ÉCOLE MATHÉMATIQUE AFRICAINE : Bases mathématiques de l'intelligence artificielle

Introduction à l'apprentissage artificiel en santé (cours 1/2)

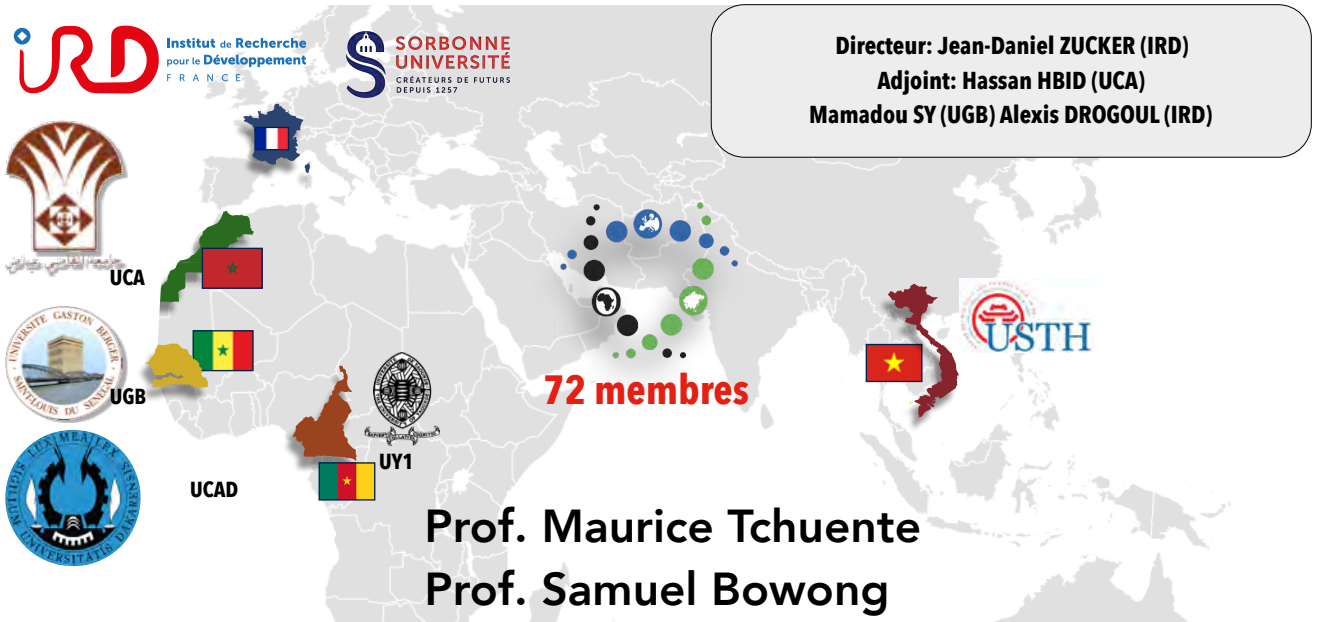
JEAN-DANIEL ZUCKER

Unité Mixte Internationale

UMMISCO

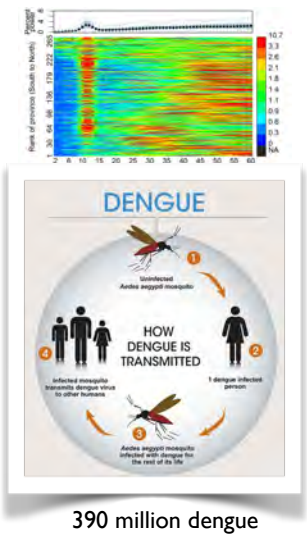
Jean-Daniel Zucker

UMMISCO : laboratoire de recherche international sous la tutelle de 7 partenaires académiques → explorer et concevoir des méthodes mathématiques et informatiques innovantes pour la modélisation et la simulation de systèmes complexes…



Modèle de partenariat scientifique multi-sites équitable, dans lequel des enseignant-chercheur(e)s de plusieurs pays du Sud et de France collaborent

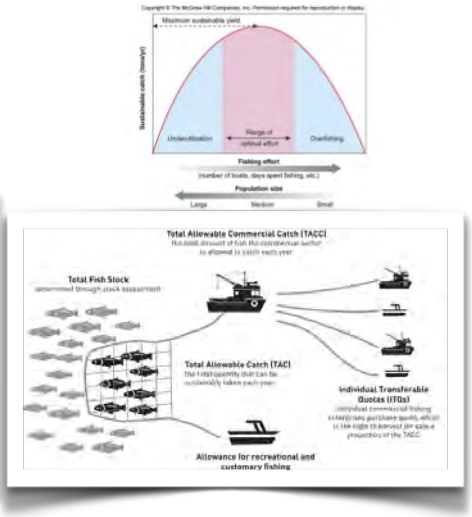
Concevoir des méthodes systémiques dont la finalité est d'être appliquées à des problématiques majeures de développement durable (ODD)



390 million dengue infections/year

Mieux contrôler les épidémies

Réduire la pollution des villes



Atteindre un effort de pêche durable

IRD is a public institute that undertakes

65 research units (jointly with other institutions)

2250 agents and a community of 7000 French researchers

Over 1500 publications/year, 50% in co-publication with partners

ECOBIO Ecology, biodiversity and functioning of continental ecosystems

OCEANS Oceans, Climate and Resources

DISCO Internal dynamics and continents surface

SOC Societies and globalisation

SAS Health and Societies

Ce cours:

Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

III. Exemple de la régression régularisée

IV. Médecine de précision et IA

V. Classeur avec rejet

VI. Conclusion

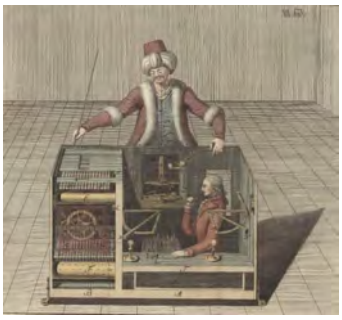
JDZucker

IA & Santé : Historique

1950 1955 1974 1980 2010 2020

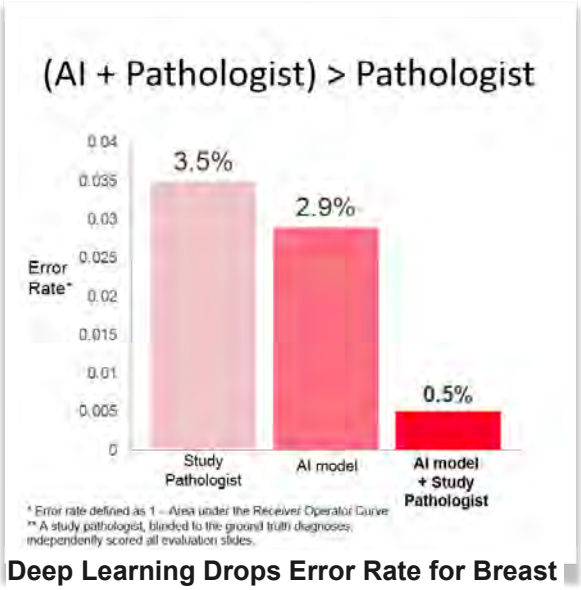
L'Intelligence artificielle (AI) est née au milieu du XXième siècle

« L'intelligence artificielle (IA) vise à reproduire au mieux, à l'aide de machines, des activités mentales, qu'elles soient de l'ordre de la **compréhension**, de la **perception**, ou de la **décision** » www.larousse.fr



1769 : von Kempelen fabrique le Turc mécanique qui joue aux échecs.

1950: le **test** de **Turing**. Un **test** d'intelligence artificielle fondée sur la faculté d'une machine à **imiter la conversation humaine.**

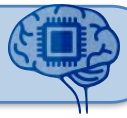


Deep Learning Drops Error Rate for Breast Cancer Diagnoses by 85%

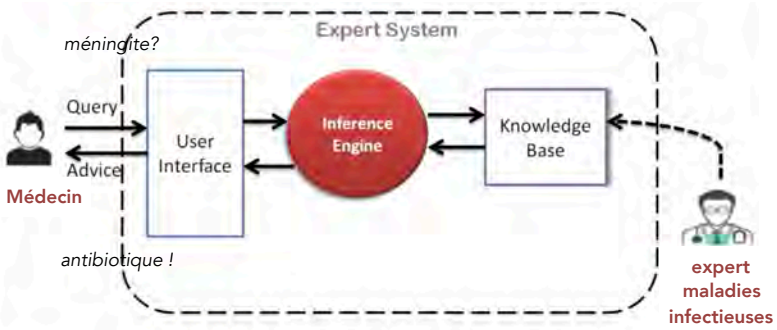
JAMA, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

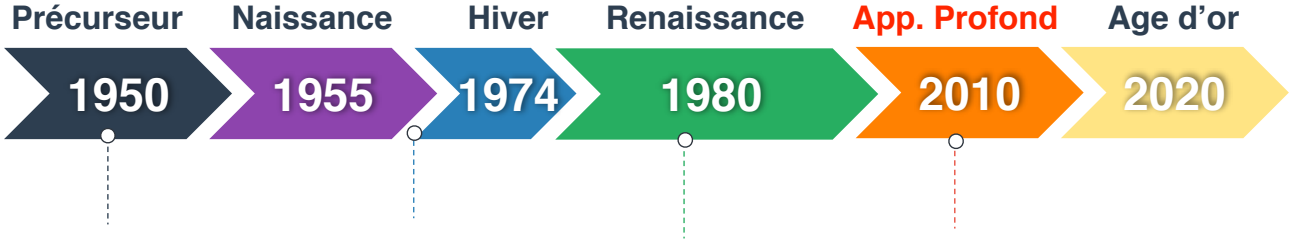
IA et médecine… une longue histoire

Intelligence Artificielle : compréhension/perception/décision



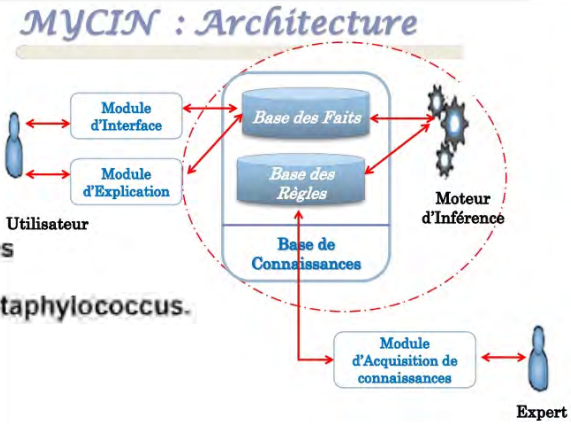
Le système expert MYCIN (1973-)

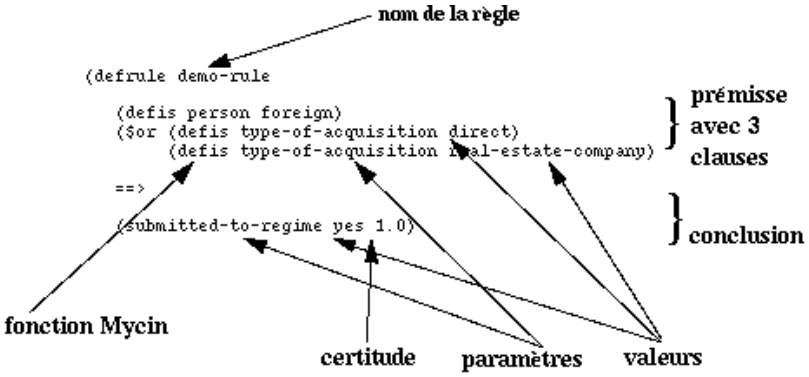


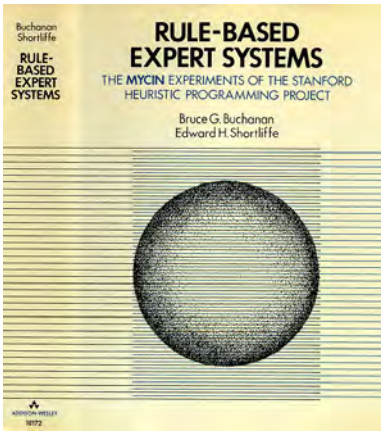


Mycin (1973-)

SI
la coloration de l'organisme est GRAM+
et si la morphologie de l'organisme est Cocci
et si le mode de développement de l'organisme est en colonies
ALORS
il existe une évidence (0.7) que l'identité de l'organisme soit Staphylococcus.



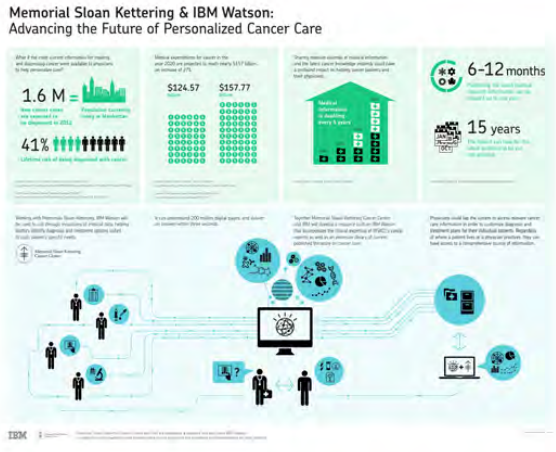




Aujourd'hui IBM a développé Watson

Intelligence Artificielle : compréhension/perception/décision

Watson for Oncology (2013)



https://www.youtube.com/watch?v=UpFHNGF4F8o&pbjreload=10

Précurseur	Naissance	Hiver	Renaissance	App. Profond	Age d'or
1950	1955	1974	1980	2010	2020
Test de Turing	Systèmes experts		Réseaux de neurones	Réseaux convolutifs	

De nombreuses sociétés s'intéressent à l'IA en santé

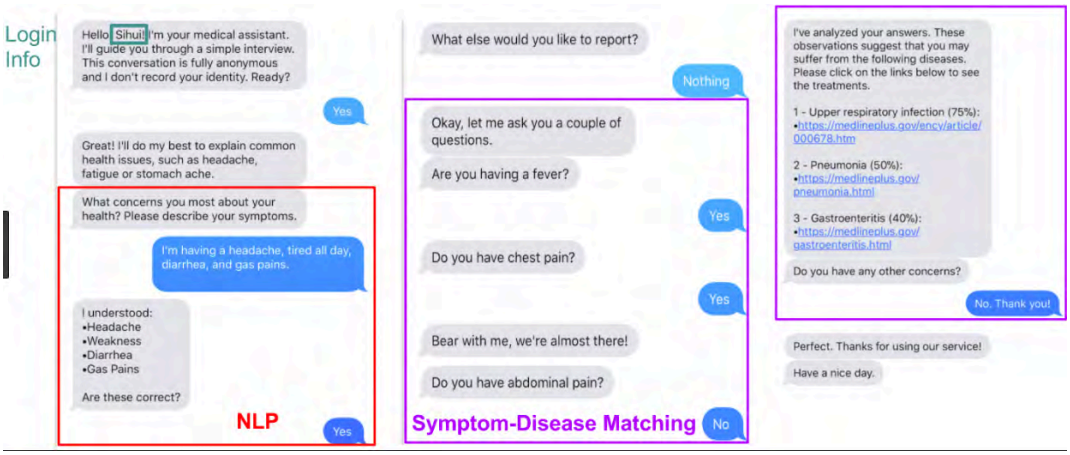
Aujourd'hui Chatterbot et Medecine

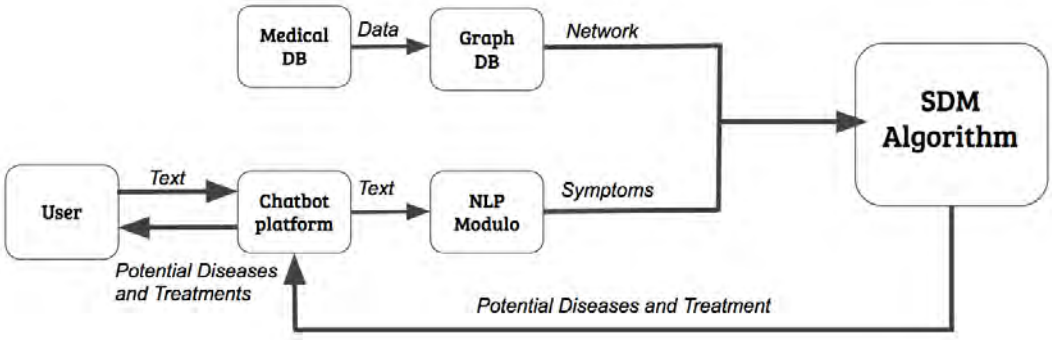
	Your.MD	Sensely	Buoy Health	Infermedica	Florence
Funds raised	$17.3 MM	$11.8 MM	$9 MM	$5 MM	Unspecified
Year founded	2012	2013	2014	2012	2016
HQ location	London, UK	San Fran., CA	Boston, USA	Wroclaw, POL	Dortmund, DEU
Staff size	49	20	23	26	Unspecified
Target user	Patient, Clinician	Patient, Clinician	Patient	Patient, Clinician	Patient
Types of data processed	Chat (text)	Chat (text), image, video	Chat (text)	Chat (text), image, video	Chat (text)
Est. current users	Unspecified	Unspecified	Unspecified	Unspecified	2,000+?

Data collected from Crunchbase and LinkedIn, March 2018

techemergence

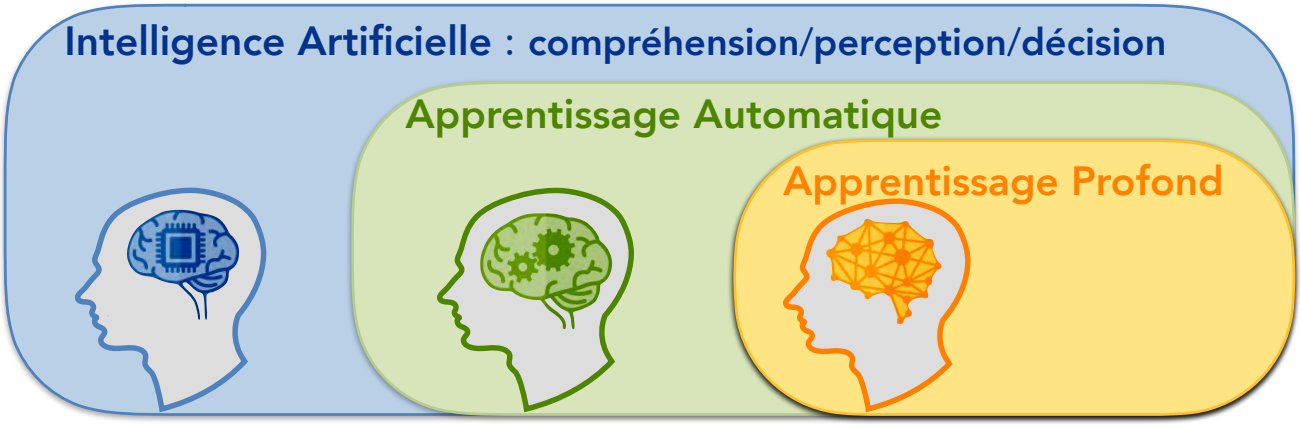
Can a Chat a Day Keep the Doctor Away?





https://chatbotslife.com/can-a-chat-a-day-keep-the-doctor-away-f7539fea289a

Le rôle croissant de l'apprentissage profond en médecine



1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.

1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.

1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.

1967
The Nearest Neighbor Algorithm was written.

1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.

1997
IBM's Deep Blue beats the world champion at Chess.

2002
A software library for Machine Learning, named Torch is first released.

2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go

Dans la médecine, de l'IA venant de 3 grands champs

S. Russell and P. Norvig, The Artificial Intelligence a Modern Approach (AIMA), 3e Preview Edition. Prentice Hall, 2009.

Raisonner/Comprendre

Percevoir/Interagir

Décider/Agir

1er champs de l'IA & Santé:

Raisonner/Comprendre

Systèmes Experts

Coach/Application santé

Google Im2Calories project Compter les calories à partir d'une image !

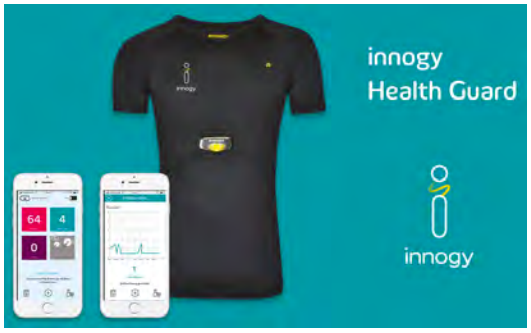
2ème champs de l'IA:

Percevoir/Interagir

Chatterbots médicaux

Membres artificiels

Capteur (e.g. T-Shirt Cardio)

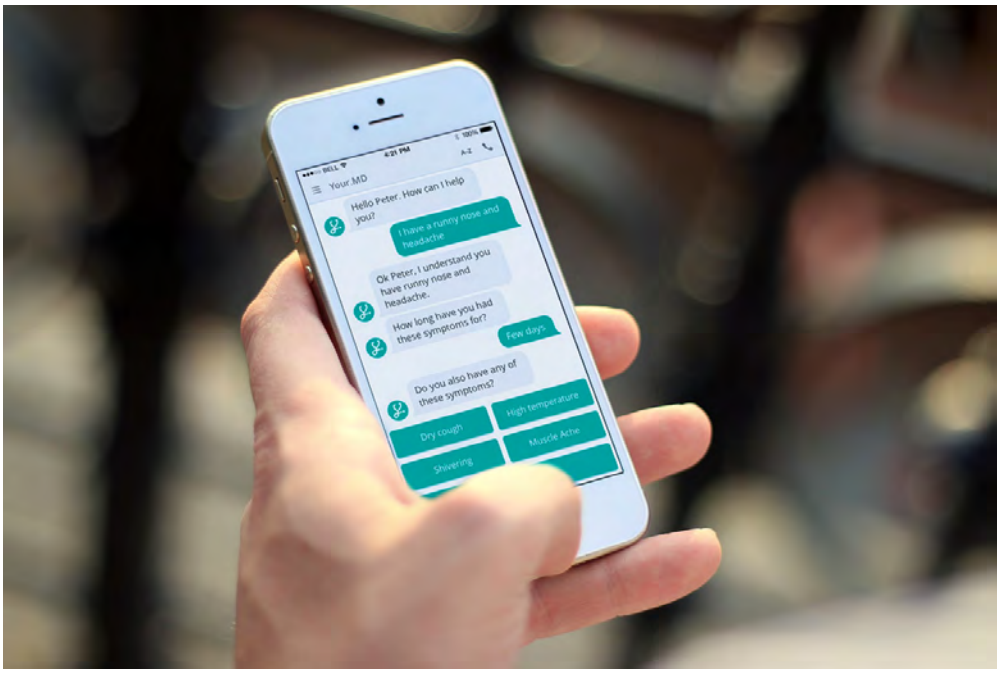


Exosquelette

JDZucker

Les techniques qui vont introduire des ruptures : l'IA joue les docteurs grâce aux **Chatterbots**.

