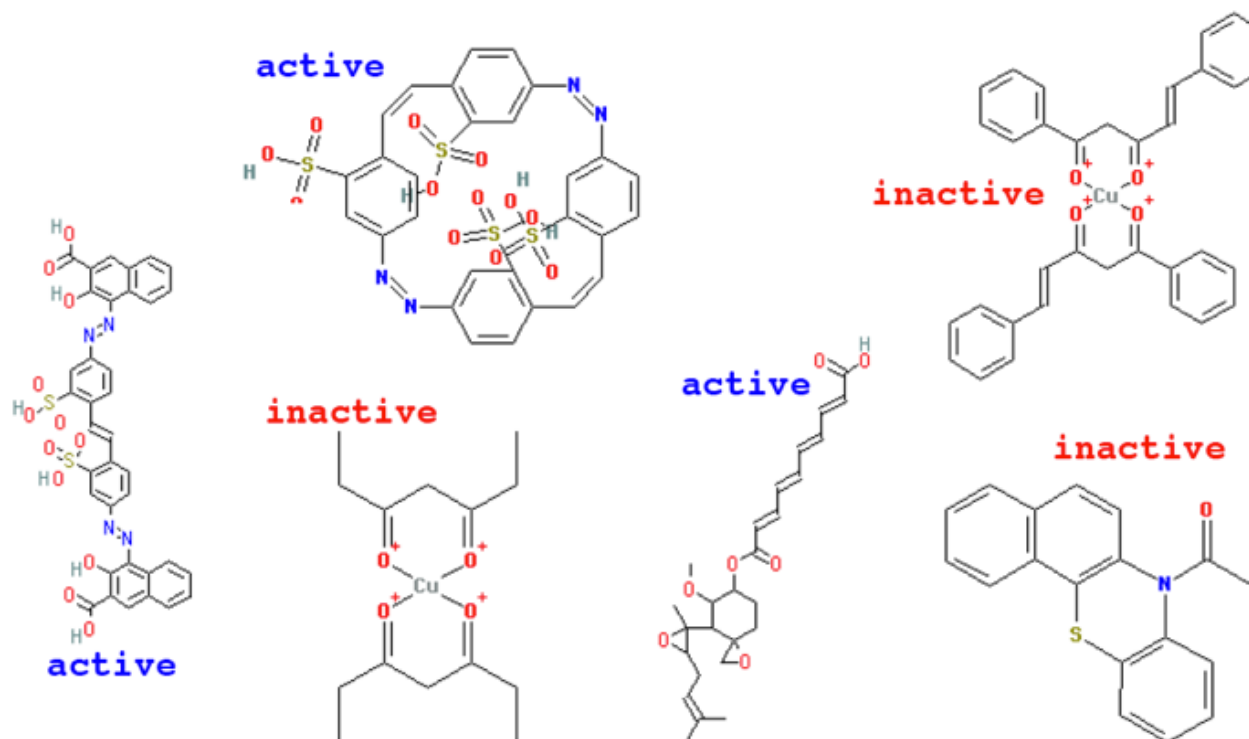
- **Objet** présent dans une image



# Association / Prédiction

Apprentissage  
supervisé




- Prédire si une molécule est bio-active ou pas



NCI AIDS screen results (from <http://cactus.nci.nih.gov>).

# Analyse de textes

- Reconnaissance de **sentiments** exprimés dans des textes

	Electronics	Video games
	(1) <u>Compact</u> ; easy to operate; very good picture quality; looks <u>sharp</u> !	(2) A very <u>good</u> game! It is action packed and full of excitement. I am very much <u>hooked</u> on this game.
	(3) I purchased this unit from Circuit City and I was very <u>excited</u> about the quality of the picture. It is really <u>nice</u> and <u>sharp</u> .	(4) Very <u>realistic</u> shooting action and good plots. We played this and were <u>hooked</u> .
	(5) It is also quite <u>blurry</u> in very dark settings. I will <u>never_buy</u> HP again.	(6) It is so boring. I am extremely <u>unhappy</u> and will probably <u>never_buy</u> UbiSoft again.

# GIEC : filtrage de documents

---

- Estimation de l'émission de gaz à effet de serre par les sols agricoles
  - En particulier N<sub>2</sub>O (influence des engrais azotés)
- Par une méta-analyse des **articles scientifiques pertinents**
  - **Plus de 10<sup>6</sup> articles** scientifiques publiés / an
  - (plus ou moins) disponibles sur Internet

**Filtrage** nécessaire de ces articles

En optimisant **précision** et **rappel**

(et **interprétabilité** du filtre)

# Supervised learning

---

- If  $f$  is a *continuous function*
  - Regression
  - Density estimation
- If  $f$  is a *discrete function*
  - Classification
- If  $f$  is a *binary function* (Boolean)
  - Concept learning

# Supervised learning

---

- ***Discrimination***

- One can predict that
  - clients
    - Adding up international calls for more than 300€/month
    - and who have made more than 3 reclamations in the past
  - Are likely to change for another provider

- ***Regression***

- The number of accidents declared by a driver is
  - inversely proportional to the duration of its driver's license,
  - with coefficients that are specific to each gender.

# Apprentissage prescriptif pour « intervenir »



# Apprentissage prescriptif

---

- Apprentissage « **prescriptif** » (recherche de *causalités*)

1. J'observe que les gens qui mangent des glaces

sont souvent en maillot de bain

2. Je voudrais vendre davantage de glaces

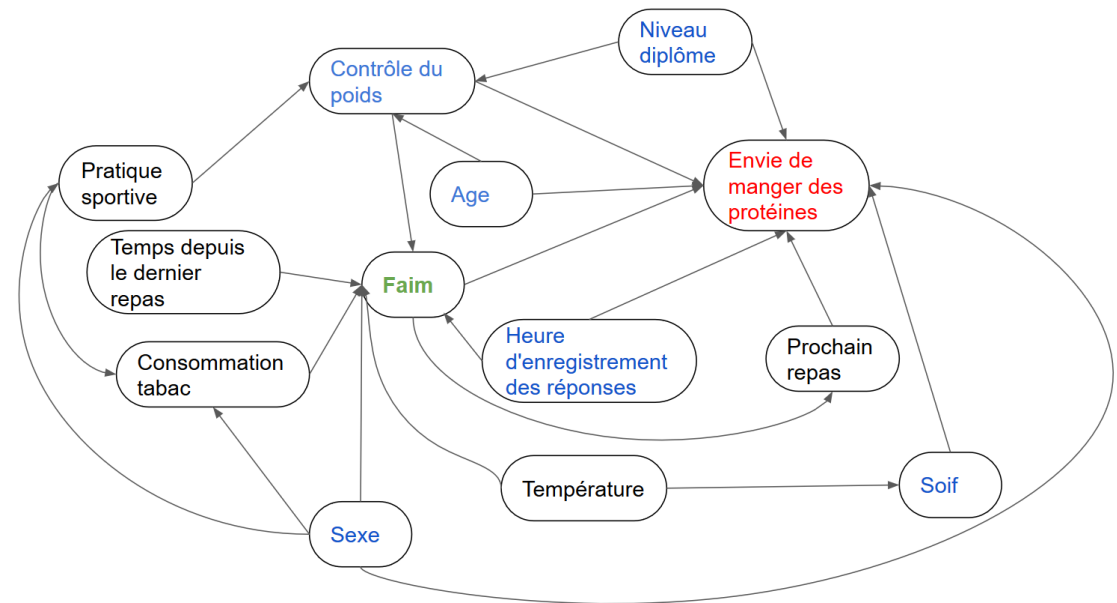


Je demande aux gens de se mettre en maillot de bain

# La recherche de relations causales

Qu'est-ce qui **cause** l'appétence pour des plats protéinés ?

- La **faim** ?
- L'**heure** dans la journée ?
- Le **genre** ?
- L'**aspect visuel** ?
- L'**aspect olfactif** ?
- La richesse en **protéines** des **repas précédents** ?
- ...