Dans (A routine check–up at the digital medical centre. Microbial Biotechnology 2016) les auteurs imaginent dans un futur proche une discussion entre Mme Tristam-Wellness et un robot Dr. Siri von App



3ème champs de l'IA:

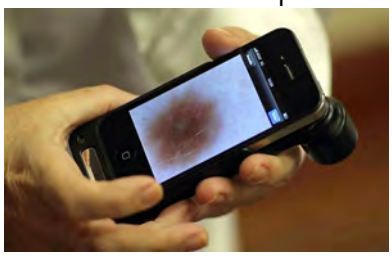
Décider/Agir

Aide au diagnostic

… de plus de 4 000 maladies génétiques, dont beaucoup sont extrêmement difficiles à identifier

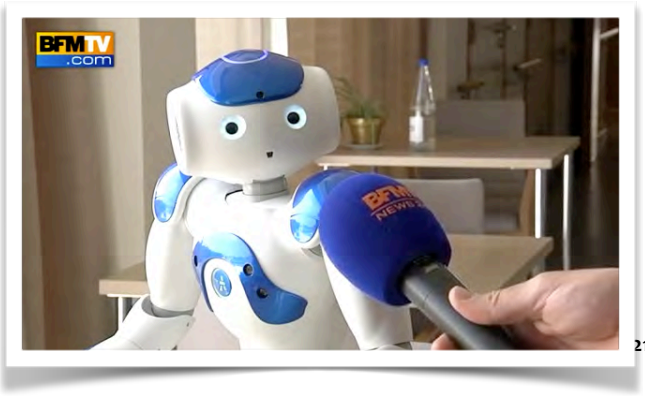
https://www.face2gene.com/

… de cancer de la peau



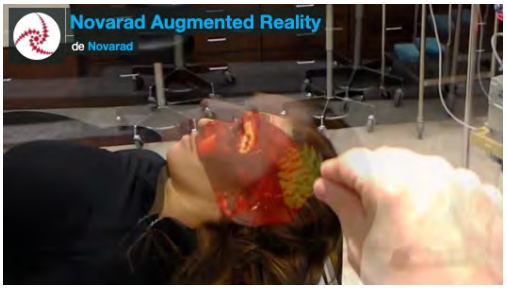
https://fmh-association.org/dermoscreen-application-diagnostiquer-cancer-peau/

Robots aide-soignant



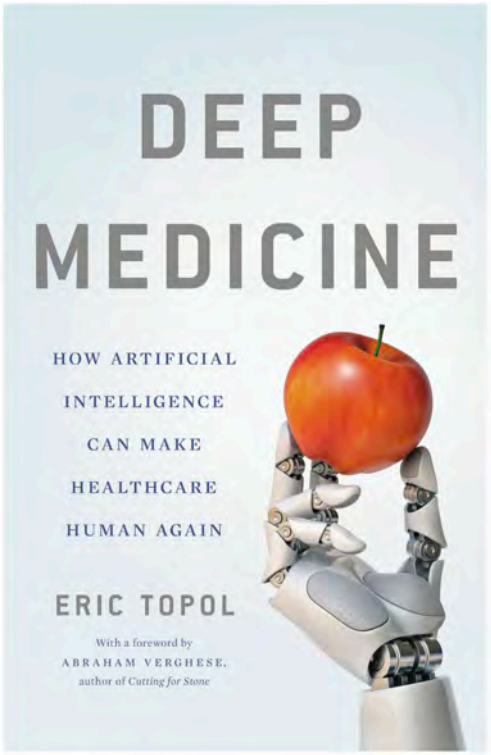
21

Chirurgie assistée par ordinateur



JDZucker

Le rôle croissant de l'apprentissage profond en médecine



JDZucker

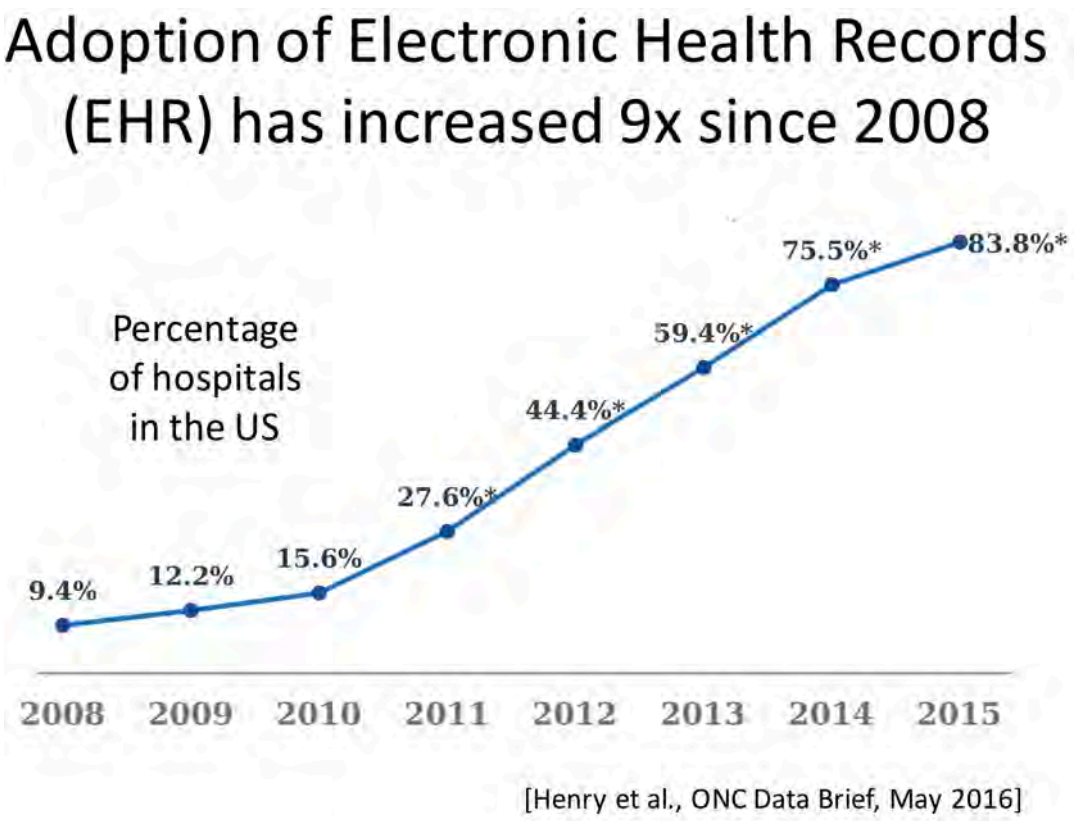
Très prometteuse, en 2019 l'IA en médecine reste superficielle et n'a pas encore tenu ses promesses

FIGURE 3.3: IBM Watson ad.

L'informatique peut faire une grande différence, mais jusqu'à présent→ pas tenu ses promesses.

Difficulté→ l'assemblage et l'agrégation des données a été sous-estimée, non seulement par Watson, mais aussi par toutes les entreprises de technologie qui s'intéressent aux soins de santé.

Et la médecine est en retard



[Henry et al., ONC Data Brief, May 2016]

Et la médecine est en retard

"Bien que la plupart des secteurs du monde soient bien avancés dans la **quatrième révolution industrielle**, qui est centrée sur l'utilisation de l'IA,
la médecine est encore bloquée dans la première phase de la troisième, qui a vu la première utilisation généralisée des ordinateurs et de l'électronique.
Que les fichiers MP3 soient compatibles avec toutes les marques de lecteurs de musique, par exemple, alors que la médecine n'a encore jamais vu de dossiers médicaux électroniques largement compatibles et conviviaux illustre la lutte du secteur pour le changement. » Eric Topol Deep Medicine 2019

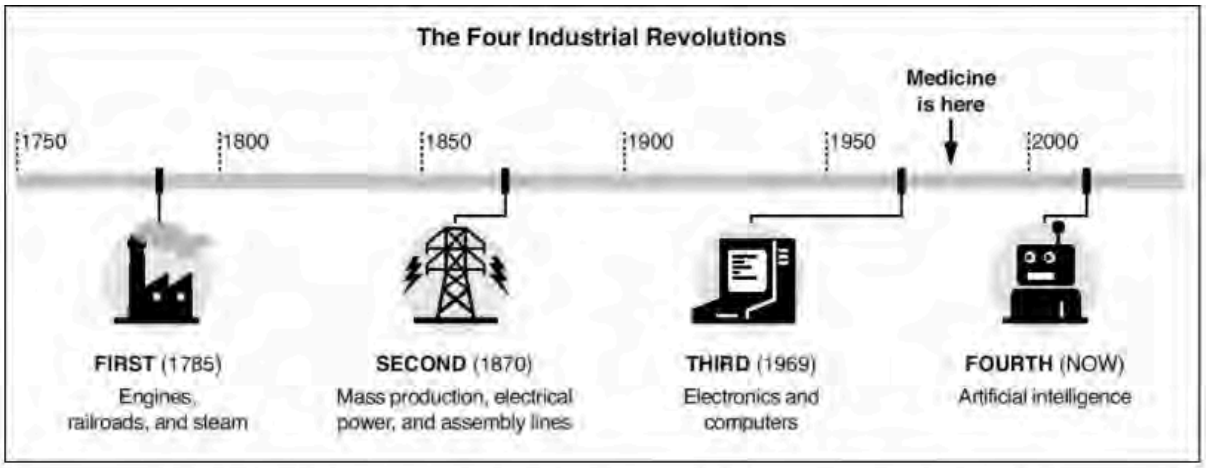
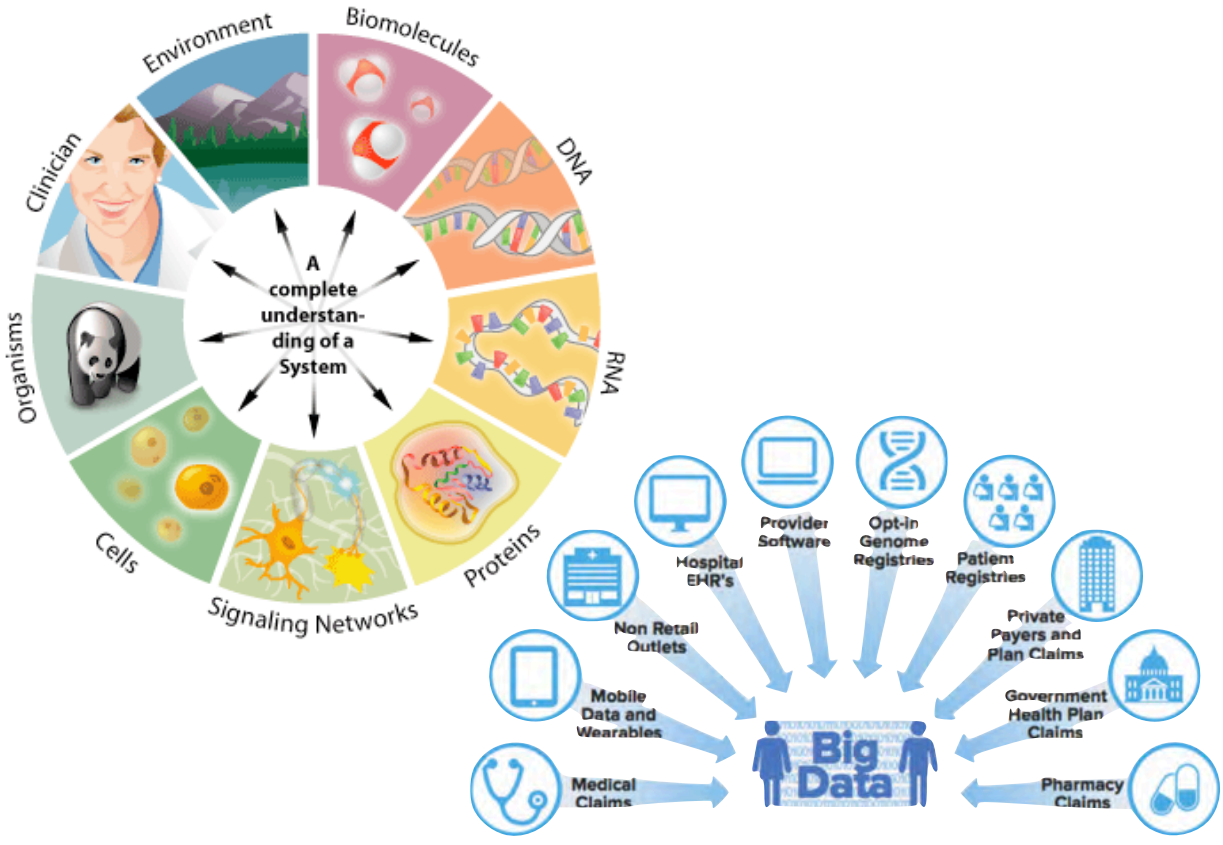


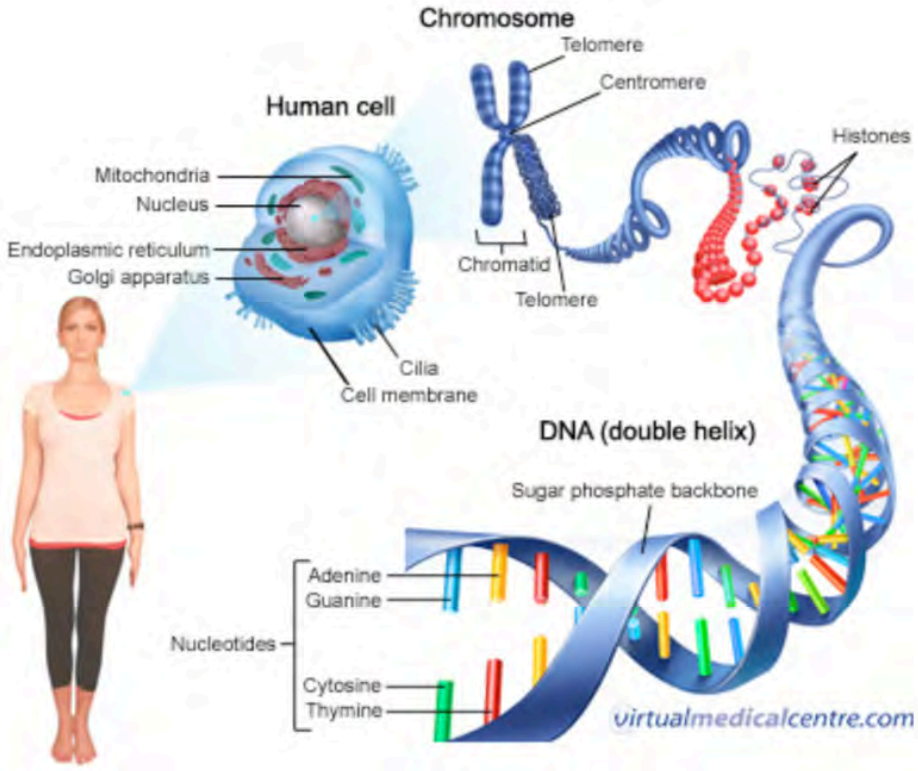
FIGURE 1.4: The four Industrial Revolutions. Source: Adapted from A. Murray, "CEOs: The Revolution Is Coming," *Fortune* (2016): http://fortune.com/2016/03/08/davos-new-industrial-revolution.

IA & Santé : Types de données



Une grande diversité de données

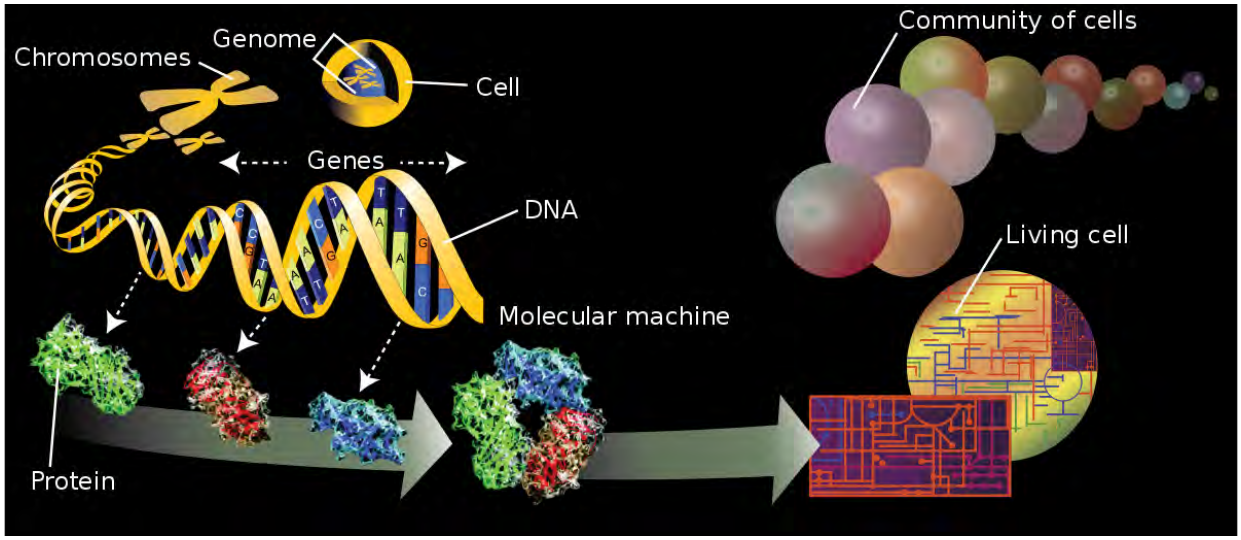
De l'ADN à la vie chez l'humain



1 corps = 10^{14} cellules humaines (et 100 x plus non humaines)

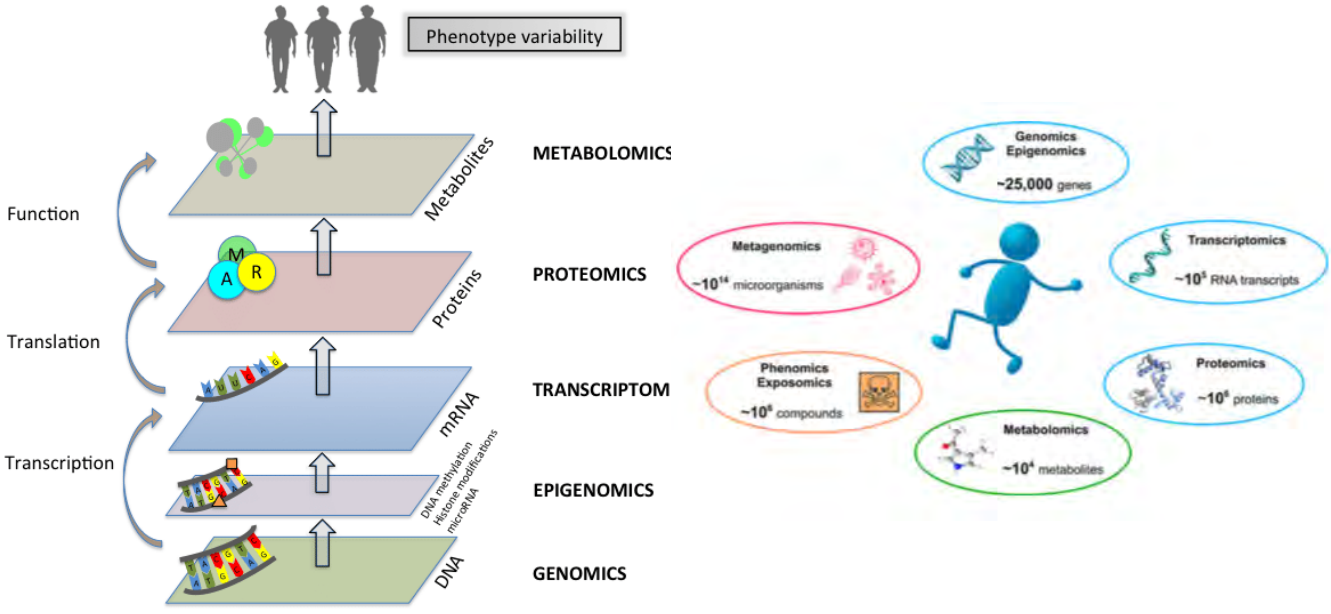
1 cellule = 6×10^{9} ACGT codant pour 20 000+ gènes

Quantification des omes→ omiques



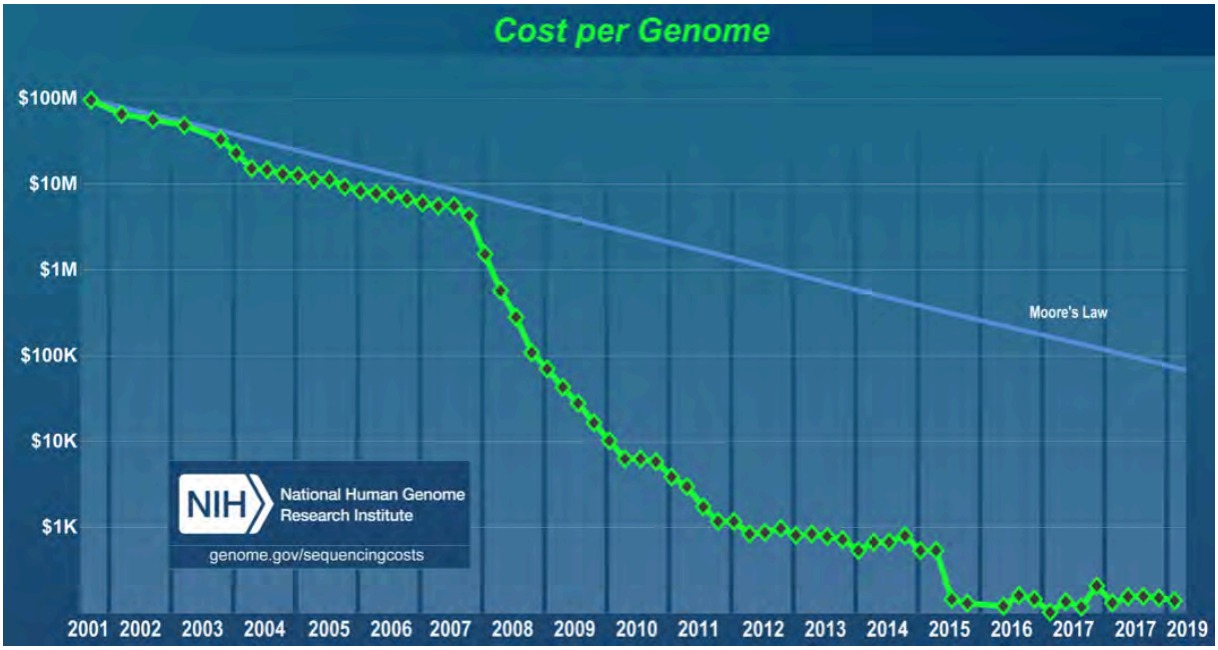
Le néologisme omics de langue anglaise fait référence de façon informelle à un domaine d'étude en biologie se terminant par -omique, comme la génomique, la protéomique ou la métabolomique.
Omics vise la caractérisation et la quantification collectives de pools de molécules biologiques qui se traduisent par la structure, la fonction et la dynamique d'un organisme ou d'organismes. Wikipedia

Les données « Omics » au phenotypage profond



L. Yu, K. Li, and X. Zhang, "Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review.," Oncotarget, 8(10) 2017.

Coût des génomes



Cost Per Genome

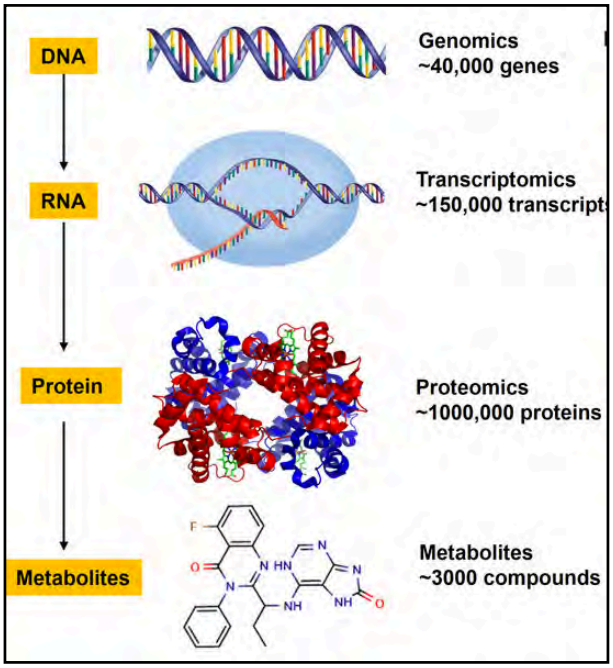
$3,000,000,000	Human Genome Project	2003
$20,000,000	1st individual genome	2006
$2,000,000	1st NGS Genome	2007
$200,000	1st 30x genome	2008
$10,000	1st sub-10K genome	2010
$1,000	1st $1,000 genome	2014
	1st $200 genomes	2018

https://www.cnbc.com/2019/07/01/for-600-veritas-genetics-sequences-6point4-billion-letters-of-your-dna.html

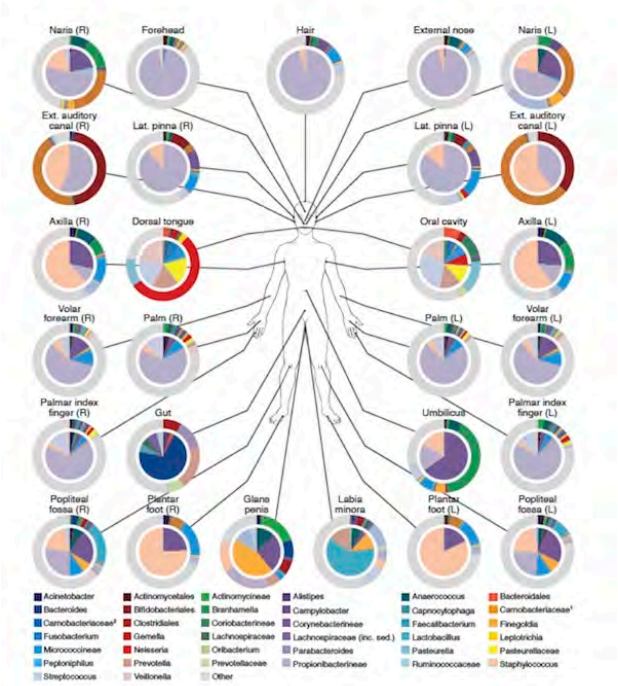
1st $100 genomes 202?

Les données « Omics » permettent de nous caractériser très finement, nous et… nos hôtes.

Analyser nos propre cellules



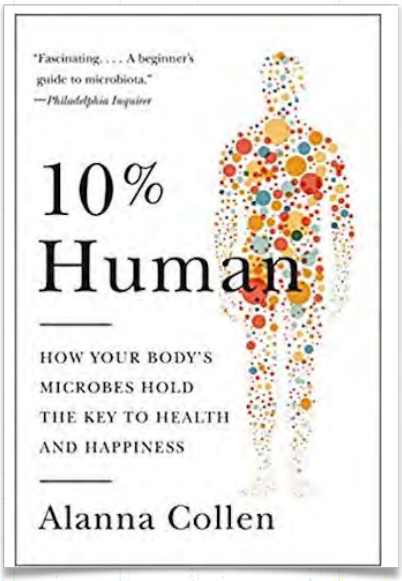
Analyser nos bactéries

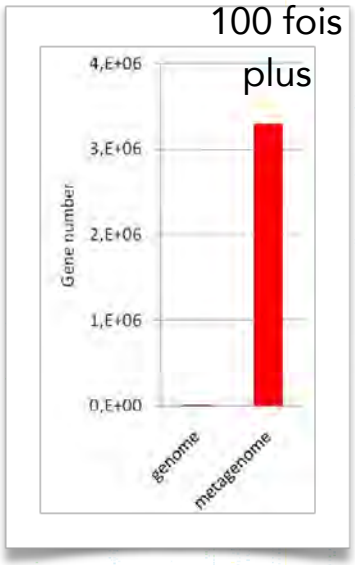


F. S. MD, et al., "Translational Radiomics: Defining the Strategy Pipeline and Considerations for Application-Part 1: From Methodology to Clinical Implementation," Journal of the American College of Radiology, 2018

Microbiote intestinal humain : un organe oublié

Du bébé "stérile" à la naissance → 2 kg de micro-organismes,
sur les 100 billions de cellules du corps humain, seule 1 sur 10 est humaine.





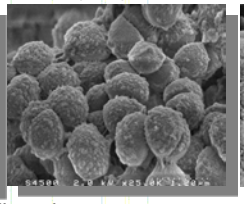
Métabolisme
production de vitamines
dégradation des aliments
extraction énergétique

Système immunitaire
l'"éducation" des défenses immunitaires innées

Quantification de notre microbiome

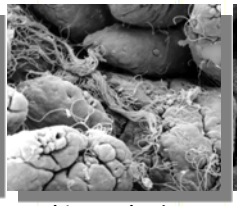
La plupart des micro-organismes sont inconnus et non cultivables…

Faecalibacterium prausnitzii



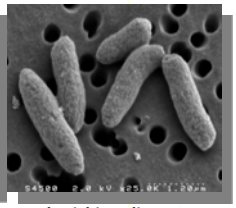
Ruminococcus spp

Clostridium difficile en caecum souris



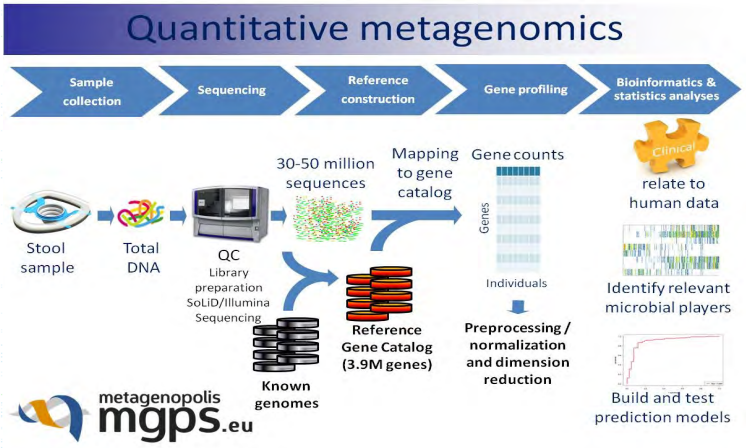
Bactéries ancrées dans une Plaque de Peyer, Intestin de souris

Bacteroides dorei



Escherichia coli

Photos UEPSD

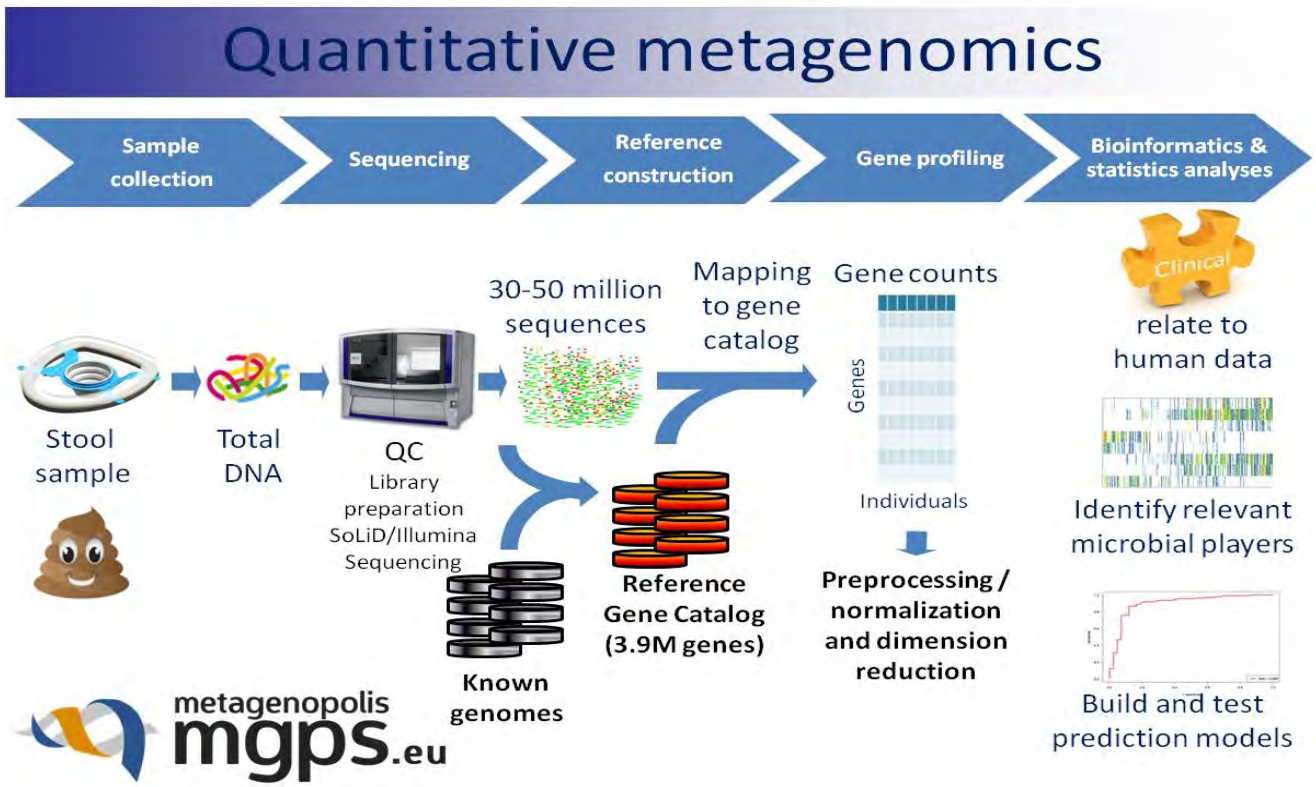


A Powerful Microscope to Scan the neglected organ

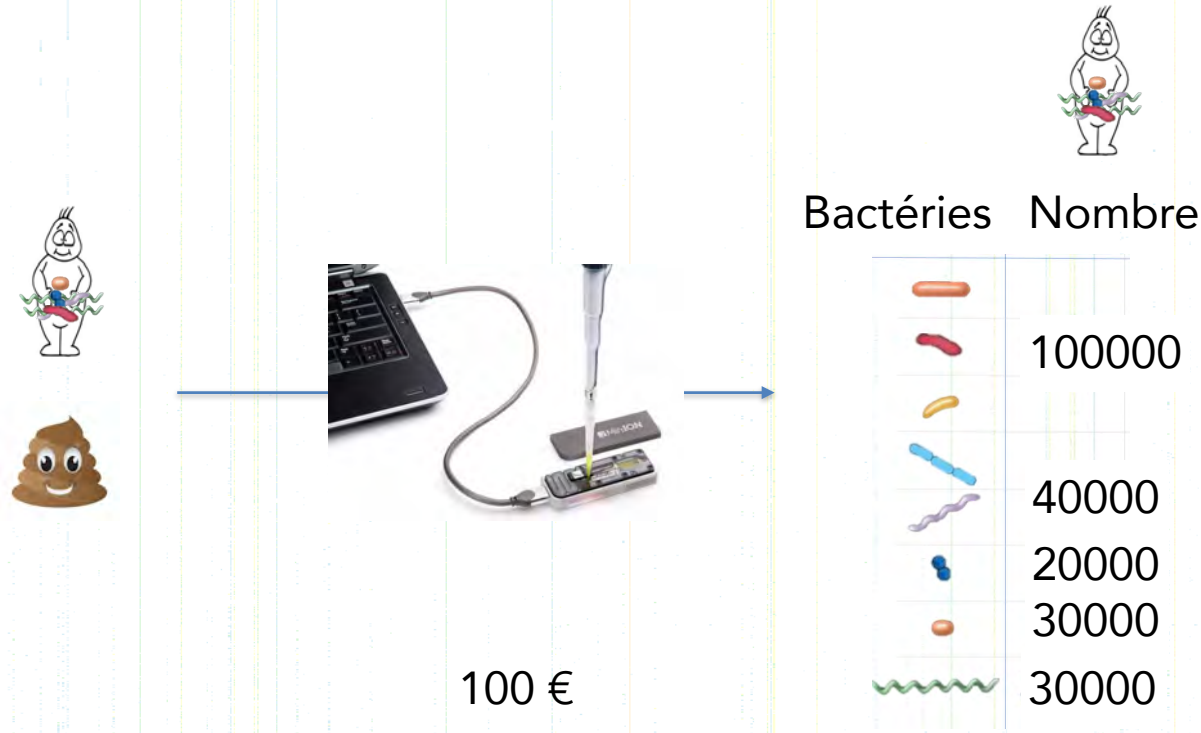
Ehrlich SD ©

Quantification de notre microbiome

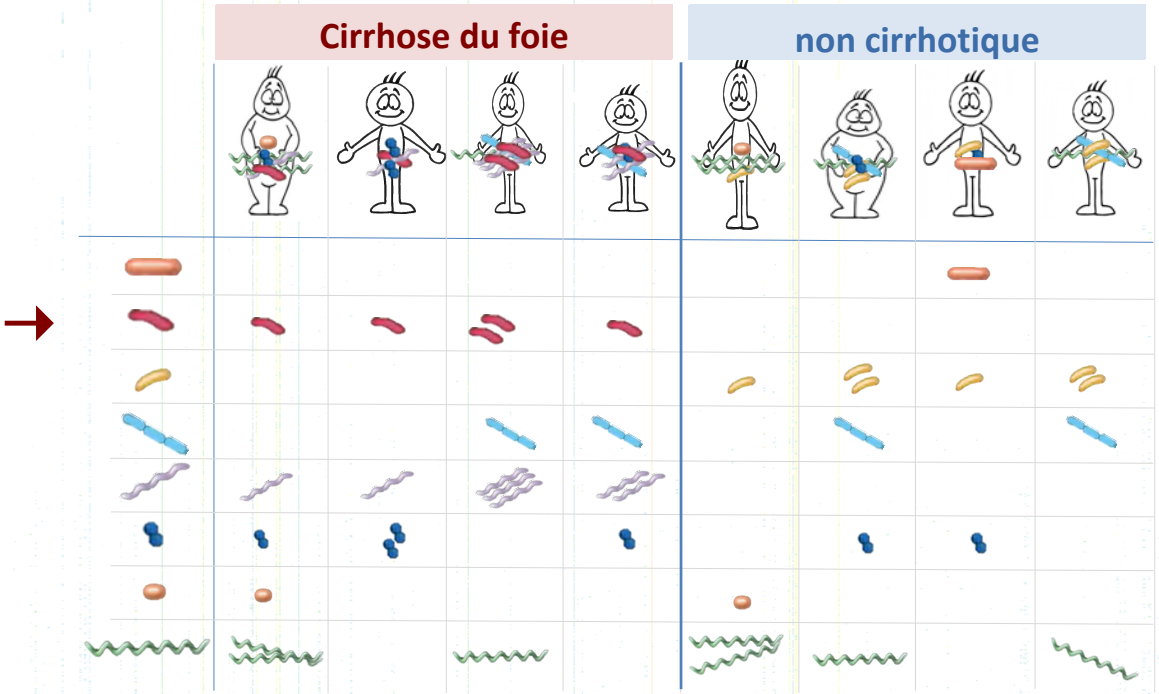
La plupart des micro-organismes sont inconnus et non cultivables…



Quantification de notre microbiome (vision simple)



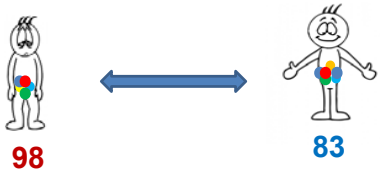
Vers une médecine de précision des maladies intégrant la métagénomique.



Signature géniques impliquant plusieurs gènes bactériens parmi des **millions**…

Le microbiote intestinal comme moyen de diagnostic

Cirrhose du foie

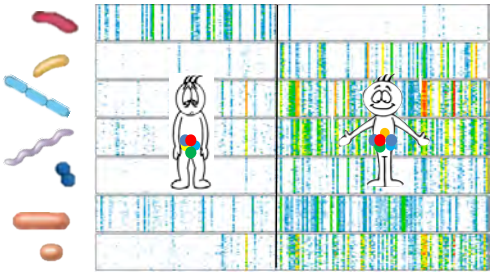


Diagnostic actuel:
- biopsie 40%
- symptômes cliniques ou imagerie 60%

↓

Apprentissage Automatique

↓

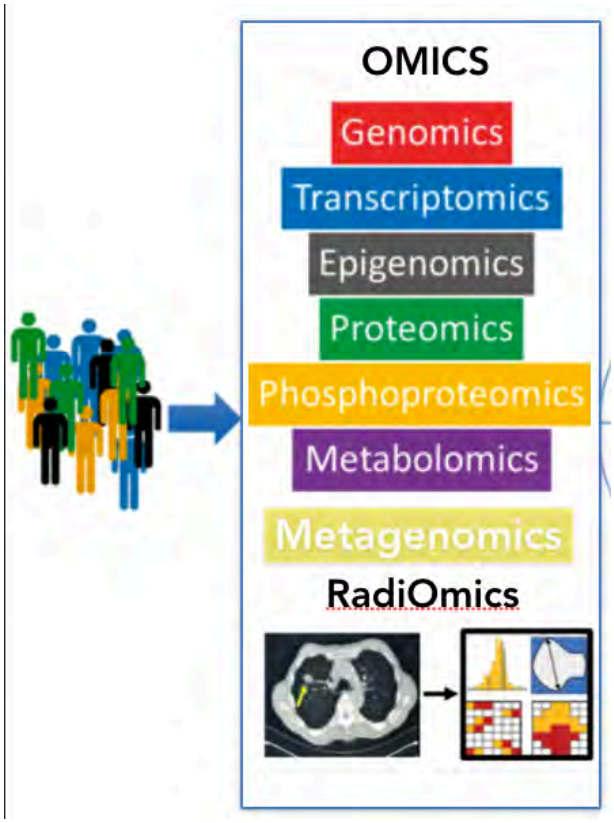


Modèle prédictif à partir de sept espèces

Diagnostic très performant (AUC=94%), **non invasif**

Qin N. et al. Nature 2014

Des données massives



Both N and p are two important characteristics of the Data Mining task we address...

p attributes/features/variables

Example	Att1	Att2	Att3	Att4	Att5	Att6	Att7	Att8	Att9	Att10

N

instances/input vectors/ observations/ examples/ records/ ...

... but N/p is another key characteristic. In UCI repository the Median is 42 → [0,0032 , 990114]

UCI Machine Learning Repository

THE UNIVERSITY of EDINBURGH School of informatics

YouTube COMEDY WEEK

Slam Preference Data Data Set

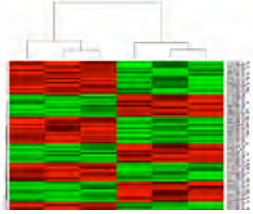
	Left	Right	More Fun
t1	sNabaB-eb3Y	wHkPb68dxEw	left
t2	sNabaB-eb3Y	y2emSXSE-N4	left
t3	fY_FQMQpjok	sNabaB-eb3Y	left
t4	Vr4D8xO2lBY	sNabaB-eb3Y	right
t1	sNabaB-eb3Y	dDtRnstrefE	left
t2	wSMh4E3xxdw	fY_FQMQpjok	right
t3	fY_FQMQpjok	No-ZR7-X76s	left
t4	HZPUQQNRvOg	2ruyMxxfAJg	right
t1	vX95JgKGu0o	wu4SU70w7LA	left
t2	sNabaB-eb3Y	2FHH-9teZP0	left
t3	sP3YdfG_Uio	fY_FQMQpjok	right
...	...	...	...
t912967	oPDM1BjwZg8	Z2cpqRpiTqk	right
t912968	6utE5APd4wA	t4KvXsglvMg	right
t912969	W9y6nwBwwyQ	t4KvXsglvMg	right

N =1138562
p = 3
N/p =379,521

Iris Dataset

N =150
p = 4
N/p =38

Leukemia Dataset



N = 72
p = 7729
N/p =0,009=1%

… but N/p is another key characteristic.
In UCI repository the Median is 42 → [0,0032 , 990114]

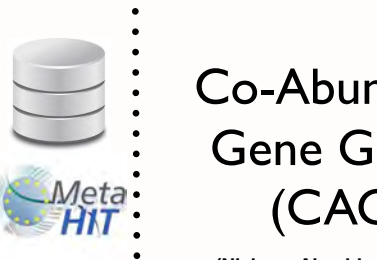
Integrated Gene Catalalog (IGC)

Li et al. 7/6/2014)

N =1267
p = 9,879,896
N/p =0.00013

Co-Abundance Gene Groups (CAGS)

(Nielsen, Almeida & al., 7/6/2014)



N =49
p = 7381
N/p =0.0066

MetaGenomics Species (MGS)

N = 49
p = 741
N/p =0.067

Precision Medicine from Metagenomics requires scaling up Data Mining algorithms

- **Learning prognostics/diagnostics models** ? estimate their performances ?
- **reconstructing (causal) networks between genes or Metagenomic Species** ?