- Quelles **recommandations** faire à un consommateur pour qu'il baisse sa consommation d'aliments carnés ?
- Quel impact **si on double le prix** de ... ?
- Quel rendement aurais-je eu l'année dernière **si j'avais** planté du ... au lieu de ...

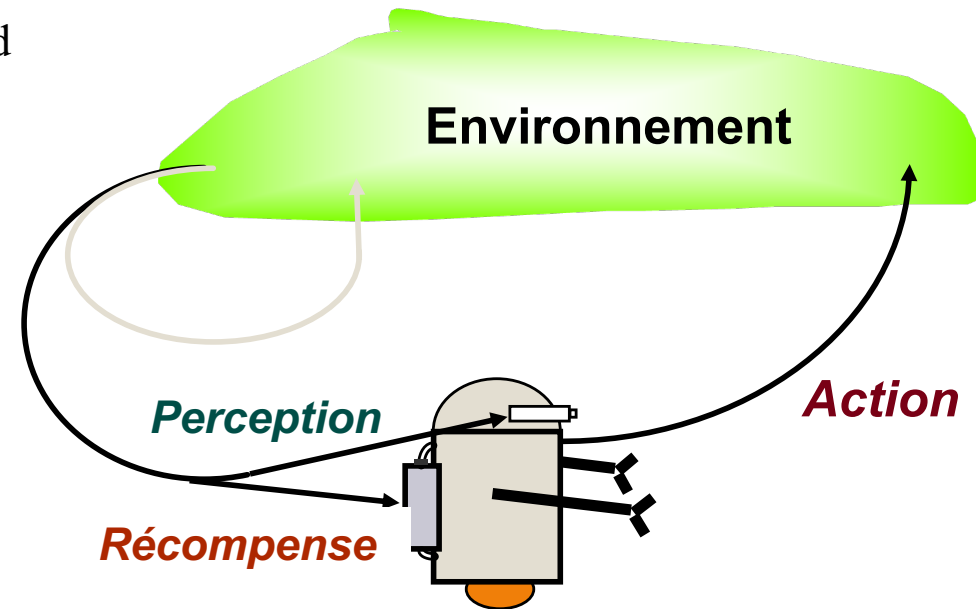
# Apprentissage par renforcement

# L'apprentissage par **renforcement**

...

The **learning data** are

- A **sequence** of perceptions, actions and rewards  $(s_t, a_t, r_t)_{t=1, \infty}$ 
  - With a **reinforcement**  $r_t$
  - That can be related to past actions made far before  $t$



**The problem:** infer a function

*perceived situation* → *action*

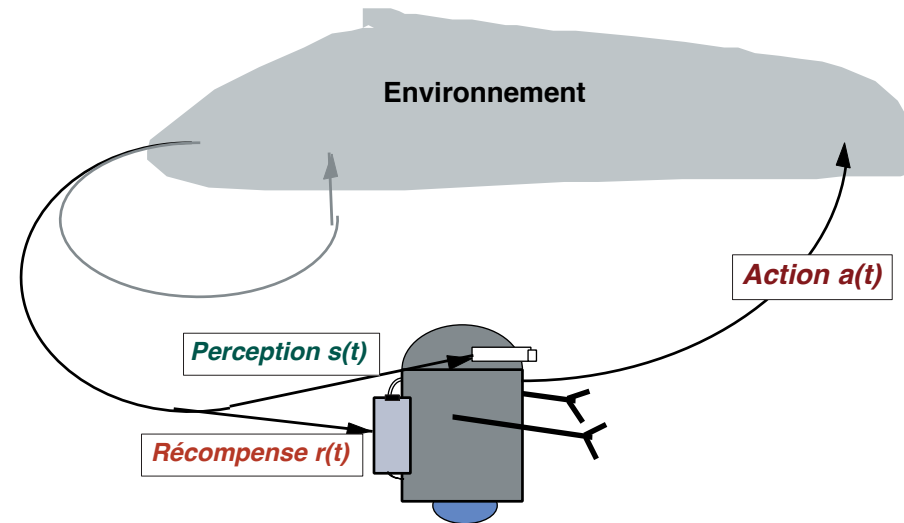
so as to maximise a gain over long term

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots + r_T = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

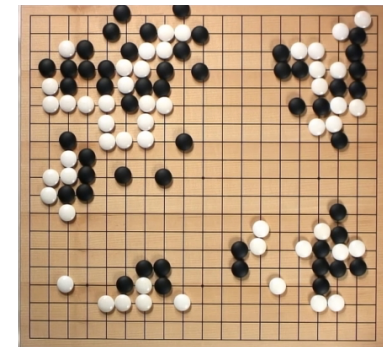
Akin to **learning reflexes**

# Les grands types d'apprentissage

- Apprentissage « **par renforcement** » (comment (ré)agir)



1. **Piloter un hélicoptère**
2. Apprendre à **jouer au tennis de table**
3. Battre le champion de **back-gammon** (1992), de **Go** (2016)
4. **Gérer un porte-feuille** d'investissements
5. **Contrôler une usine** de production électrique

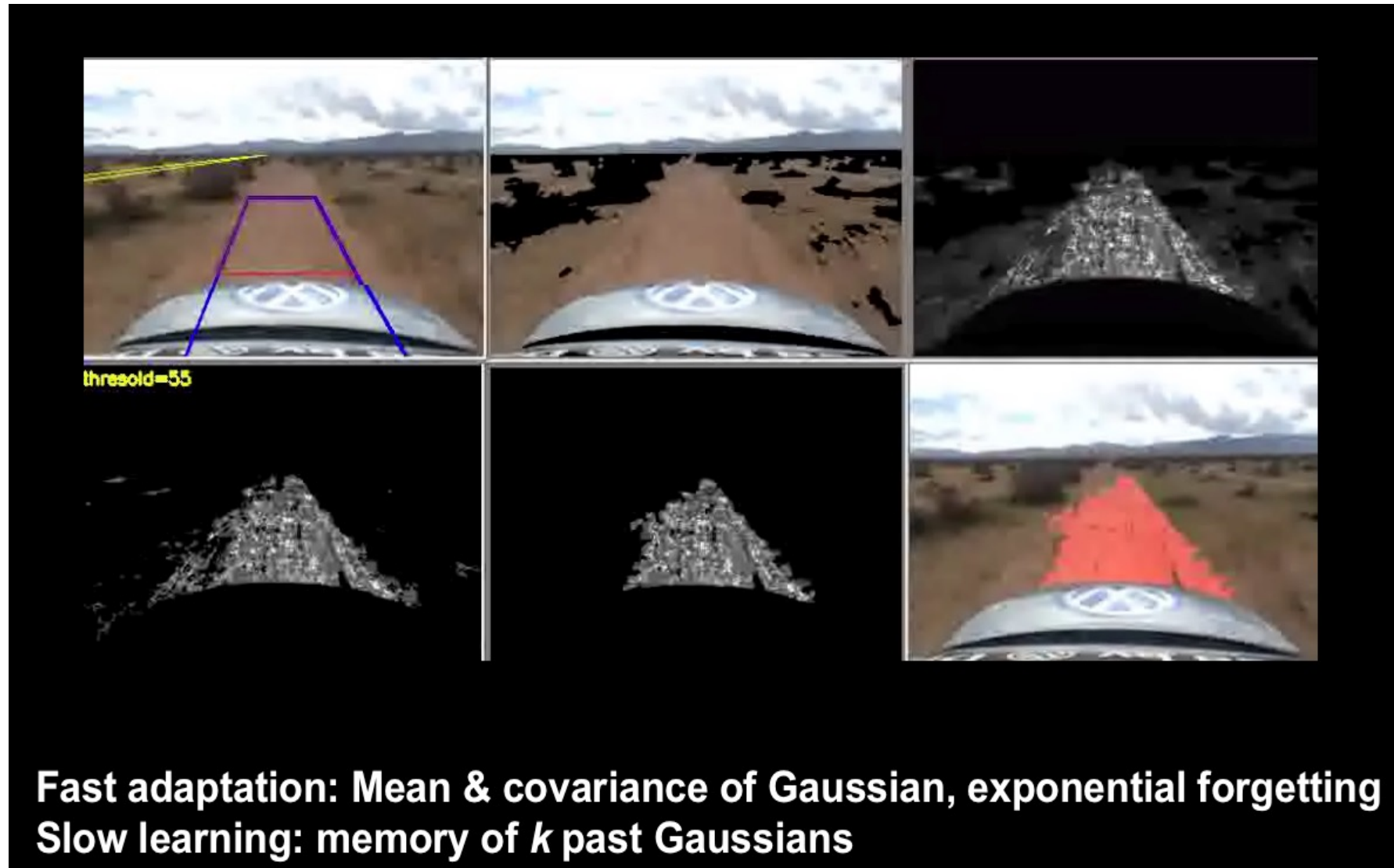


# Illustration : Grand DARPA challenge (2005)





## Illustration : Grand DARPA challenge (2005)



C'est quoi des données ?