Clinical Data
p=100
$p^2=10^6$

Transcriptomics Data
p=10,000
$p^2=10^8$

Metagenomics Data
p=10,000,000
$p^2=10^{14}$

Des données "massives" pour décrire chaque patient - Big Data

Données Clinique
p=100
$p^2=10^6$

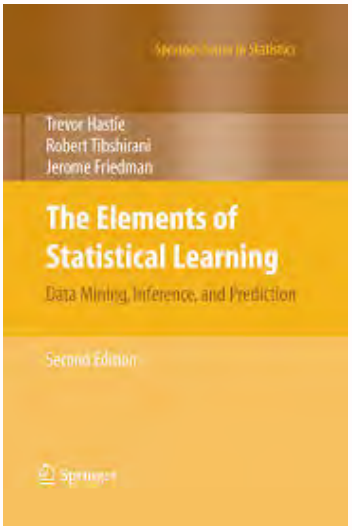
Données
Transcriptomiques
p=10,000
$p^2=10^8$

Données
Métagénomiques
p=10,000,000
$p^2=10^{14}$

Also referred as High-Dimensional Problems: p ≫ N

New chapter in the second edition of The Elements of Statistical Learning [Hastie et al., 2009]

High-Dimensional Problems: p ≫ N



- **Hard problems** because the required number N is about the VCDim of the concept space
- If concepts are linear then VCDim=#attributes=p and we need $N \geq p$ ($N/p \geq 1$) to learn...

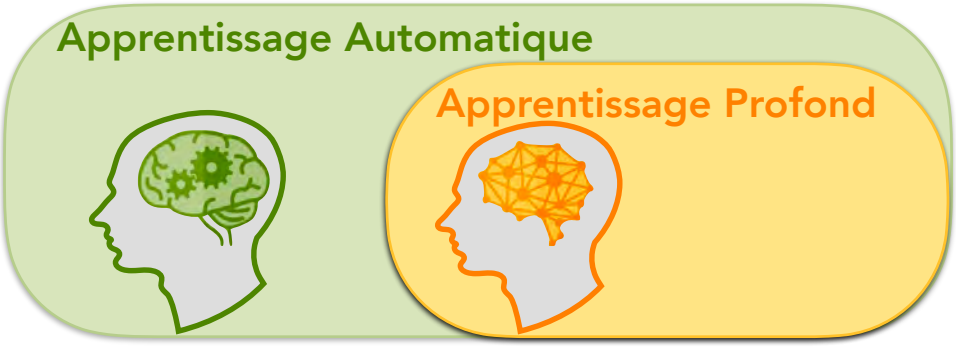
Problèmes liés à l'apprentissage en santé

☑ la quantité de données
☑ l'intégration de données
☑ **l'anonymisation des données**

☑ la compréhension des biais des algorithmes (fairness)
☑ **la robustesse des algorithmes**
☑ **l'efficacité et les garanties des algorithmes sur des problèmes n << p**

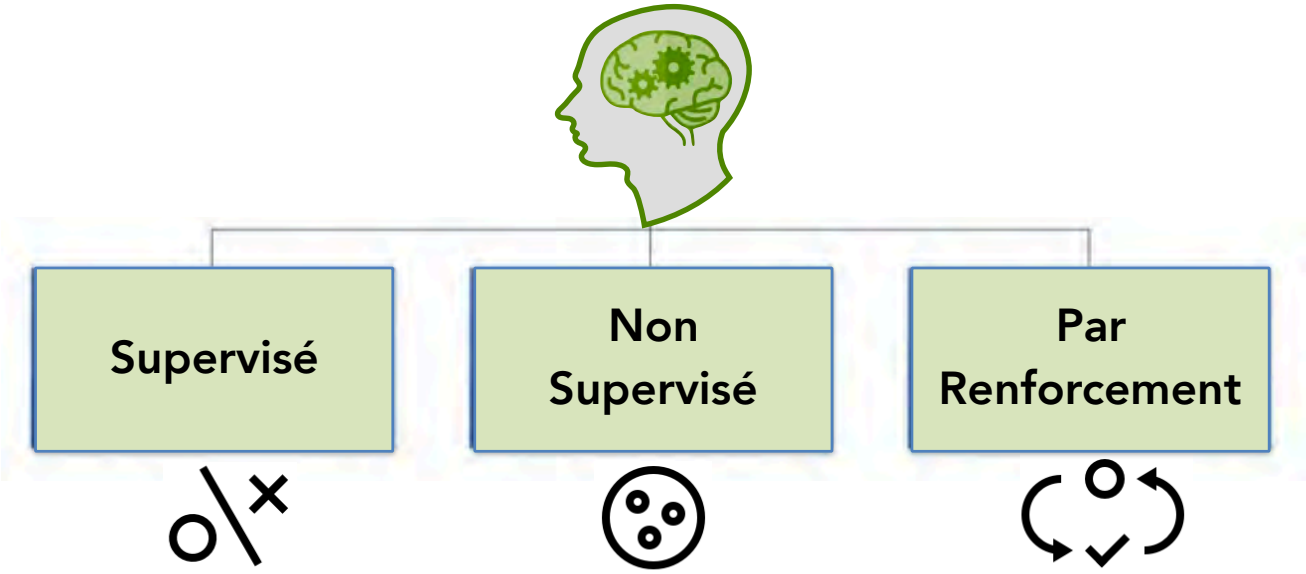
☑ **l'interpretabilité des modèles**
☑ **la prise en compte du coût dans les modèles…**

IA & Santé : l'importance de l'apprentissage



Décliner les trois domaines du Machine Learning en santé

Les 3 grandes tâches du Machine Learning



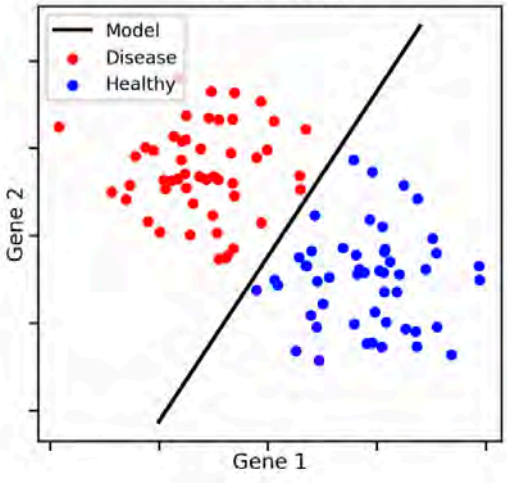
Apprendre à partir d'exemples

Apprendre à organiser des observations

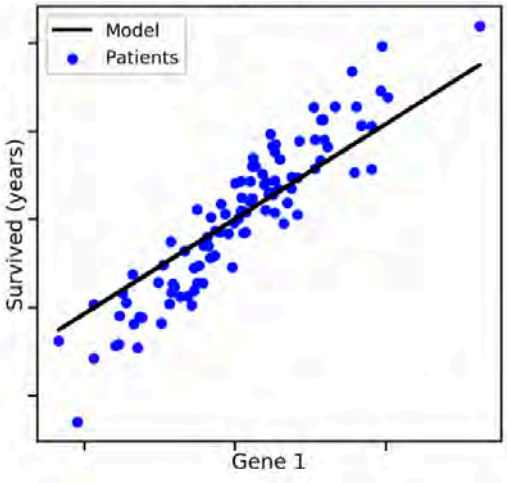
Apprendre à décider par essais/erreur

Apprentissage supervisé : à partir de données géniques

Classer des patients selon l'expression de leur gènes



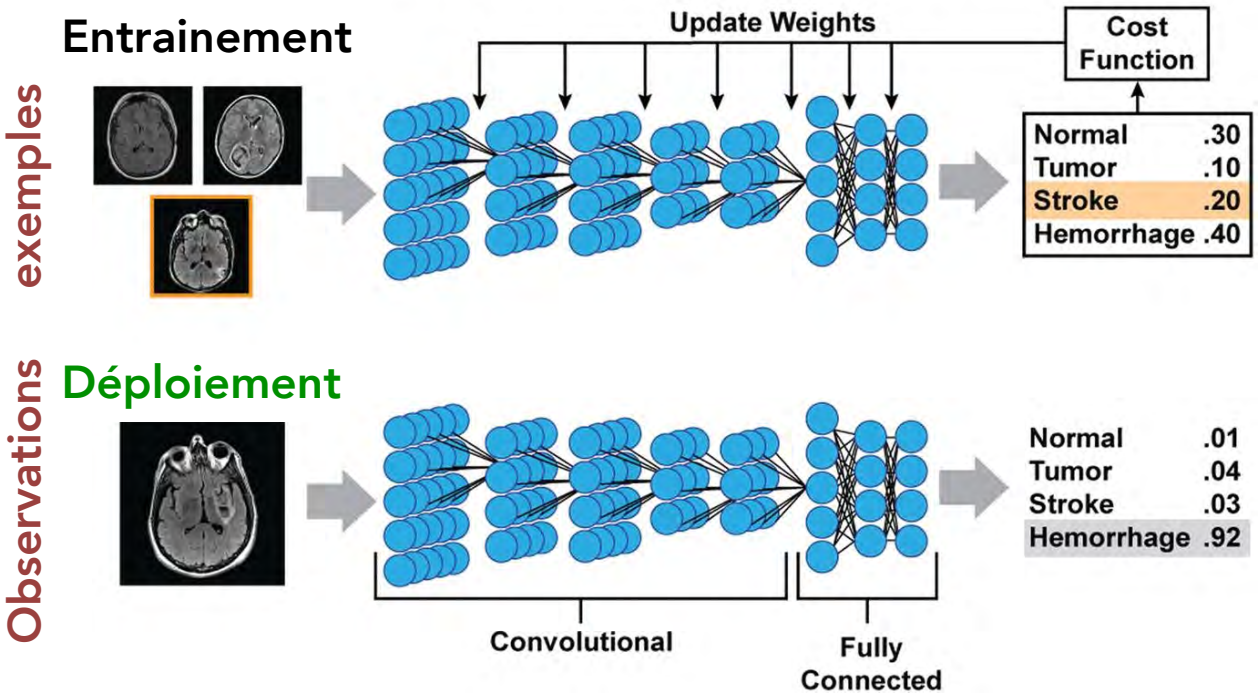
Prédire la survie des patients en fonction de l'expression d'un gène



Séparateur linéaire
SVM Gradient Boosting RF

Régression linéaire
Régression logistique
Régression polytonique

Apprentissage supervisé : à partir d'images

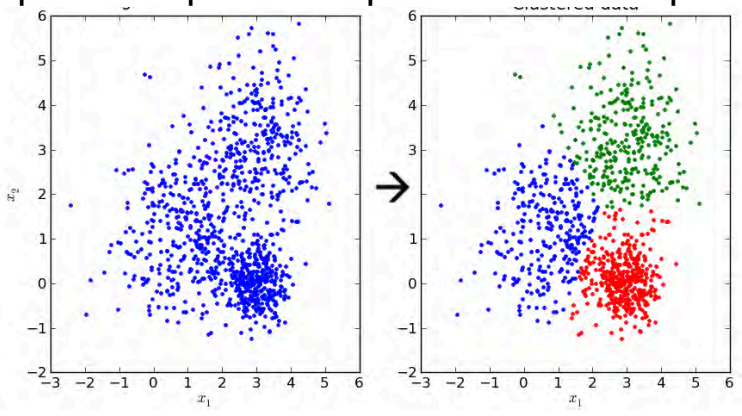


Zaharchuk et al. "Deep Learning in Neuroradiology," AJNR Am J Neuroradiol, (39)10. 2018.

JDZucker

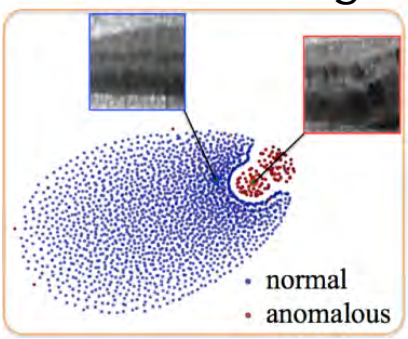
Apprentissage non supervisé: clinique, omique ou images

Regrouper des patients proches: exemple k-means



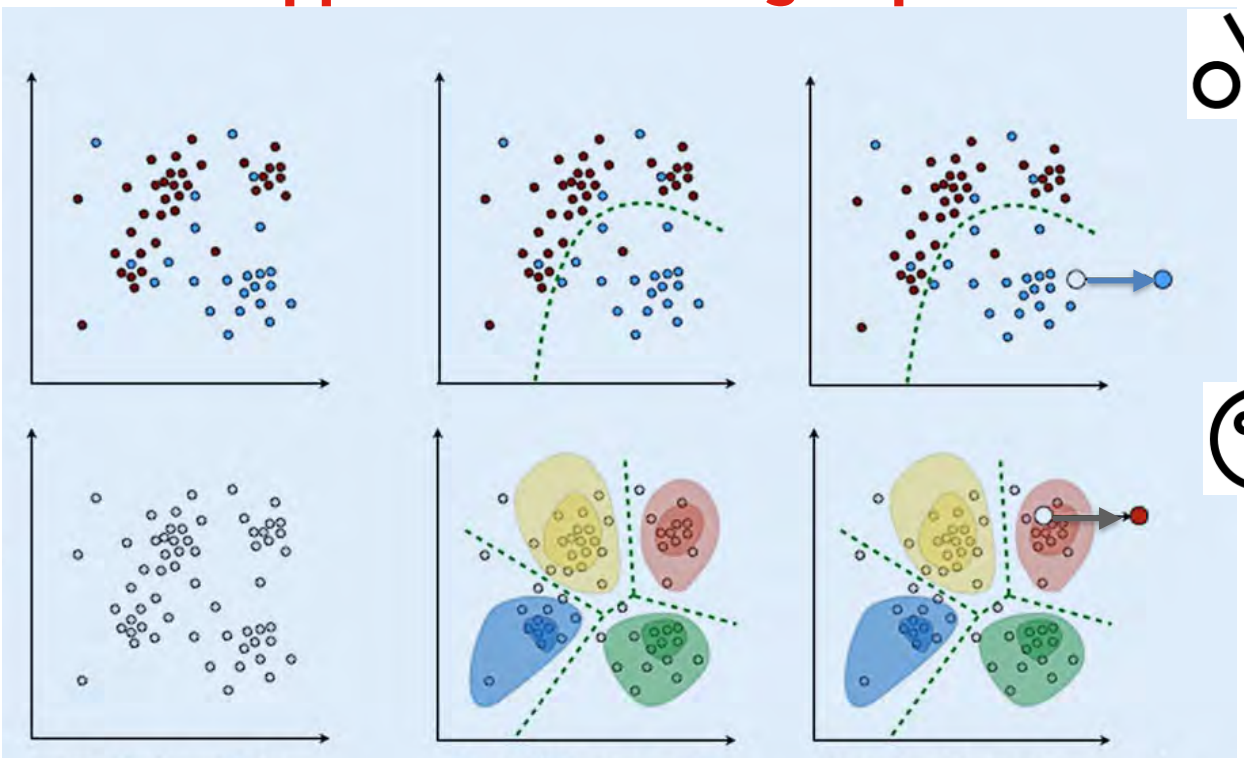
k ?

détecter des anomalies d'images d'examens



seuil ?

Utilisation des modèles pour prédire une classe ou l'appartenance à un groupe

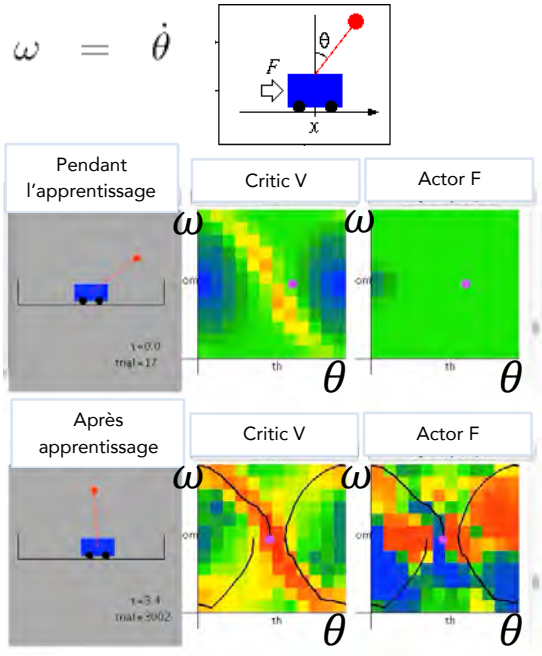


Données patients | Modèles | Utilisation du modèle sur nouveau patient

S. Röhrich, et al. "Machine learning: from radiomics to discovery and routine,"2018.

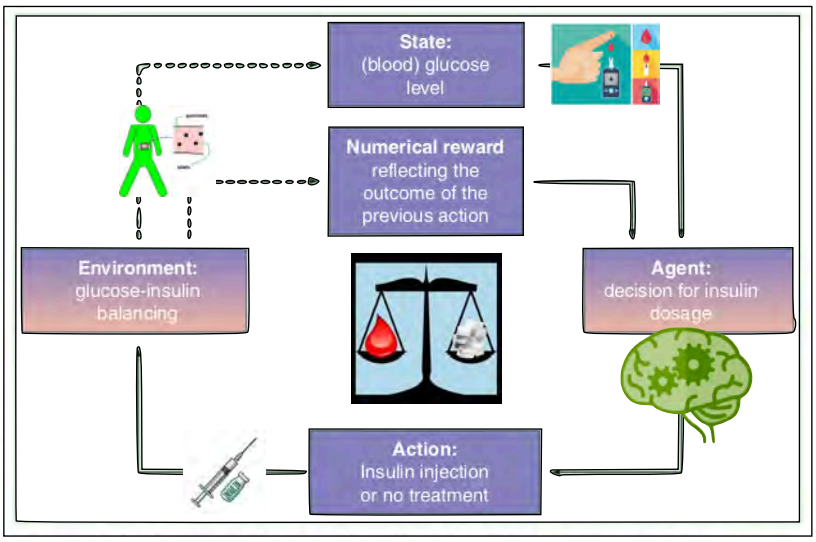
Apprentissage par renforcement : programmer le contrôle de l'injection optimale d'insuline

Quelle force F pour maintenir l'équilibre ?



https://brain.cc.kogakuin.ac.jp/~kanamaru/NN/CPRL/

blue: negative, green: 0, red: positive.



Bothe et al., "The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas," Expert Review of Medical Devices, (10)5,2014.

https://towardsdatascience.com/a-review-of-recent-reinforcment-learning-applications-to-healthcare-1f8357600407

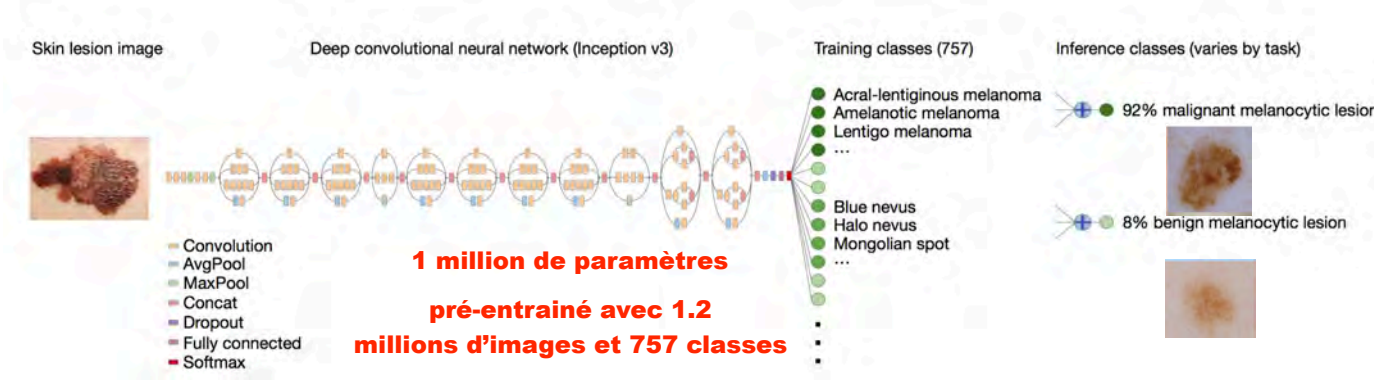
3 predictions in **2016** on Machine Learning as a disruptive technology for Medicine in the next few years.

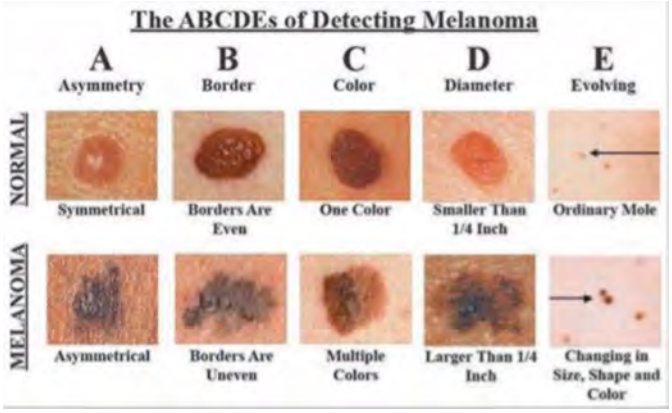
Obermeyer, Z. & Emanuel, E.J., **2016**. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, 375(13), pp.1216–1219.

- First, ML will **dramatically improve prognosis**. We can precisely identify large patient subgroups with mortality rates approaching 100% and others with rates as low as 10%.
  prediction → come into use in the next 5 years.
- Second, ML will **displace much of the work of radiologists and anatomical pathologists**. Algorithms will also monitor and interpret streaming physiological data, replacing aspects of anesthesiology and critical care.
  prediction → disruptions is within next years, not decades.
- Third, ML will **improve diagnostic accuracy**.
  Obstacles: a) gold standard for diagnosis unclear → harder to train algorithms. b) high-value EHR data are often stored in unstructured formats c) models need to be built and validated individually for each diagnosis.

  prediction → to develop, over the next decade.

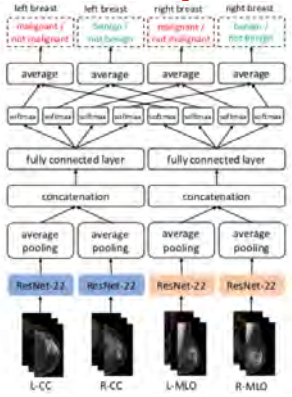
Classification des cancer de la peau du niveau d'un expert dermatologue (Nature,2017)





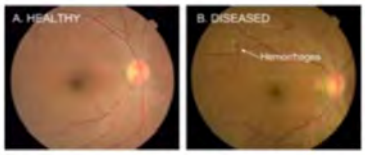
A. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, 2017.

Les performances des « IA » dépassent régulièrement celles des radiologues et anatho-pathologistes (2016-2019)



- diagnostiquer les cancers du sein mieux que les radiologues (Nan Wu, et al. 2019). Trained and evaluated on over 200,000 exams (over **1,000,000 images**). AUC of 0.895/

- diagnostiquer la rétinopathie diabétique comme les **ophtalmologistes** (Gulshan,JAMA,2016.) **128 000** images



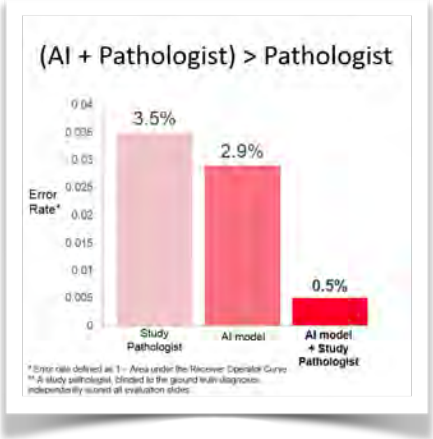
Caveat

- To avoid « **disillusionment** » a stronger appreciation of the **technology's capabilities** and limitations is needed.
- **Combining** machine-learning software with the **best human clinician "hardware"** will permit delivery of care that outperforms what either can do alone.

Chen, J. H. & Asch, NEJM 376 (2017).

Des systèmes basés sur une collaboration homme-machine peuvent faire mieux que l'IA seule…

https://siecledigital.fr/2018/10/15/une-ia-by-google-detecte-le-cancer-avec-une-precision-de-99/

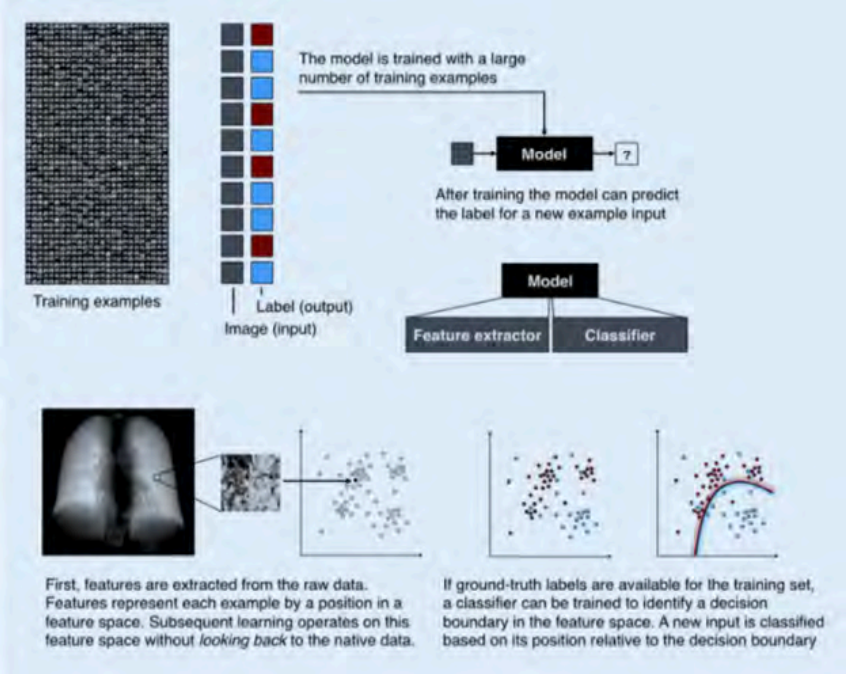


Deep Learning Drops Error Rate for Breast Cancer Diagnoses by 85%

JAMA, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

The explosion of medical imaging data creates an environment ideal for machine-learning and data-based science

Radiomics, the high-throughput mining of quantitative image features from standard-of-care medical imaging that enables data to be extracted and applied within clinical-decision support systems (CDSS) to improve **diagnostic, prognostic, and predictive accuracy**, is gaining importance in **cancer research.**



S. Röhrich,"Machine learning: from radiomics to discovery and routine," 2018.

L'approbation des usages médicaux de l'IA est en marche… forte des performances en prédiction…

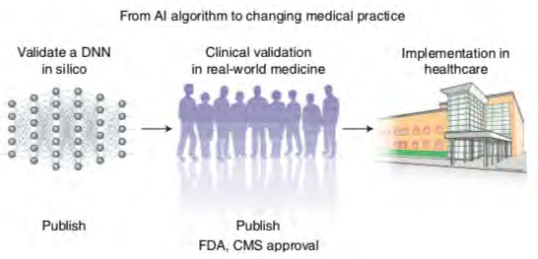


Table 1 | Peer-reviewed publications of AI algorithms compared with doctors

Specialty	Images	Publication
Radiology/ neurology	CT head, acute neurological events	Titano et al.
Pathology	Breast cancer	Ehteshami Bejnordi et al.
	Lung cancer (+driver mutation)	Coudray et al.
	Brain tumors (+methylation)	Capper et al.
Dermatology	Skin cancers	Esteva et al.
	Melanoma	Haenssle et al.
	Skin lesions	Han et al.
Ophthalmology	Diabetic retinopathy	Gulshan et al.
Gastroenterology	Polyps at colonoscopy*	Mori et al.
	Polyps at colonoscopy	Wang et al.
Cardiology	Echocardiography	Madani et al.
	Echocardiography	Zhang et al.

Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat Med, pp. 1–13, Jan. 2019.

Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

III. Exemple de la régression régularisée

IV. Médecine de précision et IA

V. Classeur avec rejet

VI. Conclusion

RGPD et modèles interprétables : droit et confiance

- Règlements de l'UE (Règlement général sur la protection des données (GDPR) en vigueur le 25 mai 2018) sur la prise de décision algorithmique et un "droit d'explication".

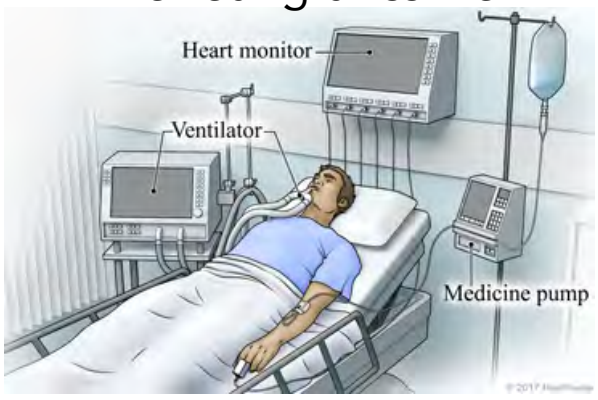
Goodman, B. & Flaxman, S. R. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". AI magazine, 2017

- Une explication de la prédiction est désirée par médecins et patients lorsque un **modèle** doit être validé avant d'être déployé en routine→ **confiance**

Vanthienen, et al. Performance of classification models from a user perspective. *Decision Support Systems* **51**, 782- 793,(2011).

erreur pas (trop) grave…

erreur gravissime…



A. Vellido, et al., "Machine learning in critical care: state-of-the-art and a sepsis case study," BMEO,2018.

Equité/Fairness : l'IA est biaisée par les données

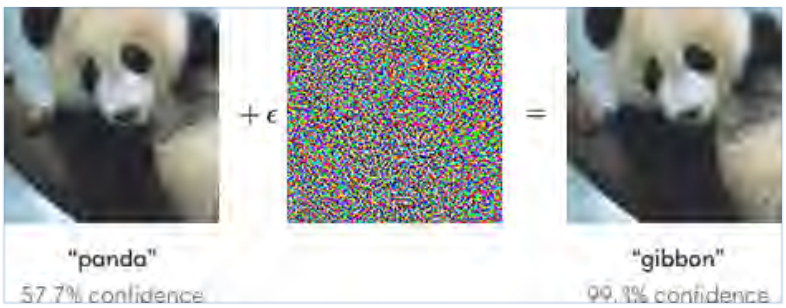
Une étude récente a révélé que certains programmes de reconnaissance faciale **classent incorrectement moins de 1 % des hommes à la peau claire**, mais plus d'**un tiers des femmes à la peau foncée**.

Que se passe-t-il lorsque l'on se fie à de tels algorithmes pour diagnostiquer le mélanome sur une peau claire ou foncée… ?

Un programme apprends à partir des données qu'on lui donne et qui peuvent être … biaisées

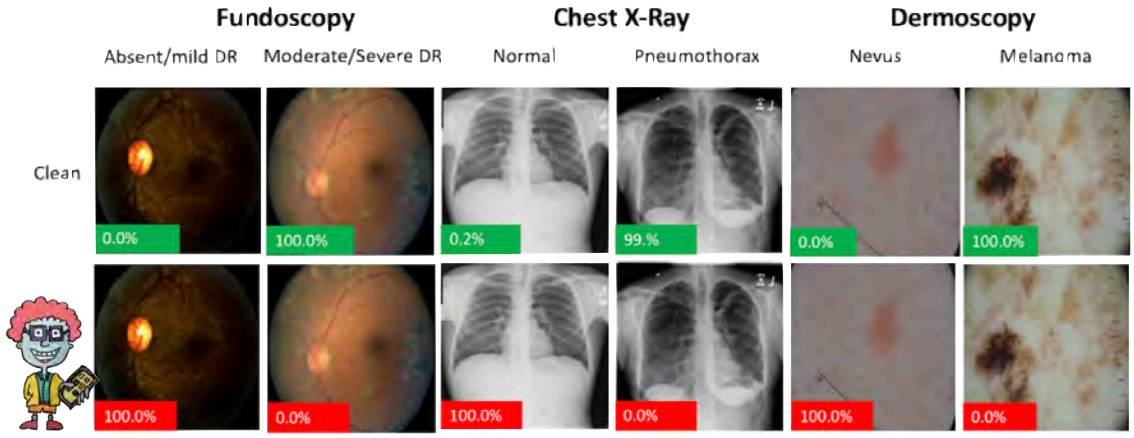
Les « aversarials attacks » sont maintenant connues, mais..

J. Su, et al. "One pixel attack for fooling deep neural networks.," CoRR, 2017.

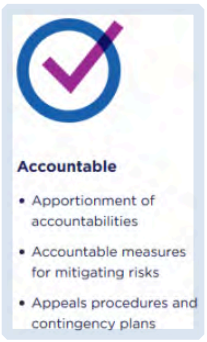


Akhtar & Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," arXiv.org, 02-Jan-2018.

…se pose la problème de la « responsabilité » des algorithmes … notamment en cas d'attaques d'images médicales.



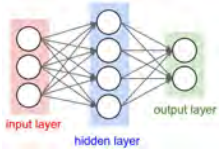
Finlayson, et al., "Adversarial Attacks Against Medical Deep Learning Systems.," arXiv, 2018.



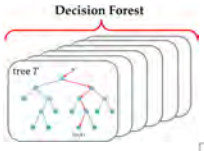
Qui est responsable en cas d'erreur ?

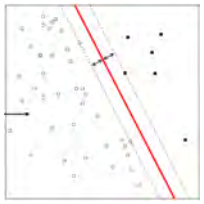
Interpretability vs Predictive power

predictive



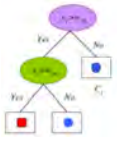
SVM XGBoost





KNN

sparse linear models



Concise decision trees

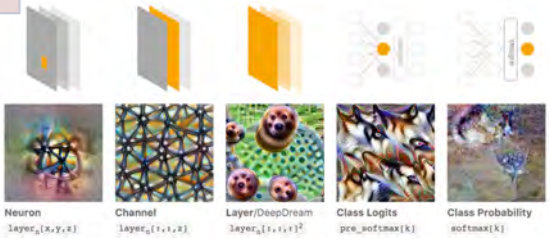
« Intrinsically » interpretable

Interpretability in Machine Learning

credits: Yann Chevaleyre

Type A - Interpreting black-box models

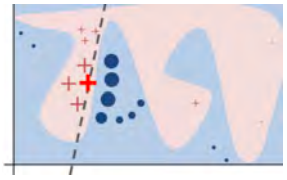
Model distillation (soft DT) [Frosst&Hinton,17]
Looking into the black box



[Olah, distill.pub]

Type B - Interpreting predictions from black-box models

Attribution methods: e.g. LIME



"Why Should I Trust You?"
Explaining the Predictions
of Any Classifier
[Ribeiro et al. '16]

Type C - Learning interpretable models

Intrinsically interpretable models: (discrete) linear model, scoring model, …

Interepreting NN using « attention maps » indicating the areas of the heat ma that the neural-network model is using to make the prediction for the image.

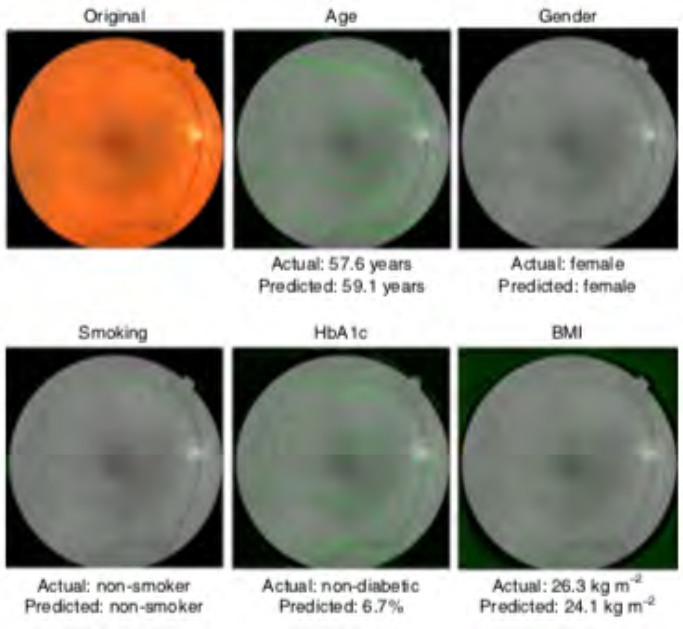


Table 6 | Percentage of the 100 attention heat maps for which doctors agreed that the heat map highlighted the given feature

Risk factor	Vessels (%)	Optic disc (%)	Non-specific features
Age	95	33	38
Gender	71	78	50
Current smoker	91	25	38
HbA1c	78	32	46
SBP	98	14	54
DBP	29	5	97
BMI	1	6	99

Heat maps (n = 100) were generated for each risk factor and then presented to three ophthalmologists who were asked to check the features highlighted in each image (n = 300 responses for each risk factor). The images were shuffled and presented as a set of 700, and the ophthalmologists were blinded to the output prediction of the heat maps and the ground-truth label. For the variables that were present in both datasets (age and gender), the most commonly highlighted features were identical in both datasets.

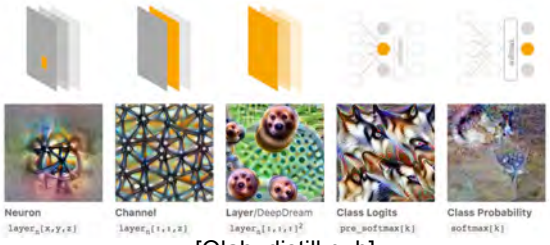
Attention maps for a single retinal fundus image. The top left image is a sample retinal image in colour from the UK Biobank dataset. The remaining images show the same retinal image, but in black and white.

Interpretability in Machine Learning

credits: Yann Chevaleyre

Type A - Interpreting black-box models

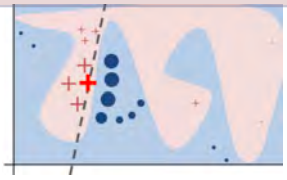
Model distillation (soft DT) [Frosst&Hinton,17]
Looking into the black box
Feature Importance (e.g. Random Forest)



[Olah, distill.pub]

Type B - Interpreting predictions from black-box models

Attribution methods: e.g. LIME



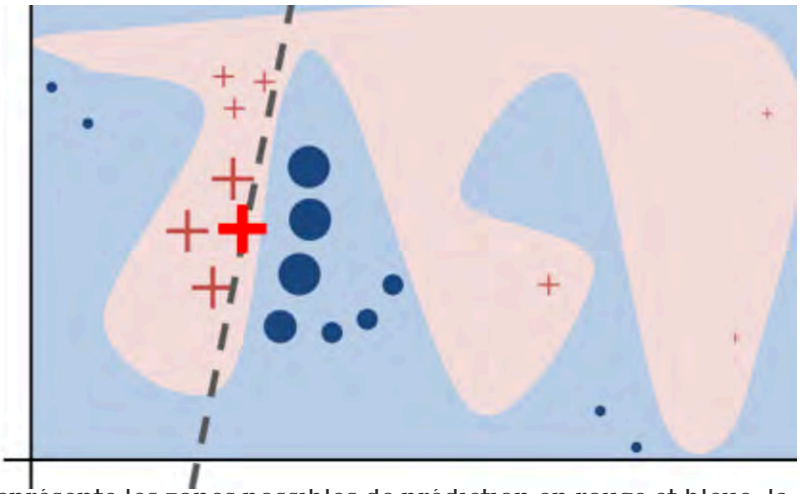
"Why Should I Trust You?"
Explaining the Predictions
of Any Classifier
[Ribeiro et al. '16]

Type C - Learning interpretable models

Intrinsically interpretable models: (discrete) linear model, scoring model, …

Interpreting predictions: LIME - Local Interpretable Model-Agnostic Explanations

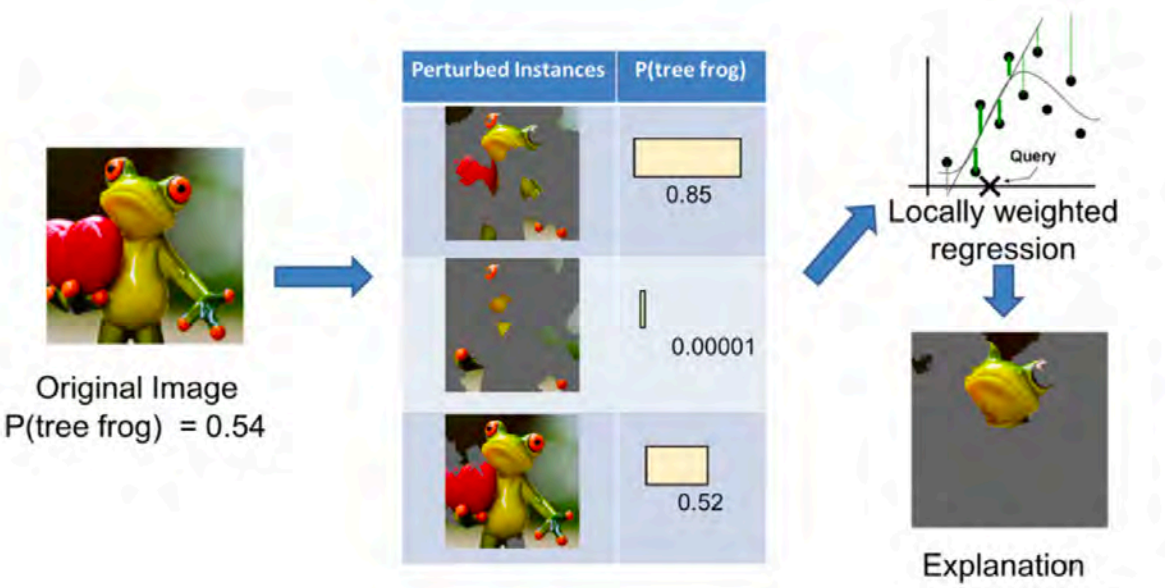
Objectif: convertir les prédictions en un modèle **interprétable : séparateur linéaire.**



Le graphique représente les zones possibles de prédiction en rouge et bleue, la croix rouge en gras et la prédiction initiale, les axes représentent des variables les autres points (rond bleu ou croix rouge) sont les **prédictions obtenues après modification des valeurs des variable**s.
Par exemple, un point situé à droite de la prédiction originale aura été modifiée uniquement sur la variable qui correspond à l'axe des abscisses.
Enfin plus un point possède une grande taille, plus il est "proche" (en distance) du point initial.

M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?'," presented at the the 22nd ACM SIGKDD International Conference, New York, New York, USA, 2016, pp. 1135–1144.

LIME



[source:]

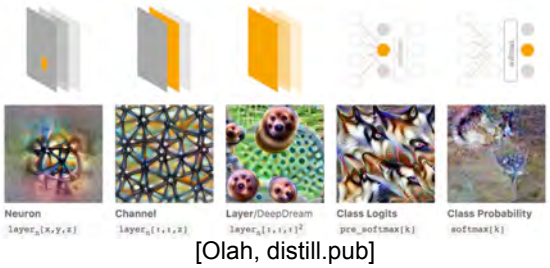
[Ribeiro et al. '16]

Interpretability in Machine Learning

credits: Yann Chevaleyre

Type A - Interpreting black-box models

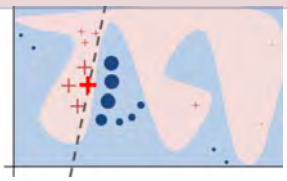
Model distillation (soft DT) [Frosst&Hinton,17]
Looking into the black box
Feature Importance (e.g. Random Forest)



[Olah, distill.pub]

Type B - Interpreting predictions from black-box models

Attribution methods: e.g. LIME



"Why Should I Trust You?" Explaining the Predictions of Any Classifier [Ribeiro et al. '16]

Type C - Learning interpretable models

Intrinsically interpretable models: (discrete) linear model, scoring model, …

Diarem score: was best score (2013); validated on independent cohorts

	Score
Age (years)	
<40	0
40–49	1
50–59	2
≥60	3
HbA_{1c} (%)	
<6·5%	0
6·5–6·9%	2
7·0–8·9%	4
≥9·0%	6
Other diabetes drugs	
No sulfonylureas or insulin-sensitising agent other than metformin	0
Sulfonylureas and insulin-sensitising agent other than metformin	3
Treatment with insulin	
No	0
Yes	10

Total score calculated by adding scores for each of the four variables.