# Les données : organisation et types

Identifieur	Genre	Age	Niveau études	Marié ?	Nb enfants	Revenu	Profession	A prospecter ?
I_21	M	43	Bac+5	Oui	3	55 000	Architecte	OUI
I_34	M	25	Bac+2	Non	0	21 000	Infirmier	NON
I_38	F	34	Bac+8	Oui	2	35 000	Chercheuse	OUI
I_39	F	67	Bac	Oui	5	20 000	Retraitée	NON
I_58	F	56	CAP	Oui	4	27 000	Ouvrière	NON
I_73	M	40	Bac+3	Non	2	31 000	Commercial	OUI
I_81	F	51	Bac+5	Oui	3	75 000	Chef d'entreprise	OUI

# Les données : organisation et types

Identifieur	Genre	Age	Niveau études	Marié ?	Nb enfants	Revenu	Profession	A prospecter ?
I_21	M	43	Bac+5	Oui	3	55 000	Architecte	OUI
I_34	M	25	Bac+2	Non	0	21 000	Infirmier	NON
I_38	F	34	Bac+8	Oui	2	35 000	Chercheuse	OUI
I_39	F	67	Bac	Oui	5	20 000	Retraitée	NON
I_58	F	56	CAP	Oui	4	27 000	Ouvrière	NON
I_73	M	40	Bac+3	Non	2	31 000	Commercial	OUI
I_81	F	51	Bac+5	Oui	3	75 000	Chef d'entreprise	OUI

Exemple  
(*example, instance*)

Descripteur  
Attribut  
(*feature*)

Étiquette  
(*label*)

# Les données

- Vectorielles
- Séquences
- Structurés
- Temporelles
- Spatiales

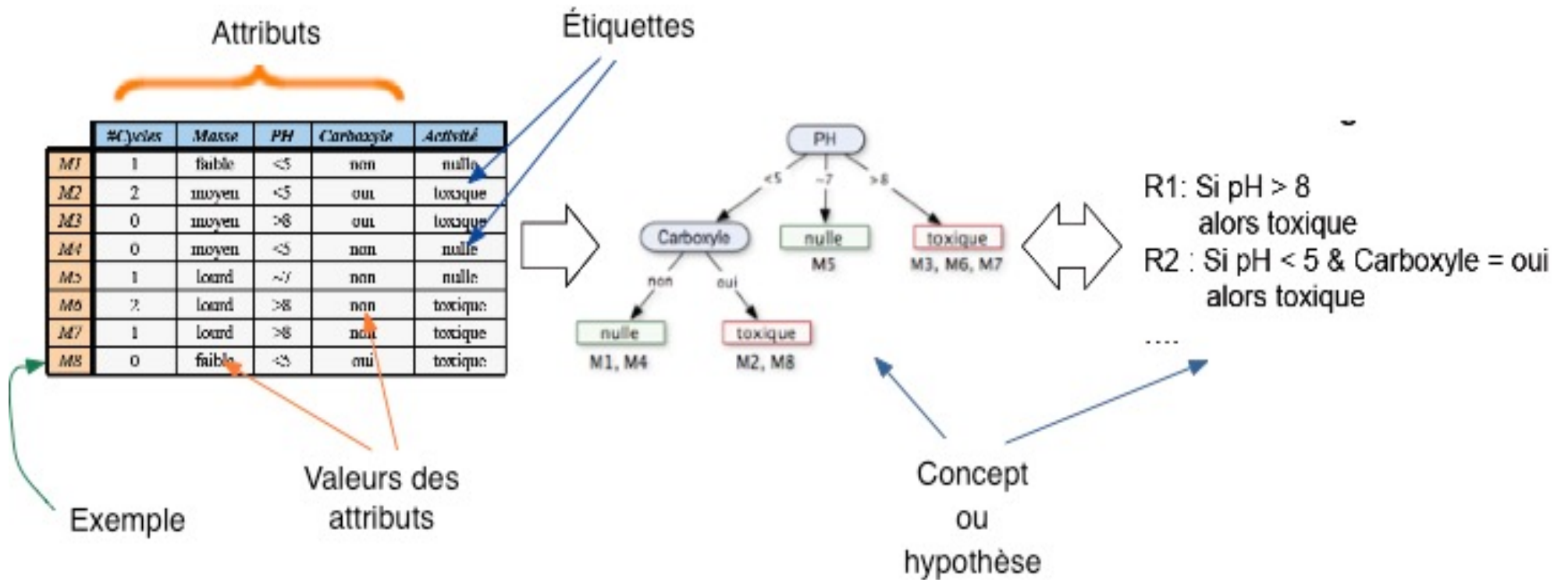
Identifieur	Genre	Age	Niveau études	Marié ?	Nb enfants	Revenu	Profession	A prospecter ?
I_21	M	43	Bac+5	Oui	3	55 000	Architecte	OUI
I_34	M	25	Bac+2	Non	0	21 000	Infirmier	NON
I_38	F	34	Bac+8	Oui	2	35 000	Chercheuse	OUI
I_39	F	67	Bac	Oui	5	20 000	Retraitée	NON
I_58	F	56	CAP	Oui	4	27 000	Ouvrière	NON
I_73	M	40	Bac+3	Non	2	31 000	Commercial	OUI
I_81	F	51	Bac+5	Oui	3	75 000	Chef d'entreprise	OUI

Exemple  
(*example, instance*)

Descripteur  
Attribut  
(*feature*)

Étiquette  
(*label*)

# Induction supervisée



# Les données

---

- Vectorielles

La protéine « [sp|P00004|CYC\\_HORSE](#) » est activée par ...

- **Séquences**

```
1  ttcagttgtg aatgaatgga cgtgccaaat agacgtgccg ccgccgctcg attgcactt
61  tgctttcggg ttgcccgtcg tttcacgcgt ttagttccgt tcggttcatt cccagttctt
121 aaataccgga cgtaaaaata cactctaacg gtcccgcgaa gaaaaagata aagacatctc
181 gtagaaatat taaaataaat tcctaaagtc gttggtttct cgttcacttt cgctgcctgc
...
4021 agaacacgcc gaggctccat tcatagcacc acttcgtcgt cttaatcccc tcctcatcc
4081 gccatggcgg tgcaaaaaat aaaaagaact c
```

- Structurés

- Temporelles

- Spatiales

```
DEVICE=eth0
BOOTPROTO=none
ONBOOT=yes
IPADDR=192.168.0.X
NETMASK=255.255.255.0
GETEWAY=192.168.0.254
search exemple.com namserver
192.168.0.254
```



# Les données

- Vectorielles

Logique du 1<sup>er</sup> ordre :

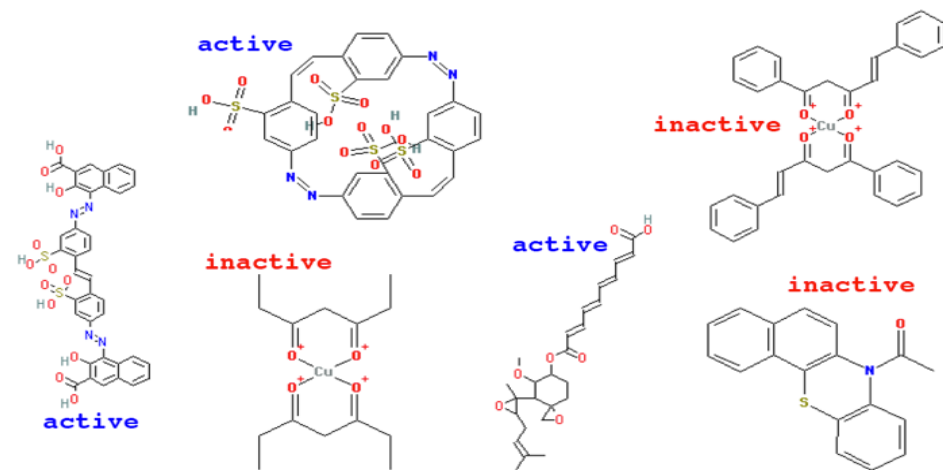
bloc(B1) & surtable(B2) & au-dessus(B1,B2) & ...

- Séquences

- **Structurés**

- Temporelles

- Spatiales

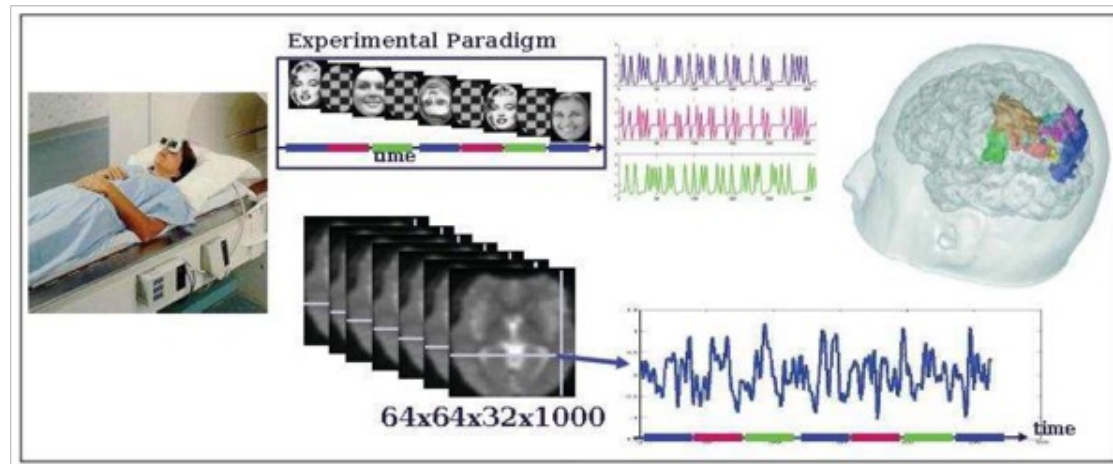


NCI AIDS screen results (from <http://cactus.nci.nih.gov>).

# Les données

- Vectorielles
- Séquences
- Structurés
- **Temporelles**
- Spatiales

• **Apprentissage supervisé** : interprétation d'IRMf



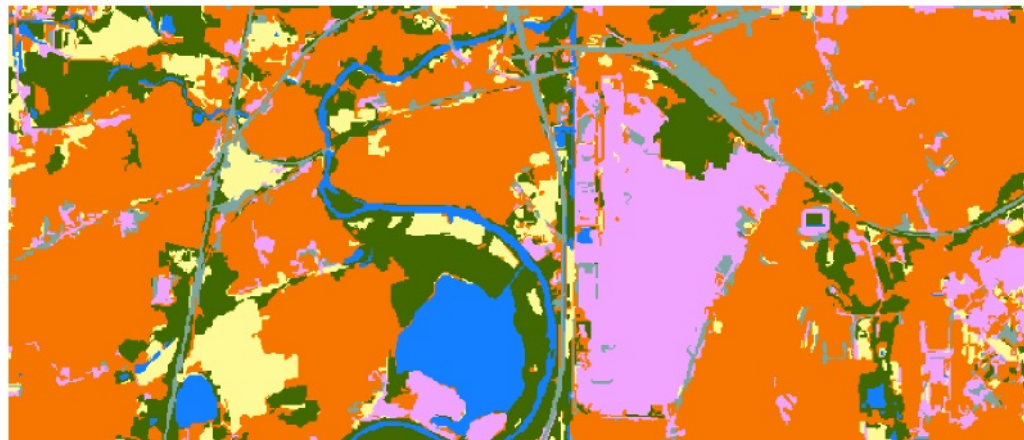
*Trouble de la reconnaissance de visage ou non*

# Les données

- Vectorielles
- Séquences
- Structurés
- Temporelles
- **Spatiales**

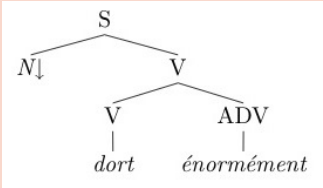



Image MRS

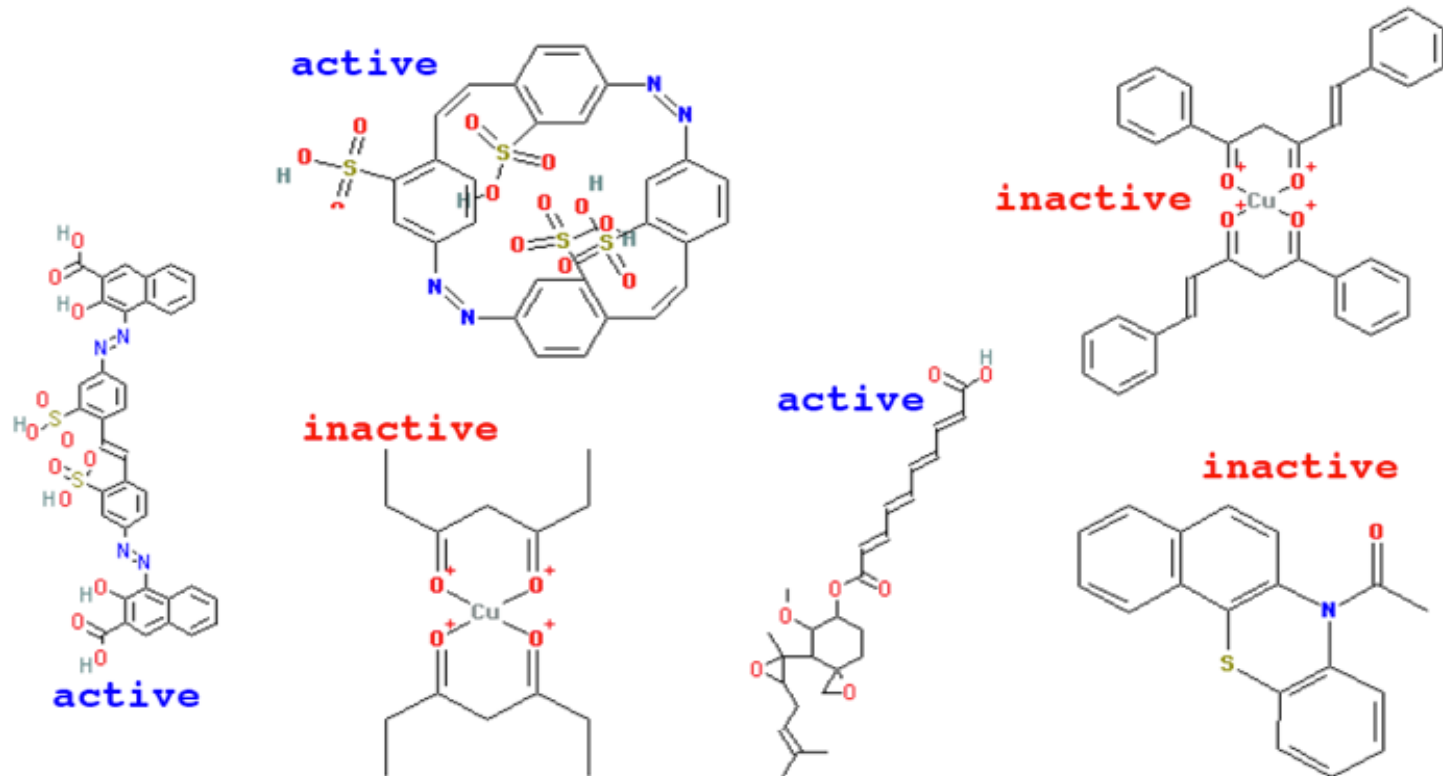


- |                          |                       |
|--------------------------|-----------------------|
| ■ Quartiers résidentiels | ■ Routes              |
| ■ Quartiers industriels  | ■ Zones agricoles     |
| ■ Surfaces d'eau         | ■ Zones de végétation |