Table 5: Calculation of DiaRem score for prediction of the probability of diabetes remission after Roux-en-Y gastric bypass surgery

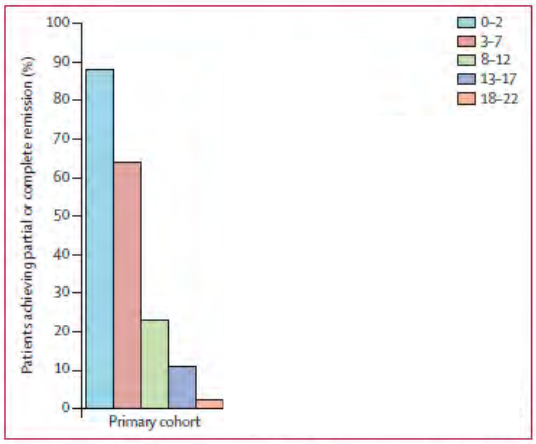


Figure 4: Proportion of patients in each cohort achieving partial or complete remission at 14 months after surgery, by DiaRem score

Table 2 Prediction errors of the different models in percentage

	Train error	Test error	Cross-validation error
Hayes - SL	13.4	31	
Hayes - J48DT	12.6	34.5	
Dixon - LR6	16	40.5	
Dixon - LR7	12	22.6	
Lee - Score	27.3[a]	16.7	16.7
Still - Score	19.4[b]	15.5	15.9
LR	7.1		19.9
DT	13.1		17.6
Lasso	14.3		18.9
EN[c]	13.1		17.2

Still et al Lancet endocrinology 2013;

74

Automated Score Re-Construction of Diarem

1. Identification of related clinical variables

age	glycated hemoglobin	insuline	other drugs

2. Meaningful thresholds for clinical variables

age				glycated hemoglobin				insuline		other drugs	
<40	40–49	50 – 59	>60	<6.5	6.5 – 6.9	7 – 8.9	> 9	yes	no	yes	no

3. Optimization of weights for sub-groups of the variables

age				glycated hemoglobin				insuline		other drugs	
<40	40–49	50 – 59	>60	<6.5	6.5 – 6.9	7 – 8.9	> 9	yes	no	yes	no
0	1	2	3	0	2	4	6	10	0	3	0

4. Find an optimal separator between two classes

Classify as `Remission` if sum of scores < 7

Classify as `Non-remission` if sum of scores ≥ 7

2 views

PERSPECTIVE
https://doi.org/10.1038/s42256-019-0048-x
nature machine intelligence

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Black box machine learning models are currently being used for high-stakes decision making throughout society, causing problems in healthcare, criminal justice and other domains. Some people hope that creating methods for explaining these black box models will alleviate some of the problems, but trying to explain black box models, rather than creating models that are interpretable in the first place, is likely to perpetuate bad practice and can potentially cause great harm to society. The way forward is to design models that are inherently interpretable. This Perspective clarifies the chasm between explaining black boxes and using inherently interpretable models, outlines several key reasons why explainable black boxes should be avoided in high-stakes decisions, identifies challenges to interpretable machine learning, and provides several example applications where interpretable models could potentially replace black box models in criminal justice, healthcare and computer vision.

NATURE MACHINE INTELLIGENCE | VOL 1 | MAY 2019 | 206–215 | www.nature.com/natmachintell

Prise de décisions à enjeux élevés **santé**, justice pénale, etc.
La voie à suivre est de concevoir des modèles qui sont intrinsèquement interprétables

Nous ne pouvons pas utiliser boîte noire AI pour trouver le lien de causalité, ou de compréhension→ l'intelligence humaine et l'IA interprétable. Acceptons les Boîte noires pour fournir de la valeur, optimiser des résultats et susciter l'inspiration.

ARTIFICIAL INTELLIGENCE

In defense of the black box

Black box algorithms can be useful in science and engineering

By **Elizabeth A. Holm**

The science fiction writer Douglas Adams imagined the greatest computer ever built, Deep Thought, programmed to answer the deepest question ever asked: the Great Question of Life, the Universe, and Everything. After 7.5 million years of processing, Deep Thought revealed its answer: Forty-two (*1*). As artificial intelligence (AI) systems enter every sector of human endeavor—including science, engineering, and health—humanity is confronted by the same conundrum that Adams encapsulated so succinctly: What good is knowing the answer when it is unclear why it is the answer? What good is a black box?

In an informal survey of my colleagues in the physical sciences and engineering, the top reason for not using AI methods such as deep learning, voiced by a substantial majority, was that they did not know how to interpret the results. This is an important objection, with implications that range from practical to ethical to legal (*2*). The goal of scientists and the responsibility of engineers is not just to predict what happens but to understand why it happens. Both an engineer and an AI system may learn to predict whether a bridge will collapse. But only the engineer can explain that decision in terms of physical models that can be communicated to and evaluated by others. Whose bridge would you rather cross?

CIMPA

ÉCOLE MATHÉMATIQUE AFRICAINE : Bases mathématiques de l'intelligence artificielle

Introduction à l'apprentissage artificiel en santé (cours 2/2)

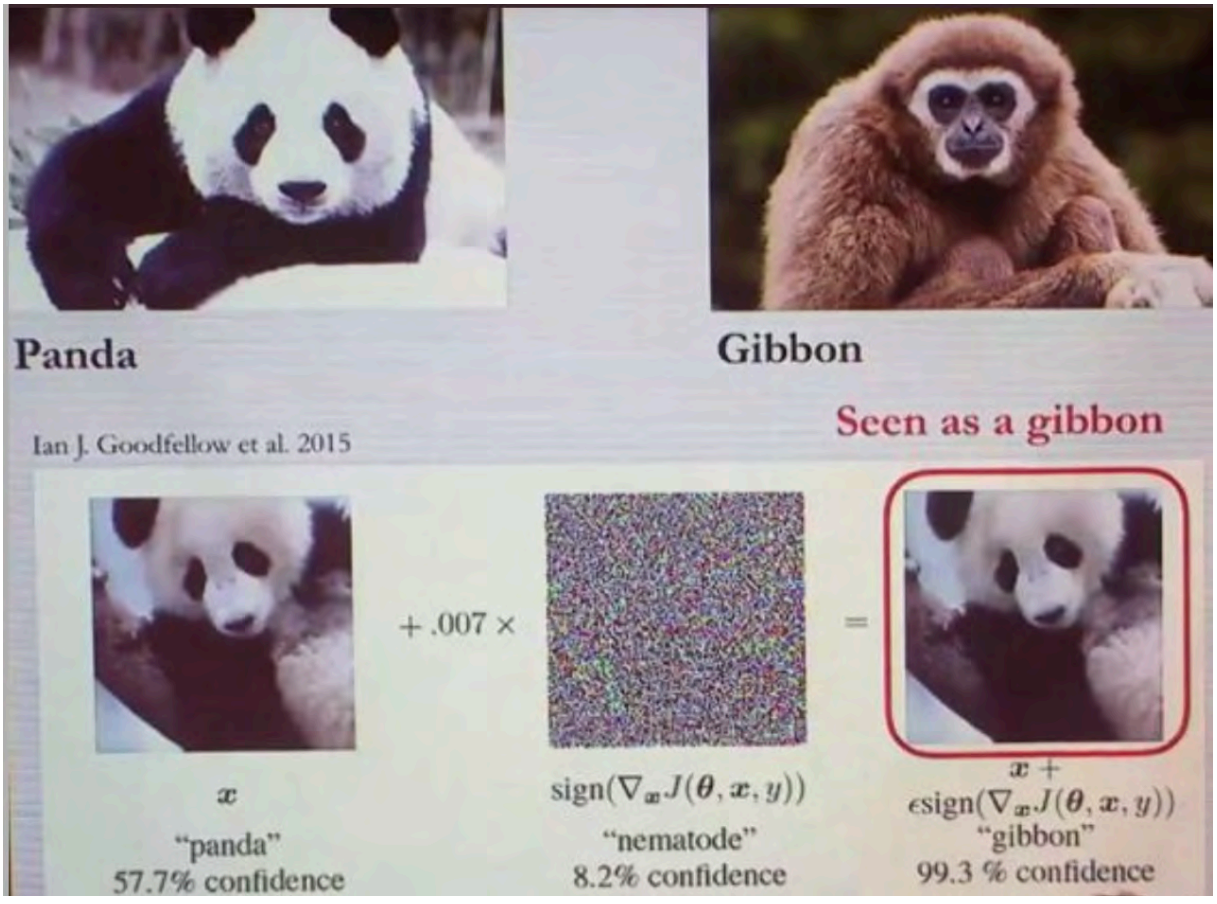
JEAN-DANIEL ZUCKER

UMMISCO Unité Mixte Internationale
IRD 75 ans
SORBONNE UNIVERSITÉ
INTEGROMICS
INVESTISSEMENTS D'AVENIR
ANR
ICAN TOGETHER

Compléments de cours suite aux questions à la fin du premier cours de la veille

1. Qu'est-ce qu'un Gibbon ?
2. Quelle est la différence entre Machine Learning et Statistical Modeling ?
3. Risque empirique, validation croisée et LOO.
4. Courbes ROC
5. Forêts aléatoires

1) A gibbon



2) What is the difference between Machine Learning and Statistical modeling?

The data modeling culture (Generative modeling)

- Assume a stochastic model (5-50 parameters).
- Estimate model parameters.
- Interpret model and make predictions.
- Estimated population: 98% of statisticians

The algorithmic modeling culture (Predictive modeling)

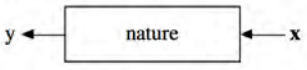
- Assume relationship between predictor vars and response vars has a functional form (10^2 -- 10^6 parameters).
- Search (efficiently) for the best prediction function.
- Make predictions [and interpret model]
- Interpretation / causation - mostly an after-thought.
- Estimated population: 2% of statisticians (many in other fields).

Leo Breiman Statistical Modeling / the two cultures

Statistical Science, 2001

1. INTRODUCTION

Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** (independent variables) go in one side, and on the other side the response variables **y** come out. Inside the black box, nature functions to associate the predictor variables with the response variables, so the picture is like this:



There are two goals in analyzing the data:

Prediction. To be able to predict what the responses are going to be to future input variables;
Information. To extract some information about how nature is associating the response variables to the input variables.

There are two different approaches toward these goals:

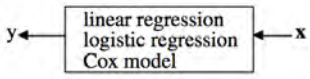
The Data Modeling Culture

The analysis in this culture starts with assuming a stochastic data model for the inside of the black box. For example, a common data model is that data are generated by independent draws from

response variables = f(predictor variables, random noise, parameters)

Leo Breiman is Professor, Department of Statistics, University of California, Berkeley, California 94720-4735 (e-mail: leo@stat.berkeley.edu).

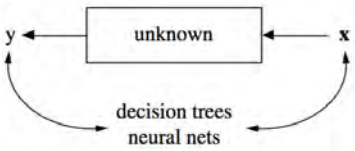
The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:



Model validation. Yes–no using goodness-of-fit tests and residual examination.
Estimated culture population. 98% of all statisticians.

The Algorithmic Modeling Culture

The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function $f(\mathbf{x})$—an algorithm that operates on **x** to predict the responses **y**. Their black box looks like this:



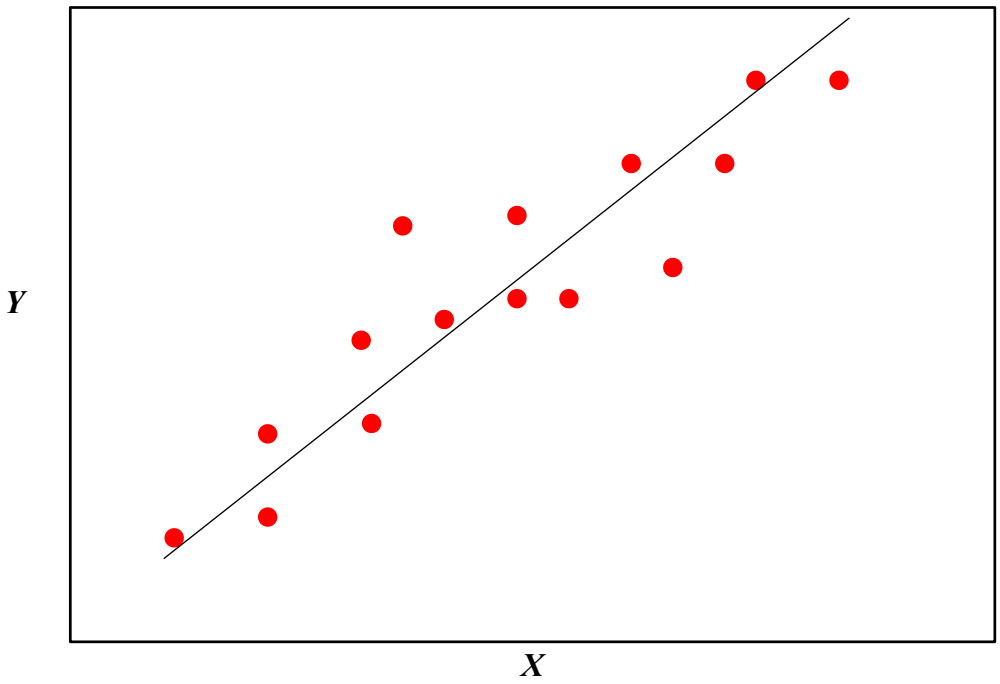
Model validation. Measured by predictive accuracy.
Estimated culture population. 2% of statisticians, many in other fields.

In this paper I will argue that the focus in the statistical community on data models has:

- Led to irrelevant theory and questionable scientific conclusions;

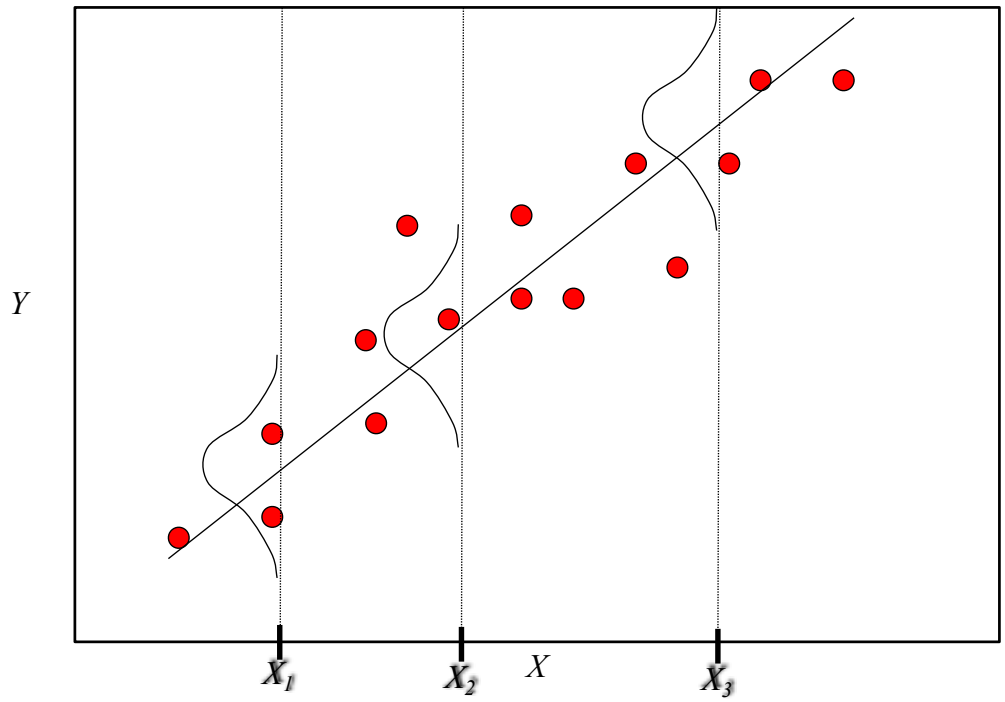
The difference between a data model and an algorithm.

Illustrative example: simple linear regression



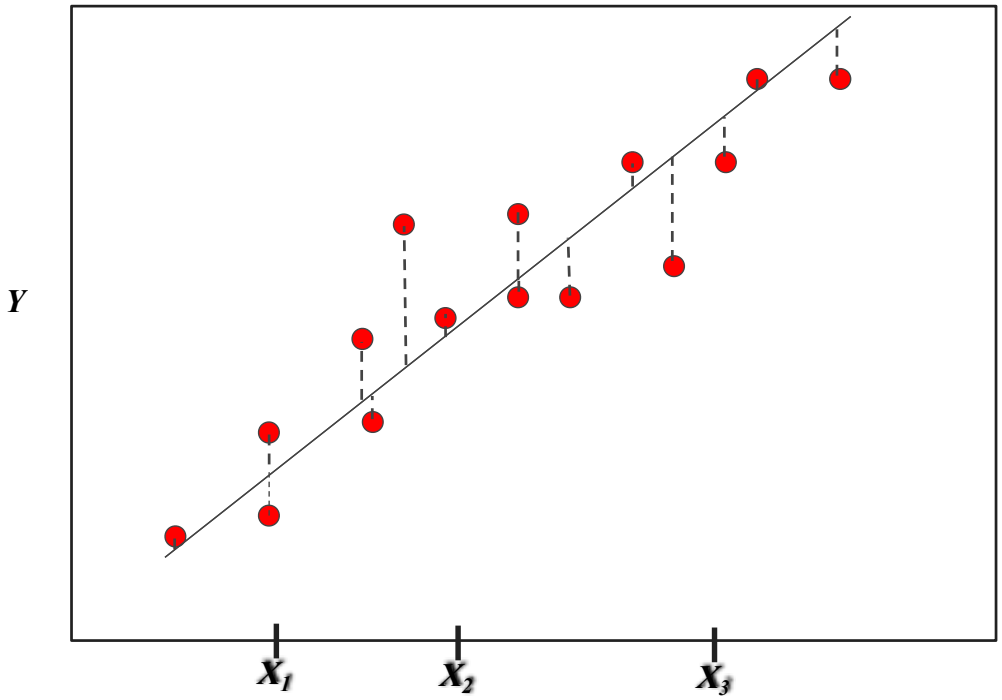
Least squares fit based on probabilistic assumptions

Chaque Y observé est supposé avoir une distribution de probabilité (p. ex. normale) pour un X donné.



Least squares fit base on geometrical assumptions

Il n'y a pas de rôle explicite pour la probabilité. Les points (X,Y) sont simplement approximés par une ligne droite. Une considération géométrique (non probabiliste) (minimiser une distance) conduit également à l'ajustement des moindres carrés.



Les 2 approches donnent tous deux des ajustements des moindres carrés. Mais ne poursuivent pas même objectif. Pour le modèle de données, la ligne est une estimation de la caractéristique probabiliste ; pour le modèle algorithmique, la ligne est un dispositif approximatif.

JDZucker

ML & SL de plus en plus similaire

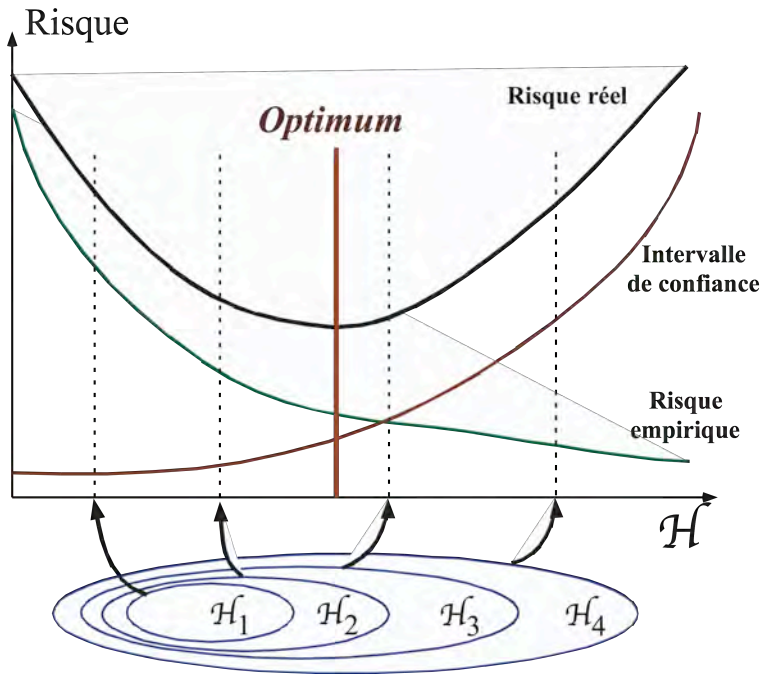
https://www.r-bloggers.com/whats-the-difference-between-machine-learning-statistics-and-data-mining/

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = $1,000,000	large grant = $50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

It should seem clear by now that machine learning, statistics, and data mining are all fundamentally similar.

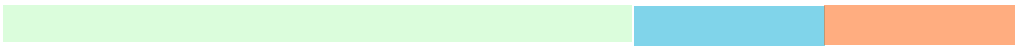
JDZucker

3) Risque Empirique et estimation de l'erreur



Risque réel, empirique et validation

Toutes les données disponibles

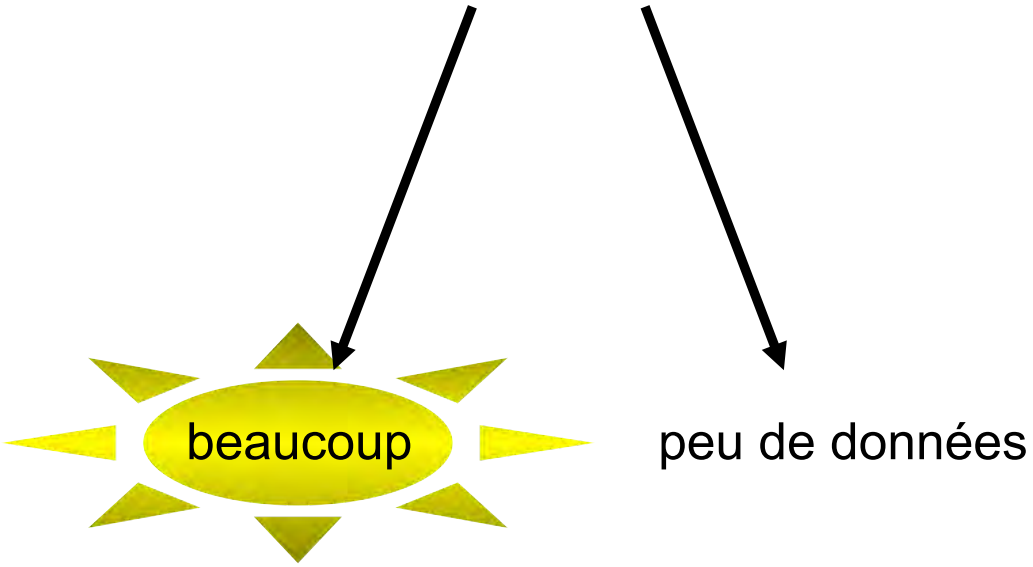


*Ensemble d'**apprentissage***

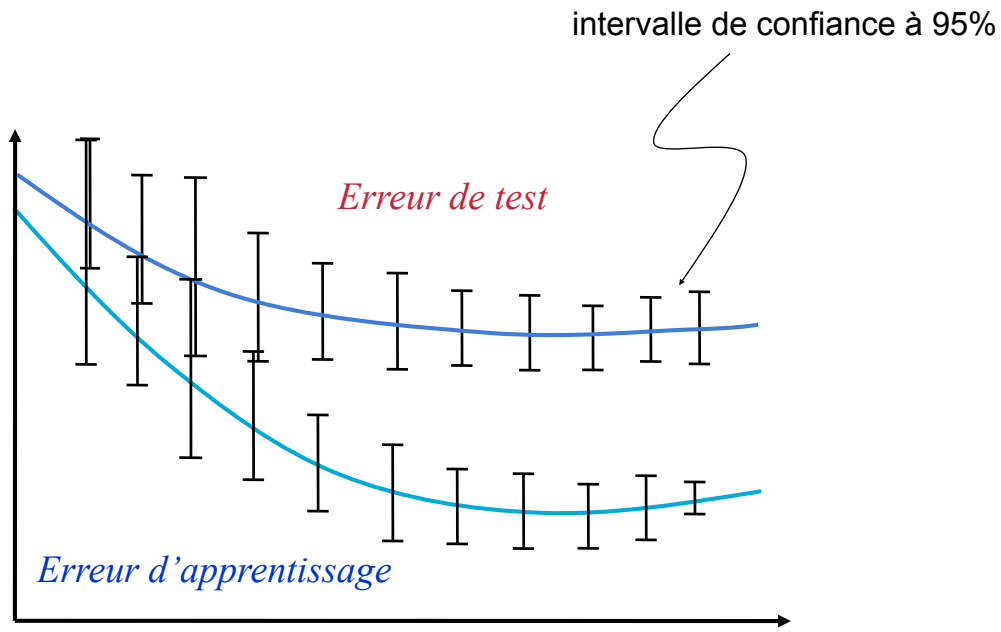
*Ensemble de **test***

*Ensemble de **validation***

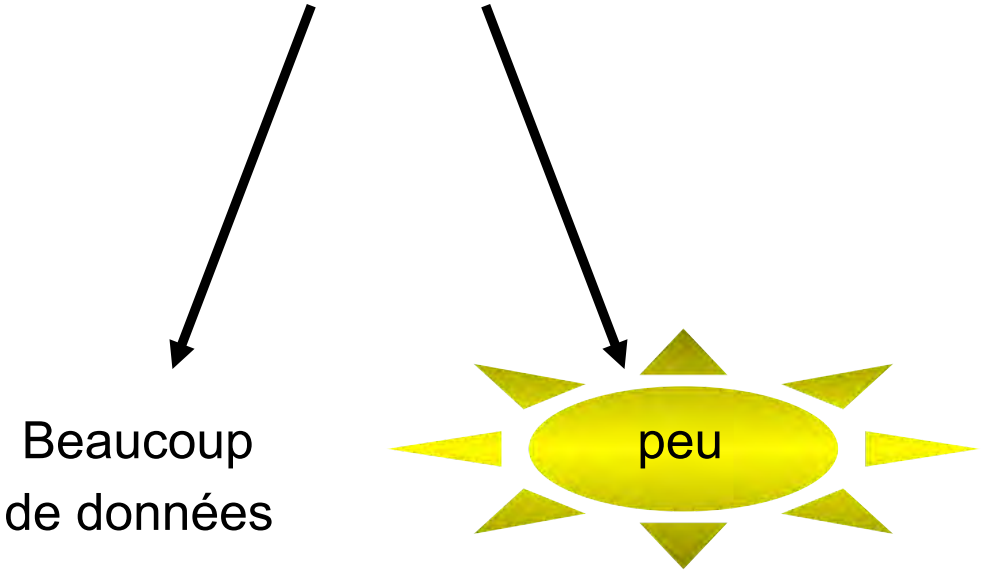
Évaluation des hypothèses produites



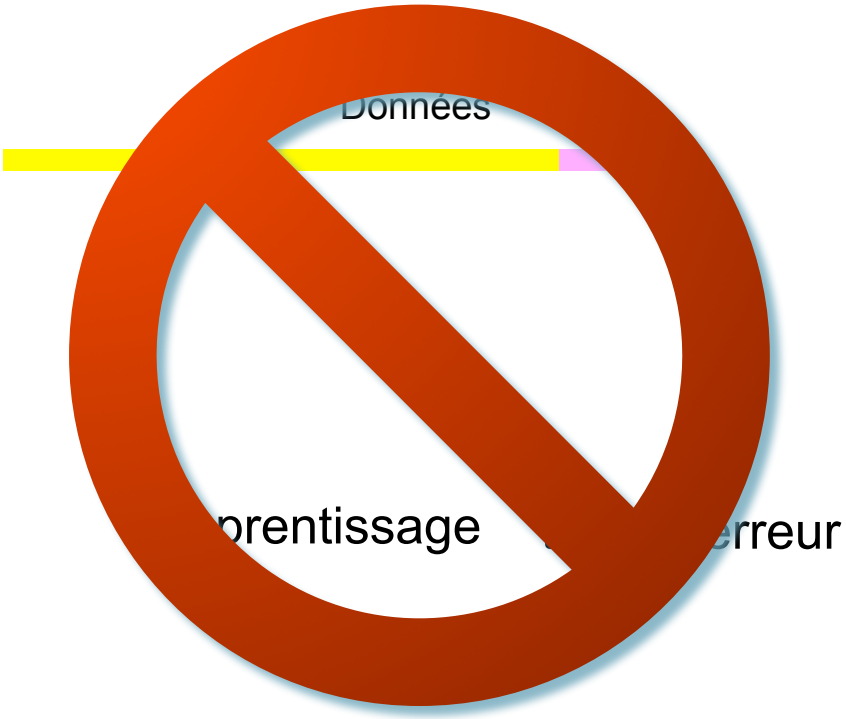
Courbes de performance



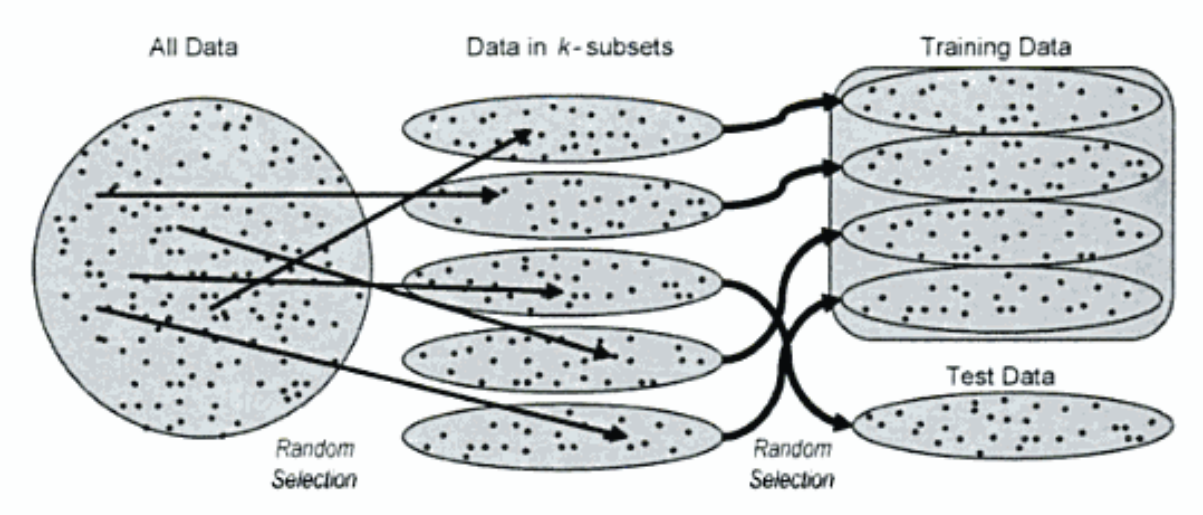
Évaluation des hypothèses produites



Différents ensembles



Validation croisée à k plis (k-fold)



Validation croisée à k plis (k-fold)

Données

k-way split

Apprend sur jaune, test sur rose → erreur_1

Apprend sur jaune, test sur rose → erreur_2

Apprend sur jaune, test sur rose → erreur_3

Apprend sur jaune, test sur rose → erreur_4

Apprend sur jaune, test sur rose → erreur_5

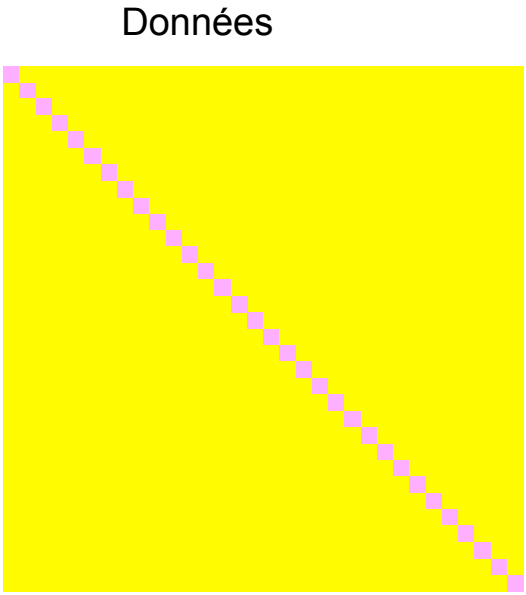
Apprend sur jaune, test sur rose → erreur_6

Apprend sur jaune, test sur rose → erreur_7

Apprend sur jaune, test sur rose → erreur_8

$$\text{erreur} = \Sigma\ \text{erreur}_i\ /\ k$$

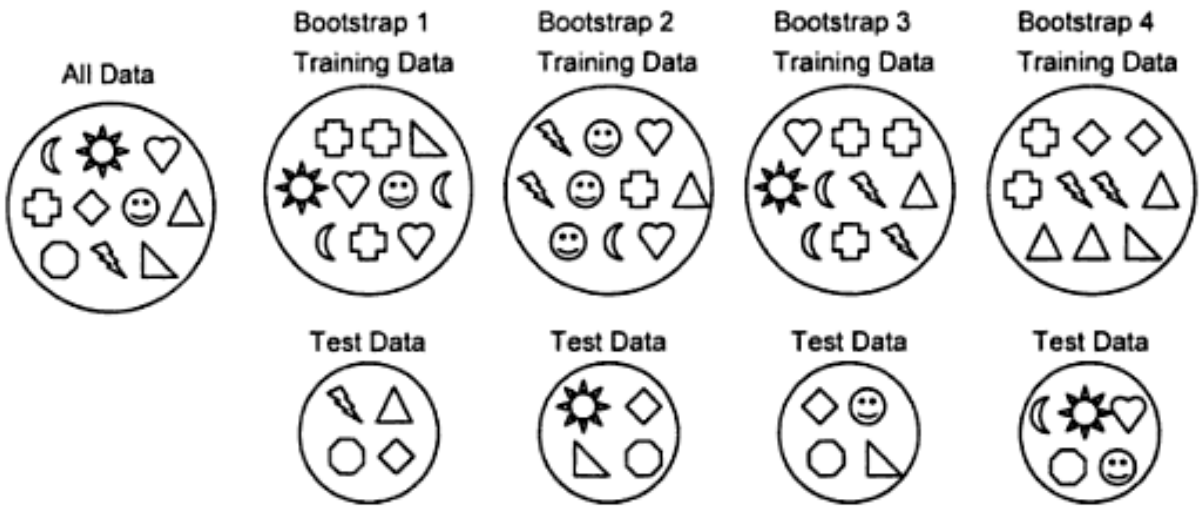
Procédure “leave-one-out” (LOO)



- Faible biais
- Haute variance
- Tend à sous-estimer l’erreur si les données ne sont pas vraiment i.i.d.

[Guyon & Elisseeff, jMLR, 03]

Le Bootstrap



Le bootstrap est biaisé

Le bootstrap est biaisé, car certaines observations **sont utilisées à la fois dans l'échantillon pour construire le modèle et dans l'échantillon pour le valider.**

Le bootstrap "hors du sac" (out-of-the-bag) et le bootstrap .632 tentent de corriger ce biais.

Le Bootstrap

Out-of-the-bag bootstrap

Le bootstrap "hors du sac" consiste à ne pas utiliser toutes les observations pour valider le modèle mais uniquement celles qui ne figurent pas déjà dans l'échantillon ayant servi à le construire (c'est d'ailleurs ce qu'on faisait pour la validation croisée).

Bootstrap .632

En fait, le bootstrap "out-of-the-bag" est quand-même biaisé, mais dans l'autre sens. Pour tenter de corriger ce biais, on peut faire une moyenne pondérée du bootstrap initial et du bootstrap oob.

```
.368 * (biais estimé par le bootstrap) +
.632 * (biais estimé par le bootstrap oob)
```

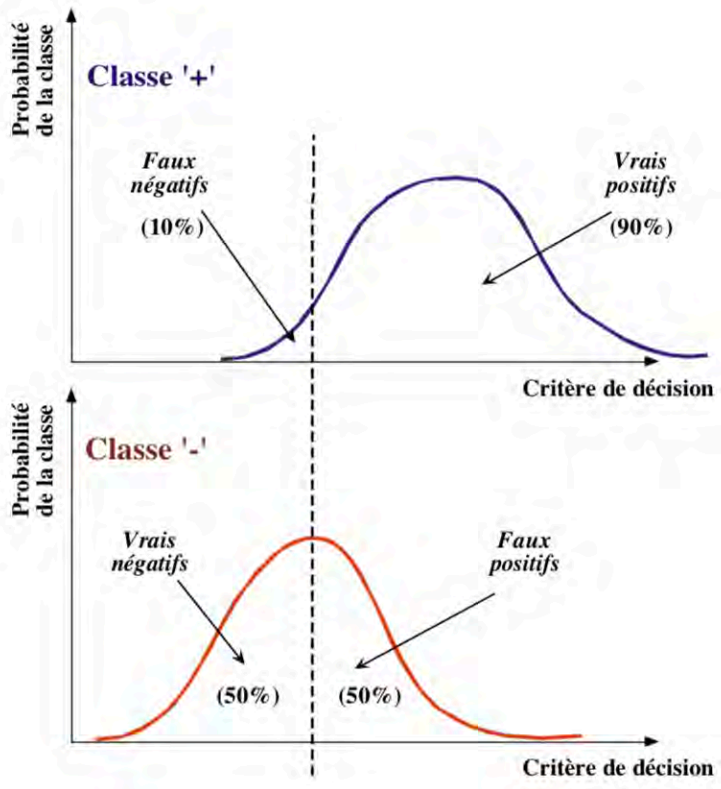
(le coefficient .632 s'interprète ainsi : pour n grand, les échantillons de bootstrap contiennent en moyenne 63,2% des observations initiales).

La probabilité de ne pas être choisi après n fois est

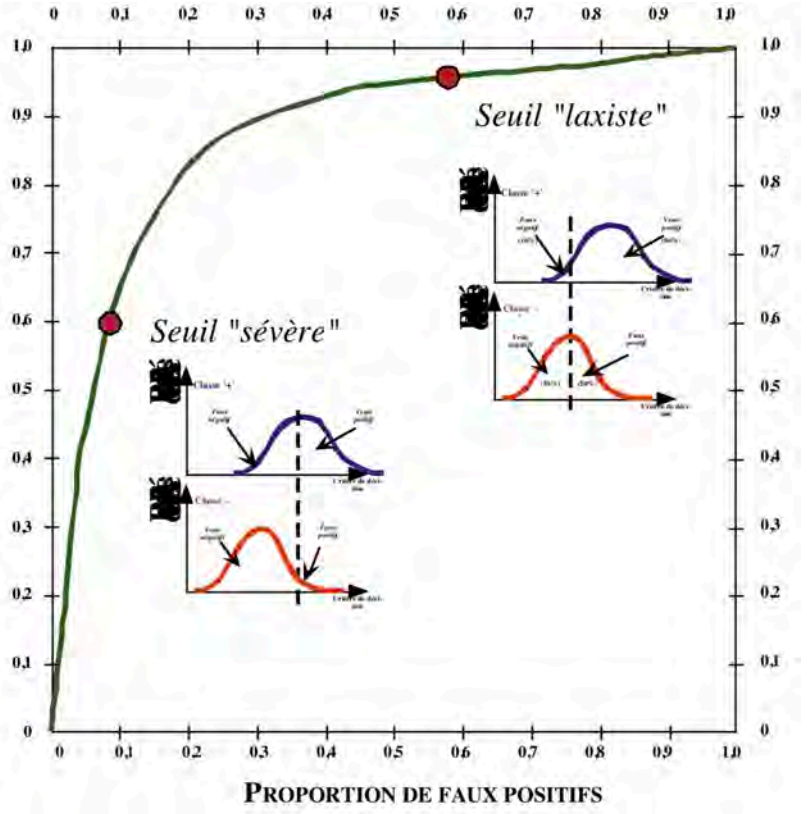
$$(1 - 1/n)^n = e^{-1} = 0.368$$

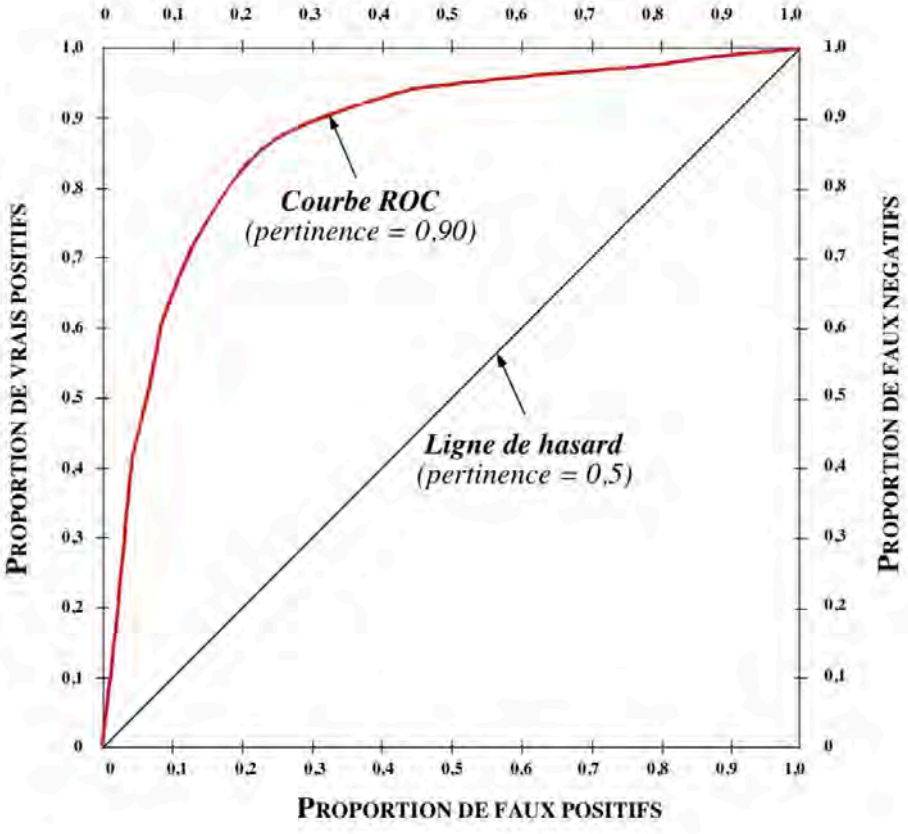
4) Courbe ROC

Receiver Operating Characteristics



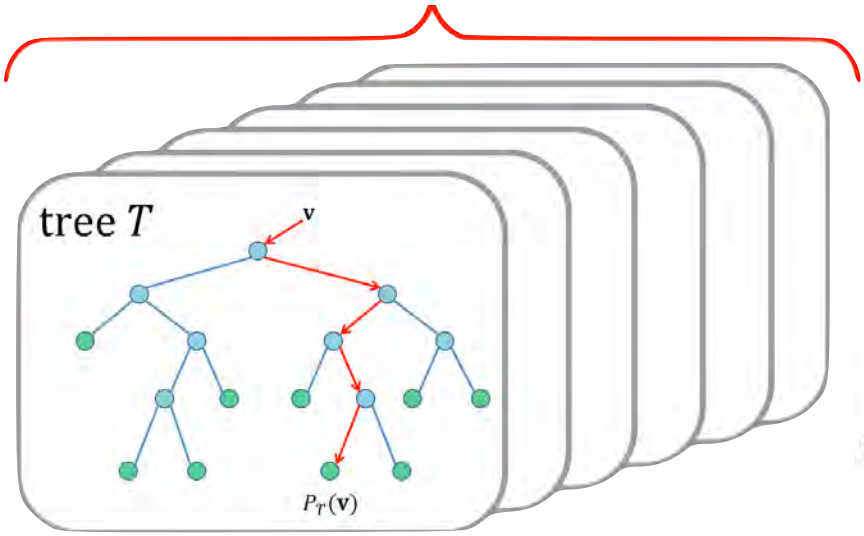
proportion de vrais positifs

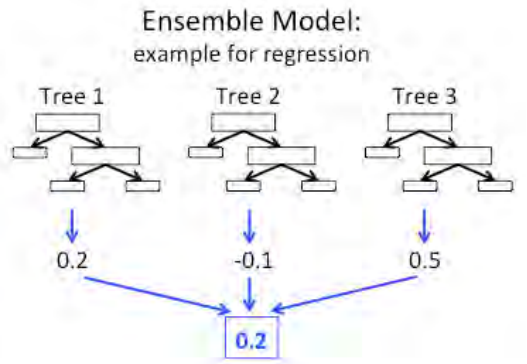




5) Random Forests: bootstrap on observations and features (m<p) for each tree

Random Forest





6) Can I export Orange workflow as a Python script?