# Types de formats

<b>Numérique continue</b> (R)	Compte en banque : 12 915,86 €
<b>Numérique discrète</b> (N ou Z)	Nombre d'enfants : 11
<b>Binaire</b>	Célibataire : vrai
<b>Catégorie</b>	Couleur dans {rouge, vert, bleu}
<b>Texte</b>	La protéine « sp P00004 CYC_HORSE » est activée par ...
<b>Données structurées</b>	<p>Arbre, expression XML, ...</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <pre> graph TD     S --&gt; N1[N]     S --&gt; V1[V]     V1 --&gt; V2[V]     V1 --&gt; ADV[ADV]     V2 --&gt; dort[dort]     ADV --&gt; enormement[énormément]             </pre> </div> <div style="border: 1px solid gray; padding: 5px; background-color: #f9f9f9;"> <pre> &lt;?xml version="1.0"?&gt; &lt;quiz&gt;   &lt;question&gt;     Who was the forty-second     president of the U.S.A.?   &lt;/question&gt;   &lt;answer&gt;     William Jefferson Clinton   &lt;/answer&gt;   &lt;!-- Note: We need to add   more questions later.--&gt; &lt;/quiz&gt;             </pre>  </div> </div>
<b>Séquences</b>	<ul style="list-style-type: none"> <li>- Génome</li> <li>- Séquence de requêtes sur site web</li> </ul>
<b>Images, vidéos</b>	

# Apprentissage supervisé



NCI AIDS screen results (from <http://cactus.nci.nih.gov>).

# Plan

---

1. Une science de l'apprentissage ?
2. Les grands types d'apprentissage
3. Le problème de l'apprentissage supervisé
4. Apprendre dans un espace d'hypothèses structuré
5. Conclusion

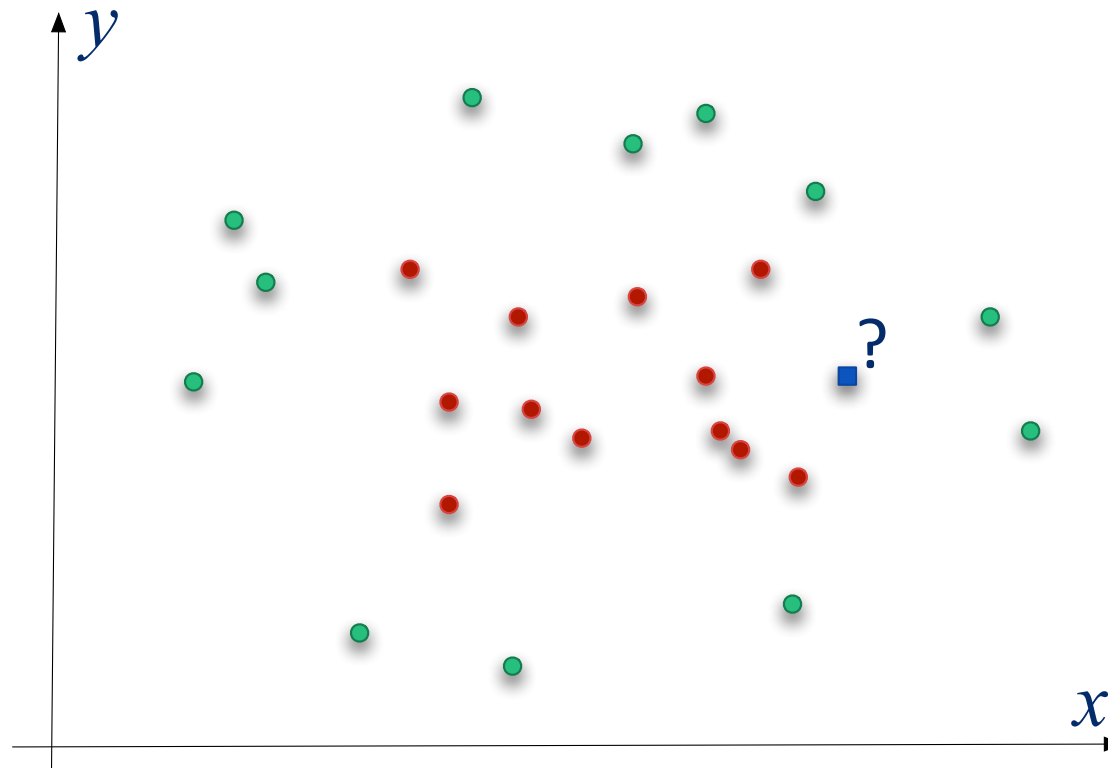
Comment fonder l'induction ?



# Supervised induction

---

- We want to be able to predict the class of unseen examples



→ A decision function

# Le problème de l'induction

- Exemples décrits par :
  - Nombre** (1 ou 2); **taille** (petit ou grand); **forme** (cercle ou carré); **couleur** (rouge ou vert)
- De classe + ou -

Description	Votre prédiction	Vraie classe

Combien de fonctions possibles de X vers Y ?

$$2^{2^4} = 2^{16} = 65,536$$

Combien de fonctions restent après 6 exemples d'apprentissage?

$$2^{10} = 1024$$

# Le problème de l'induction

- Exemples décrits par :

**Nombre** (1 ou 2); **taille** (petit ou grand); **forme** (cercle ou carré); **couleur** (rouge ou vert)

Description	Votre prédiction	Vraie classe
1 grand carré rouge		-
1 grand carré vert		+
2 petits carrés rouges		+
2 grands cercles rouges		-
1 grand cercle vert		+
1 petit cercle rouge		+
1 petit carré vert		-
1 petit r carré ouge		+
2 grands carrés verts		+
2 petits carrés verts		+
2 petits cercles rouges		+
1 petit cercle vert		-
2 grands cercles verts		-
2 petits cercles verts		+
1 grand cercle rouge		-
2 grands carrés rouges	?	

15

Combien de  
fonctions  
restantes ?



# Le problème de l'induction

- Examples described using:

**Number** (1 or 2); **size** (small or large); **shape** (circle or square); **color** (red or green)

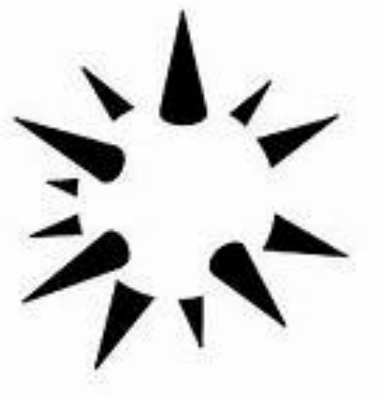
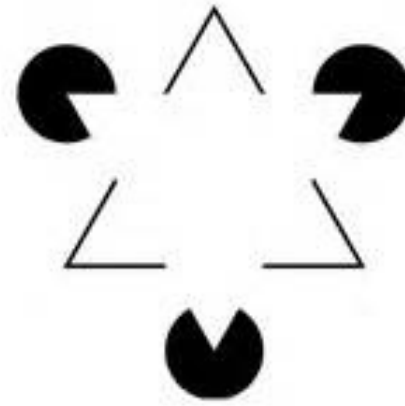
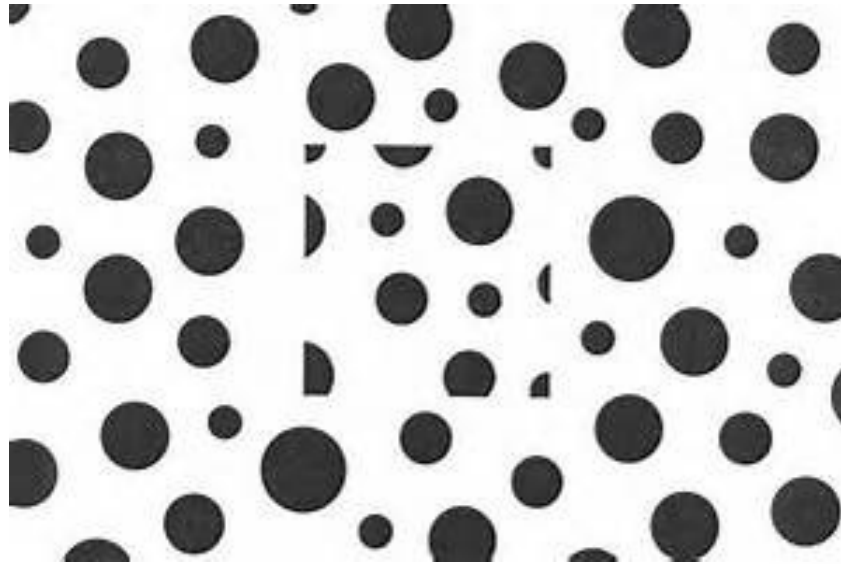
Description	Your prediction	True class
1 grand carré rouge		-
1 grand carré vert		+
2 petits carrés rouges		+
2 grands cercles rouges		-
1 grand cercle vert		+
1 petit cercle rouge		+

Combien de fonctions possibles avec 2 descripteurs de X à Y ?  $2^{2^2} = 2^4 = 16$

Combien de fonctions restent après 3 exemples différents ?  $2^1 = 2$

# L'apprentissage – une extrapolation nécessitant des a priori

---



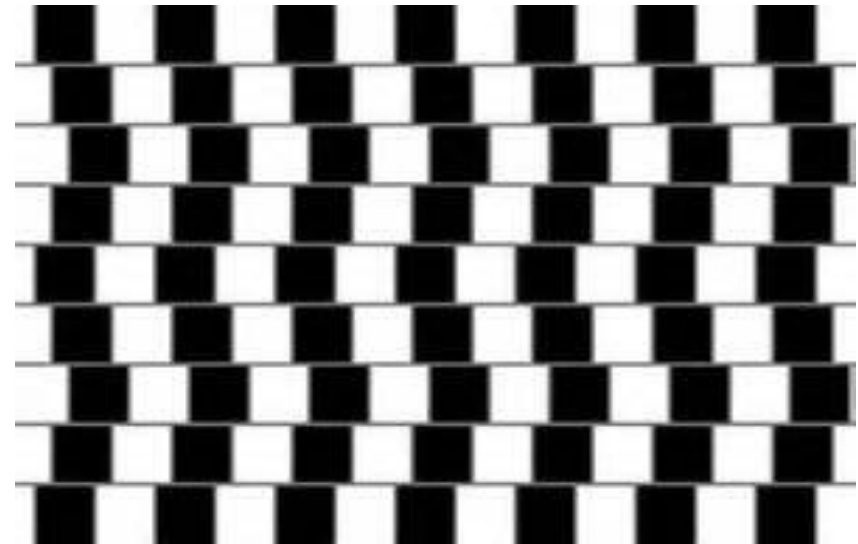
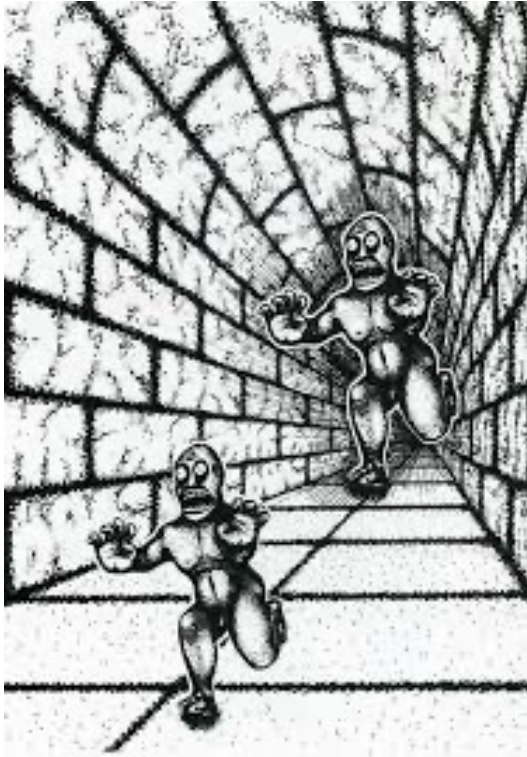
## Interpreting – completion of percepts

---



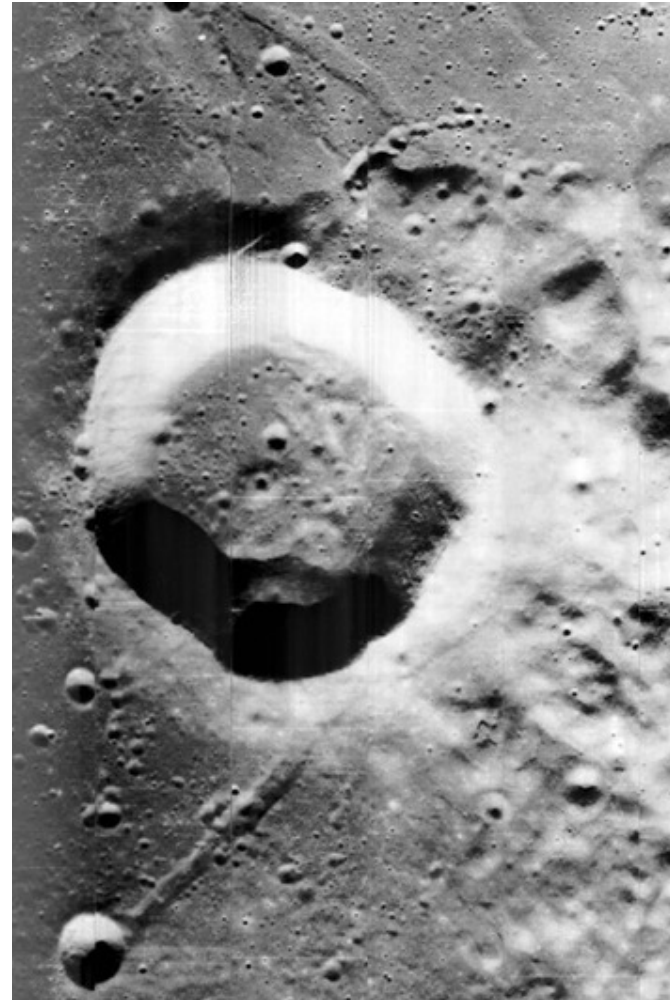
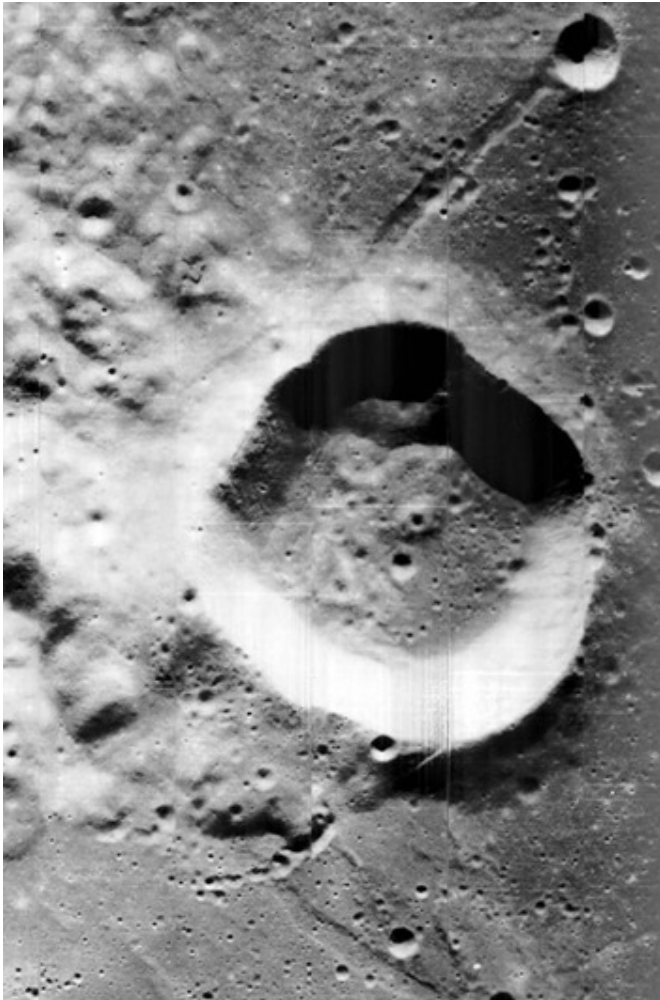
# Induction and its illusions

---



# Induction et illusions

---



Cratère ou colline ?



# Role of induction

---

- [Leslie Valiant, « *Probably Approximately Correct. Nature's Algorithms for Learning and Prospering in a Complex World* », Basic Books, 2013]

« From this, we have to conclude that **generalization** or **induction** is a **pervasive phenomenon** (...). It is as routine and reproducible a phenomenon as objects falling under gravity. It is **reasonable to expect a quantitative scientific explanation** of this highly reproducible phenomenon. »

# Induction: a double question

---

Some green emeralds => all emeralds are green

In each case:

observations => laws / general rules or ways to adapt to new situations

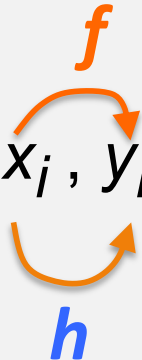
1. How to find such rules? The problem of invention
2. Can we guarantee something about those “generalizations”?

The problem of justification

# Supervised learning

---

- A *learning set*

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_m, y_m)\}$$


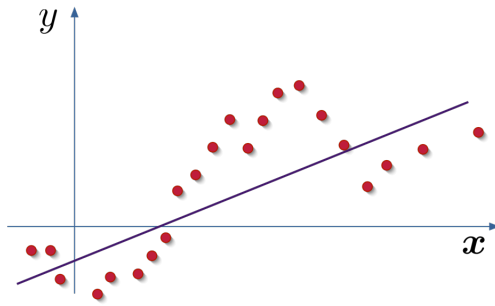
- **Prediction** for new examples  $\mathbf{x} \xrightarrow{h} \mathbf{y} ?$

# Induction: an impossible game?

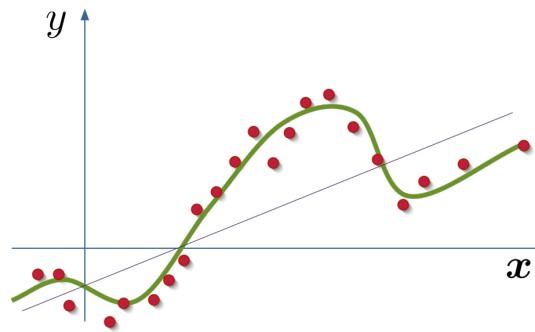
---

- **A bias is need**
- **Types of bias**
  - **Representation bias** (declarative)
  - **Research bias** (procedural)

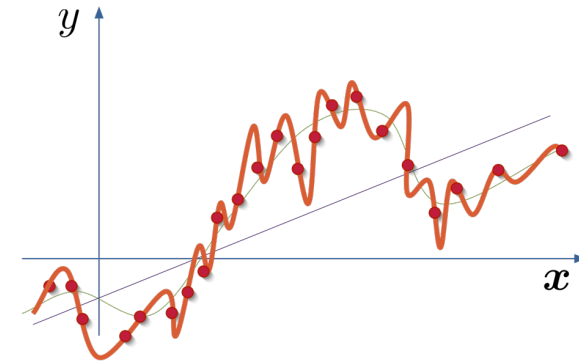
# Comment garantir un niveau de performance ?



Sous-apprentissage

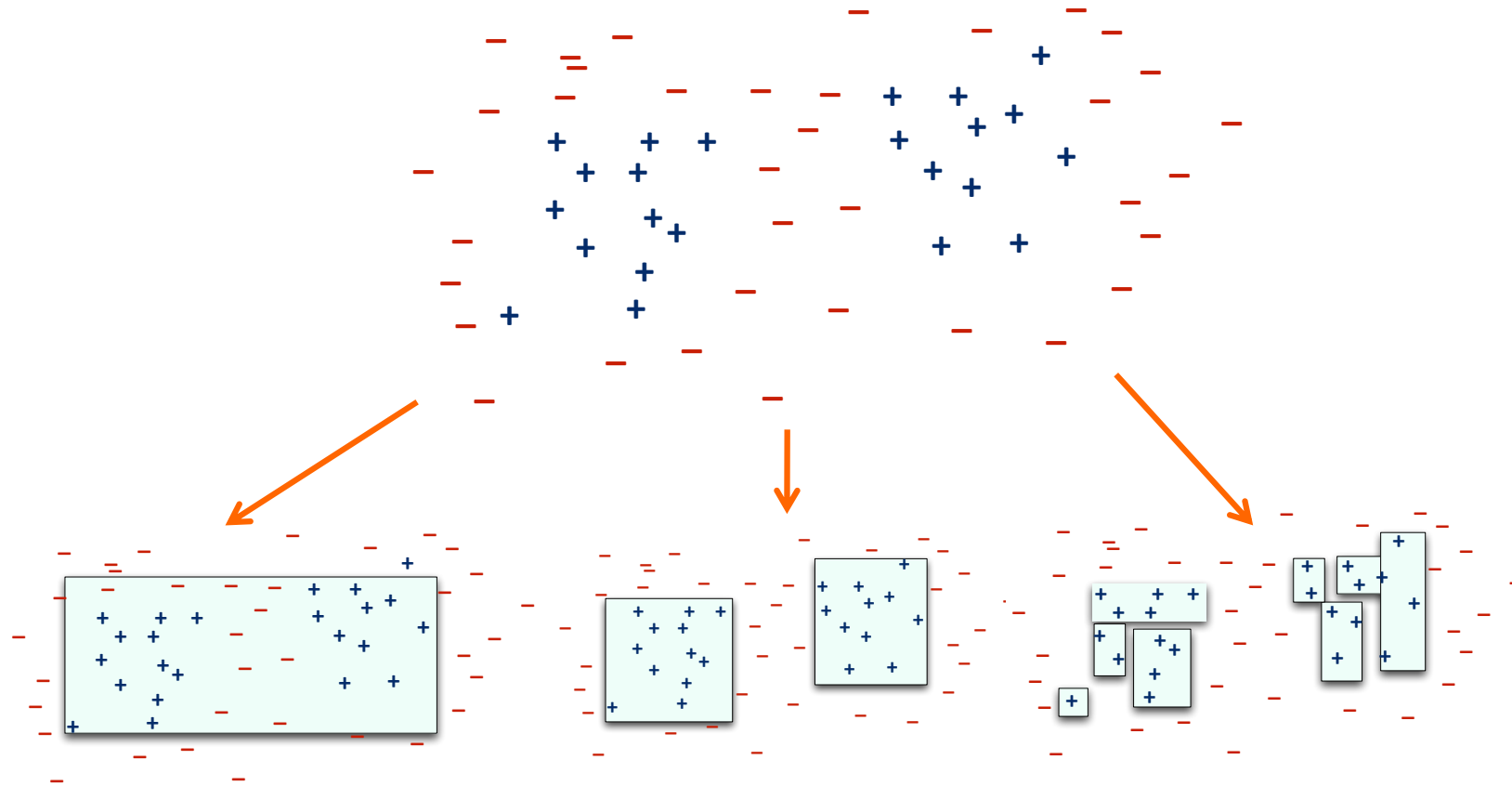


Bon-apprentissage



Sur-apprentissage

# Comment garantir un niveau de performance ?



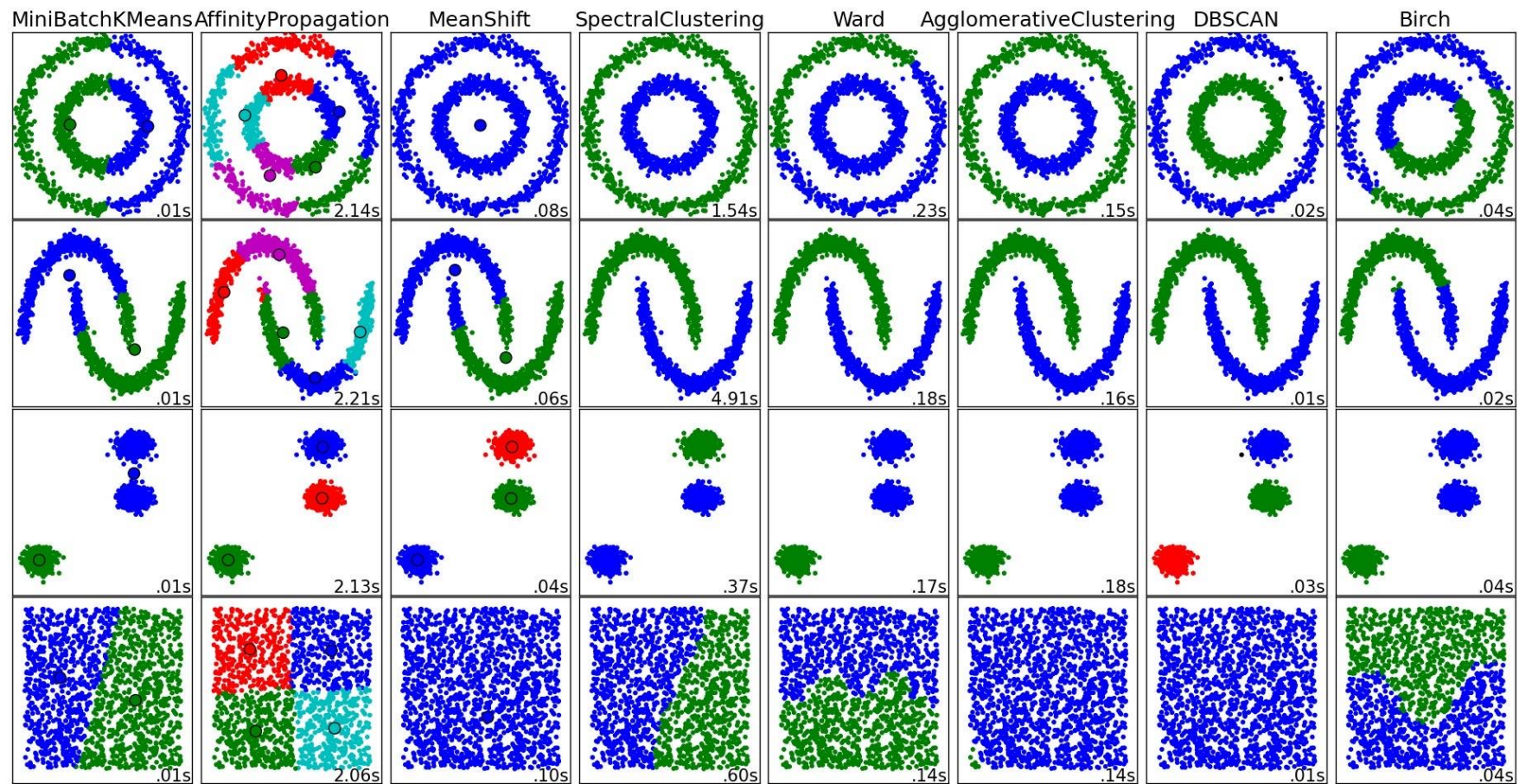
Sous-apprentissage

Bon-apprentissage

Sur-apprentissage

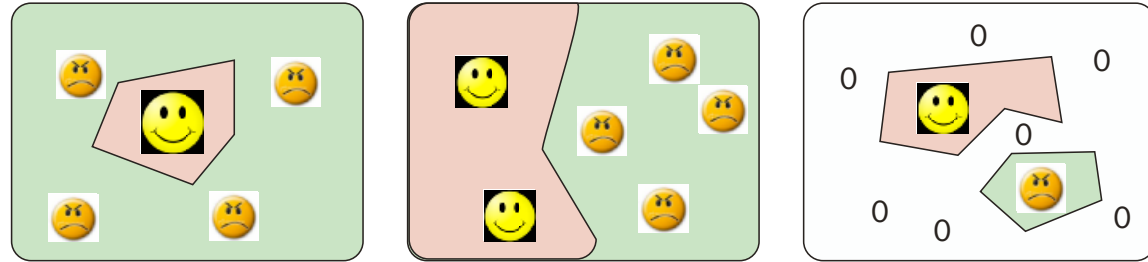
# Illustration : le clustering

Les biais a priori sont nécessaires et ... il faut bien les choisir



# Le no-free-lunch theorem

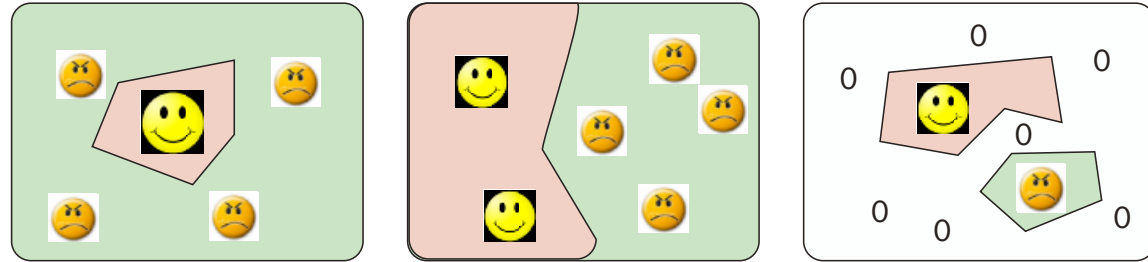
Possible



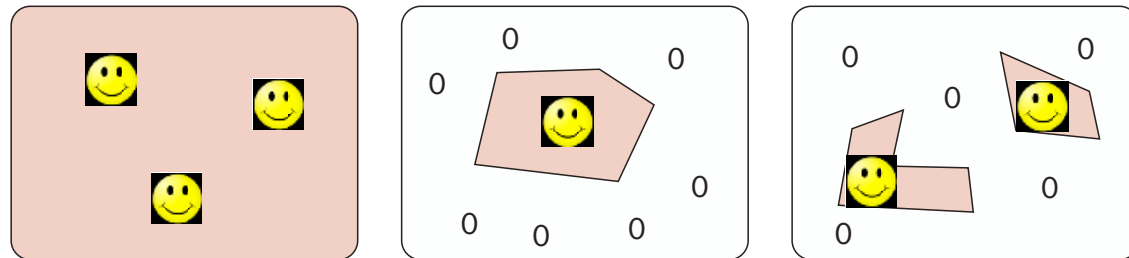


# Le no-free-lunch theorem

Possible



Impossible



Il faut **choisir** le **bon algorithme** pour la **classe de problèmes** étudiée