"Malheureusement, non. Nous en avons débattu longuement et durement. Une fonctionnalité (limitée) de ce type serait probablement possible, mais il s'agirait d'un très gros projet à mener à bien et qui coûterait plus cher que les avantages escomptés. »

https://orange.biolab.si/faq/

Je veux utiliser Orange sur de gros volumes de données.

Orange fournit le widget SQL, qui échantillonne les instances de données et permet l'analyse exploratoire des données sur de grands ensembles de données. Vous avez besoin du module psycopg2 installé pour pouvoir voir le widget.

Installez-le directement en Python avec

pip install psycopg2

Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

III. Exemple de la régression régularisée

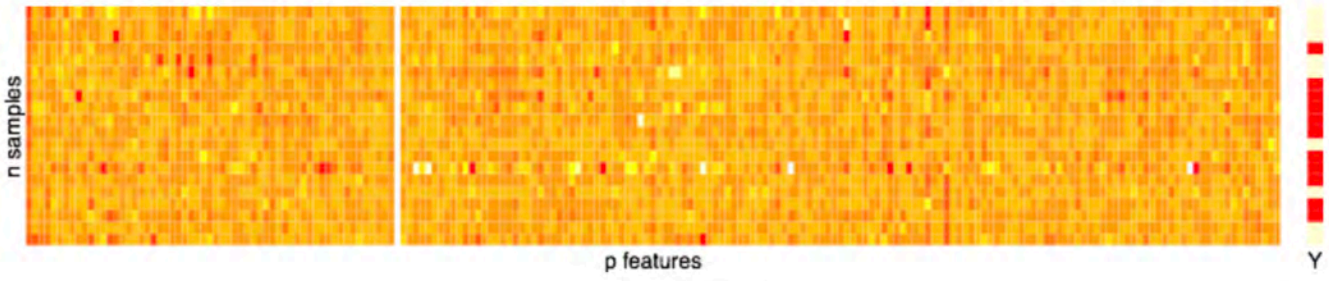
IV. Médecine de précision et IA

V. Classeur avec rejet

VI. Conclusion

Problème de régression en haute dimension

Gene expression



- n samples (patients), p features (genes)
- $X \in \mathbb{R}^{n \times p}$ gene expression profile of each patient
- $Y \in \mathcal{Y}^n$ survival information of each patient
- Fit a linear model for a sample $x \in \mathbb{R}^p$:

$$f(x) = \beta^\top x = \sum_{i=1}^{p} \beta_i x_i$$

- Standard methods (least squares or logistic regression) won't work because $n < p$

crédit: JP. Vert

I. La régression logistique simple

Variable dépendante : Y = 0 / 1

Variable indépendante : X

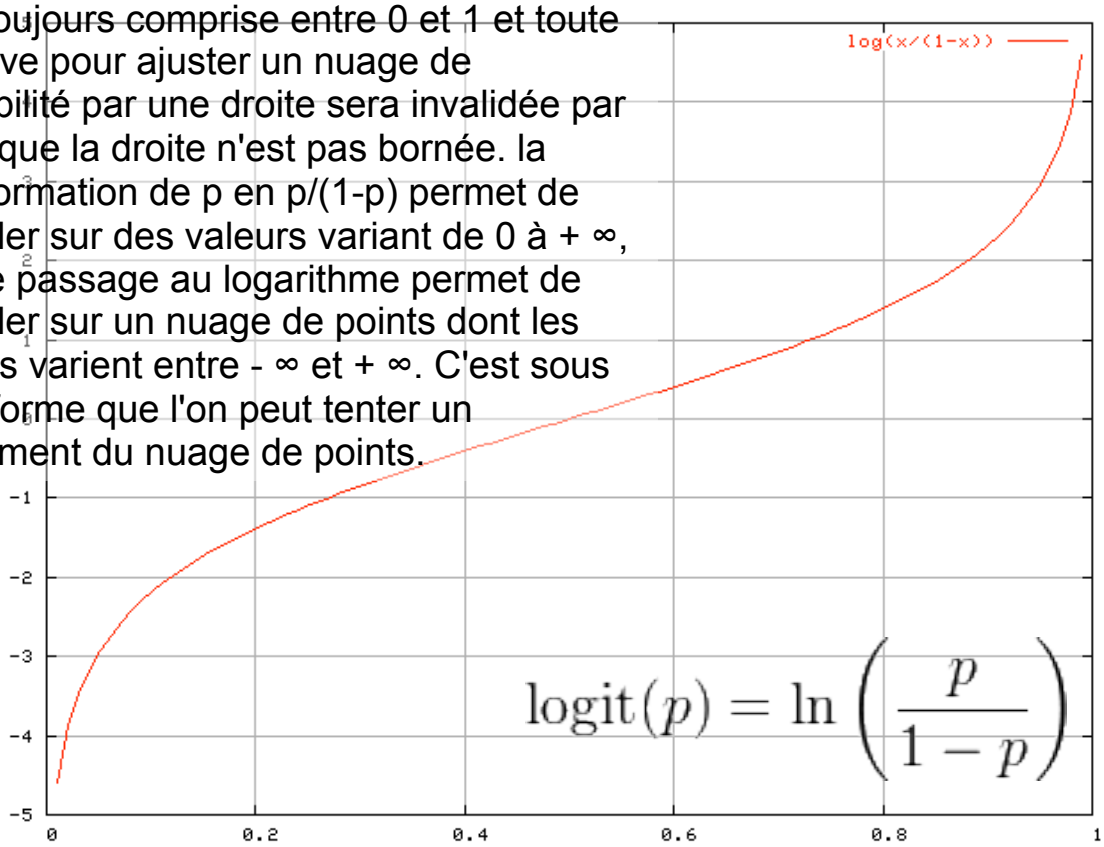
Objectif : Modéliser

$$\pi(x) = \text{Prob}(Y = 1/X = x)$$

- Le modèle linéaire $\pi(x) = \beta_0 + \beta_1 x$ convient mal lorsque X est continue.
- Le modèle logistique est plus naturel.

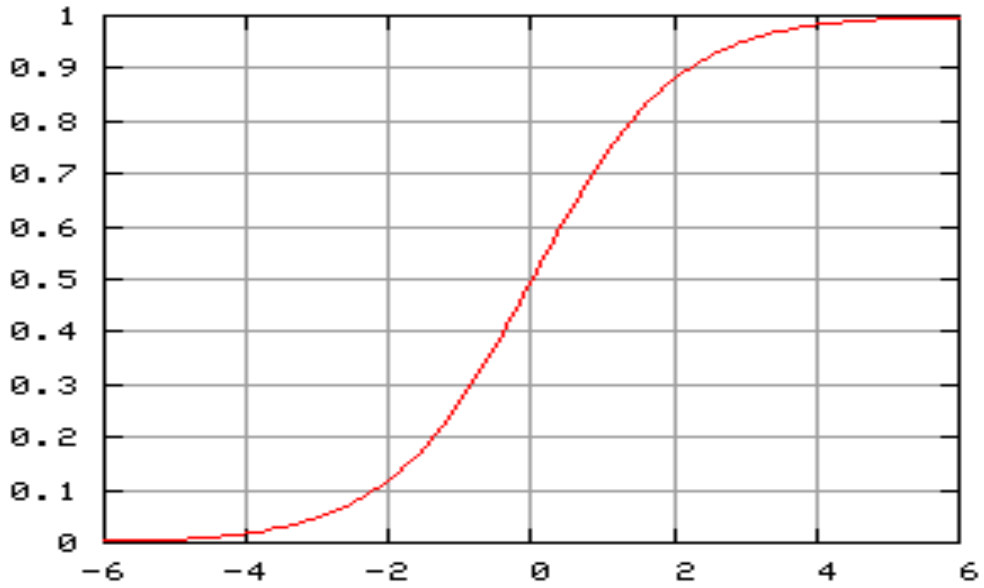
Logit(x)

Si p est une probabilité, cette probabilité sera toujours comprise entre 0 et 1 et toute tentative pour ajuster un nuage de probabilité par une droite sera invalidée par le fait que la droite n'est pas bornée. la transformation de p en p/(1-p) permet de travailler sur des valeurs variant de 0 à + ∞, puis le passage au logarithme permet de travailler sur un nuage de points dont les valeurs varient entre - ∞ et + ∞. C'est sous cette forme que l'on peut tenter un ajustement du nuage de points.



Odd ratio

$$LOGIT(p) = \ln\left(\frac{p}{(1-p)}\right) = z \quad \Leftrightarrow \quad p = \frac{\exp(z)}{1+\exp(z)}$$



Estimation des paramètres

Les données

X	Y
x_1	y_1
⋮	⋮
x_i	y_i
⋮	⋮
x_n	y_n

y_i = 1 si caractère présent,
0 sinon

Le modèle

$$\pi(x_i) = P(Y = 1 / X = x_i) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

Le modèle logistique

Logistic Regression:

$$\ln\left(\frac{P(Y)}{1\text{-}P(Y)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K$$

Linear Regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon$$

Relationship between Odds & Probability

$$\text{Odds}(\text{event}) = \frac{\text{Probability}(\text{event})}{1\text{-Probability}(\text{event})}$$

$$\text{Probability}(\text{event}) = \frac{\text{Odds}(\text{event})}{1+\text{Odds}(\text{event})}$$

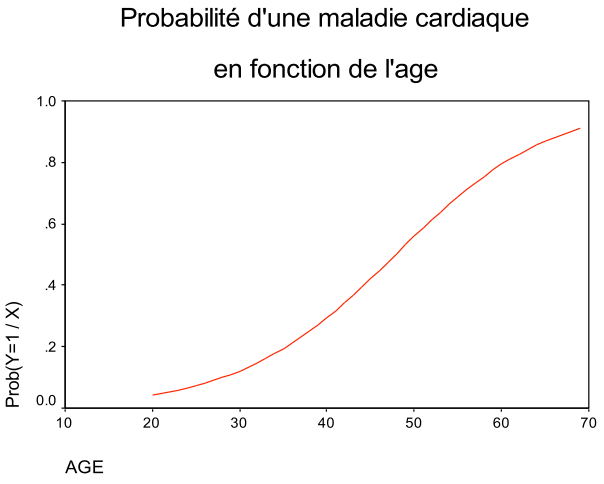
Le modèle logistique

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

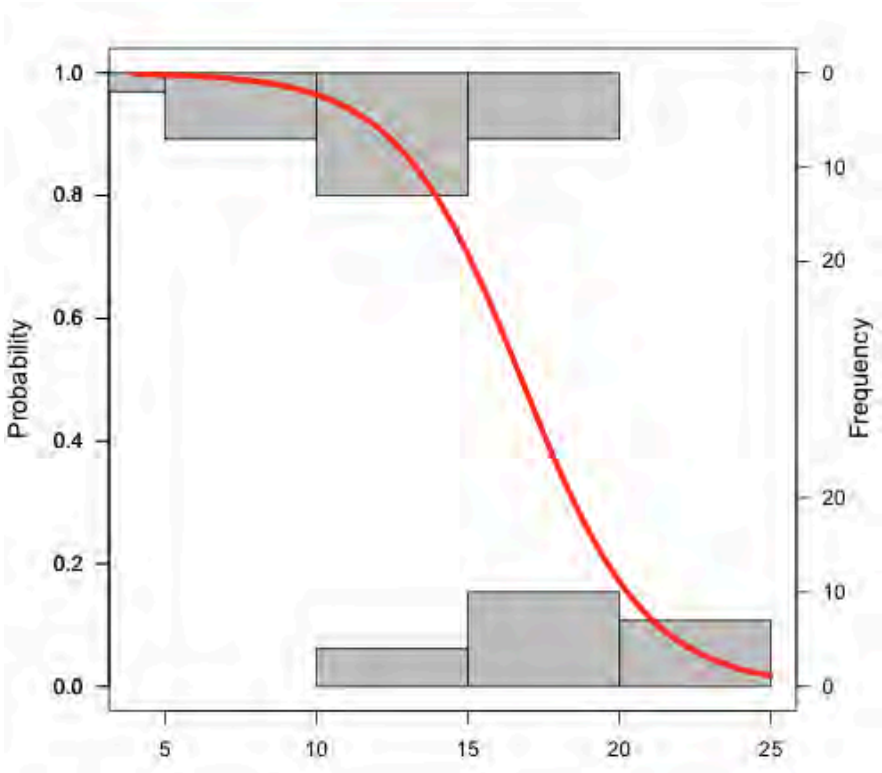
ou

$$\text{Log}\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

Fonction de lien : Logit



Exemple CARS : voiture Safe (sûre)



Regression Logistique

```
myData = cars
myData$safe = ifelse(dist < mean(dist),1,0)
attach(myData)

quartz(title="speed vs. safe")
# creates a quartz window with title

plot(speed,safe,xlab="speed",ylab="Proba of safe")

g=glm(safe~speed,family=binomial,mydata)
# run a logistic regression model
curve(exp(8-0.5*x)/(1+exp(8-0.5*x)),add = TRUE, col = "blue")
curve(predict(g,data.frame(speed=x),type="resp"),add=TRUE)
# draws a curve based on prediction from logistic regression model

points(speed,fitted(g),pch=20) # optional: you could skip this draws an
 invisible set of points of body size survival based on a 'fit' to glm model.
 pch= changes type of dots.

### or ####
library(popbio)
logi.hist.plot(speed,safe,boxp=FALSE,type="hist",col="gray")
```


Regression Logistique Prédiction

```
myData = cars
myData$safe = ifelse(dist < mean(dist),1,0)
attach(myData)
Newdata = data.frame(speed = c(13.5))

quartz(title="speed vs. safe")
plot(speed,safe,xlab="speed",ylab="Proba of safe")

g=glm(safe~speed,family=binomial,myData)
# run a logistic regression model
print(predict.glm(g,Newdata,type="response"))
```

Régression régularisée

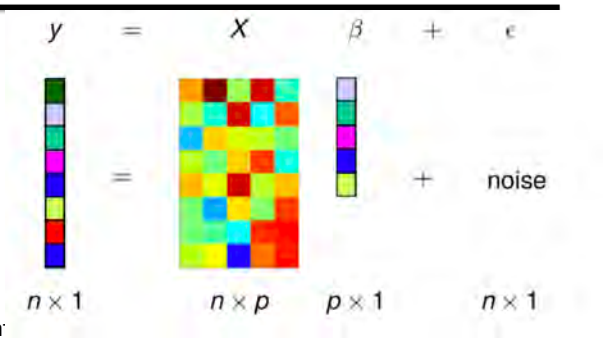
Ridge – Lasso – Elasticnet

D'après le cours de Ricco Rakotomalala

Université Lumière Lyon 2

Ricco Rakotomalala
Tutoriels Tanagra - http://tutoriels-data-mining.blogspot.fr/

1. Régression Linéaire Multiple



Modélisation prédictive d'une variable cible quantita

$$y = a_0 + a_1 x_1 + \cdots + a_p x_p + \varepsilon$$

Résume l'information de y que l'on n'arrive pas à capter avec les (x_1, ..., x_p) variables prédictives.

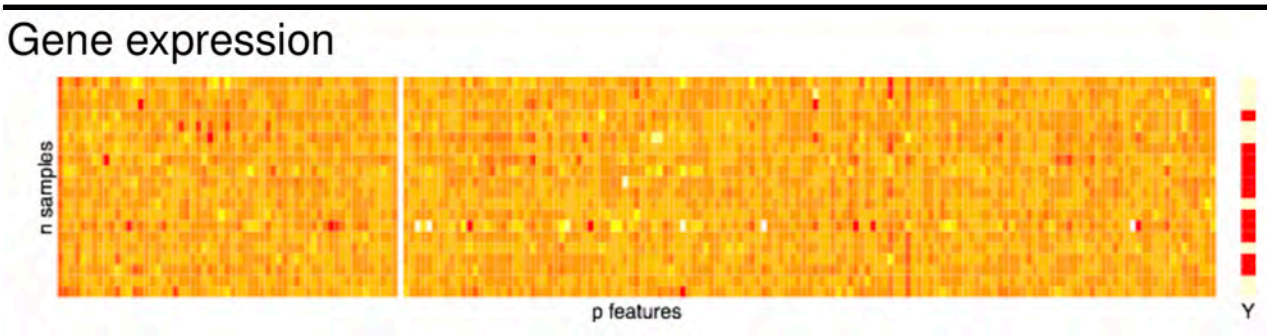
Disposant d'un échantillon de taille n, nous cherchons les paramètres estimés $\hat{a} = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p)$ qui minimise l'erreur quadratique moyenne (MSE : *mean squared error*).

$$\min_{a_0, a_1, \cdots, a_p} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{j=1}^{p} a_j x_{ij} \right) \right)^2$$

Estimateur des moindres carrés ordinaires (écriture matricielle)

$$\hat{a}_{MCO} = (X'X)^{-1} X'Y$$

Exemple expression génique



Questions: Sélectionner les gènes significatifs, trouver un algorithme avec une bonne prédiction de la réponse ?

Linear model ? Difficult to imagine : $p > n$!

$\hat{a}_{MCO} = (X'X)^{-1}X'Y$

- X^tX is an $p \times p$ matrix, but its rank is lower than n. If $n << p$, then

$$rk(X^tX) \leq n << p.$$

- Consequence : the Gram matrix X^tX is not invertible and even very ill-conditionned (most of the eigenvalues are 0 !)

Le modèle linéaire échoue

Regression Ridge

Ajouter une contrainte sur les coefficients lors de la modélisation pour maîtriser l'amplitude de leurs valeurs (« pour éviter qu'elles partent dans tous les sens »)

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j z_{ij} \right)^2$$

Sous la contrainte $\sum_{j=1}^{p} \beta_j^2 \leq \tau$

Norme L2 $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$

τ ($\tau \geq 0$) est un paramètre à fixer

Remarques :

- On parle de « shrinkage » (rétrécissement) : on rétrécit les plages de valeurs que peuvent prendre les paramètres estimés.
- Les variables x_j doivent être centrées et réduites (z_j) pour éviter que les variables à forte variance aient trop d'influence
- La variable cible y doit être centrée pour évacuer la constante de la régression (qui ne doit pas être pénalisée), la cible y peut être éventuellement réduite aussi : nous travaillerons alors sur les paramètres β_j
- ($\tau \rightarrow 0$) $\Rightarrow \beta_j \rightarrow 0$: les variances des coefficients estimés sont nulles
- ($\tau \rightarrow +\infty$) $\Rightarrow \beta_{Ridge} = \beta_{MCO}$

Regression Ridge: Fonction de pénalité

La régression ridge peut être écrite, de manière totalement équivalente : forme lagrangienne

Somme des carrés des résidus pénalisés :

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j z_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

λ (λ ≥ 0) est un paramètre (coefficient de pénalité) qui permet de contrôler l'impact de la pénalité : à fixer

Fonction de pénalité

Remarques :

- Il y a une équivalence directe entre τ et λ
- $\tau \rightarrow 0 \Leftrightarrow \lambda \rightarrow +\infty$: $\beta_j \rightarrow 0$ (tous), variances des coefficients nulles
- $\tau \rightarrow +\infty \Leftrightarrow \lambda \rightarrow 0$: $\beta_{Ridge} = \beta_{MCO}$

Question: comment choisir lambda

Utiliser la performance prédictive pour déterminer lambda

S'appuyer sur un critère de performance pure

Principe :

- Fixer une plage de valeurs de λ
- Construire le modèle sur un échantillon d'apprentissage Ω_{App}
- L'évaluer sur un échantillon test Ω_{Test}
- Choisir λ qui minimise un critère d'erreur en test

Dérivé du PRESS (predicted residual error sum of squares) :

$$ERR = \frac{1}{n_{Test}} \sum_{i \in \Omega_{Test}} (y_i - \hat{y}_i)^2$$

Les individus Ω_{Test} ne doivent pas faire partie de ceux qui ont permis de construire le modèle. Sinon, c'est λ = 0 sera la solution systématique.

Validation croisée :

Si la base est de taille restreinte, la partition apprentissage-test n'est pas judicieuse. Mieux vaut passer par la *K-fold* validation croisée (cf.

Contrainte sur la norme L1 des coefficients

Least Absolute Shrinkage and Selection Operation

RÉGRESSION LASSO

Lasso: contrainte sur la norme L1

Norme L1 : $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$

Très similaire à Ridge

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j z_{ij} \right)^2 \quad \text{s.c.} \quad \sum_{j=1}^{p} |\beta_j| \leq \tau$$

Avec la fonction de pénalité

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j z_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

λ (λ ≥ 0) est un paramètre (coefficient de pénalité) qui permet de contrôler l'impact de la pénalité : à fixer

Idem Ridge :

- x_j doivent être centrées et réduites, y au moins centré (mais peut être réduit aussi)
- Relation inverse entre τ et λ ; (λ = 0 ⇔ τ = +∞) ⇒ MCO ; plus λ élevé, plus la régularisation est forte (τ faible, les coefficients sont rétrécis)

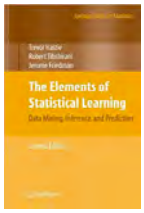
Quel intérêt par rapport à Ridge ? LASSO peut faire office de dispositif de sélection de variables en annulant certains coefficients β_j : les variables associées à ($\beta_j = 0$) sont de facto exclues du modèle prédictif.

Algorithme d'apprentissage-LARS (least-angle regression)

Il n'y a pas de calcul direct pour LASSO. Passer par la régression LARS. C'est une démarche itérative où l'on commence avec tous les (β_j=0). On fait évoluer les coefficients sélectivement. On a ainsi des scénarios de solutions (LASSO path) à mesure que $||\beta||_1$ augmente.

Algorithme LARS

« Elements of Statistical Learning », Hastie, Tibshirani, Friedman, Corrected 12th, Jan. 2017 ; page 74.



Légère modification permettant d'obtenir LASSO

Algorithm 3.2 *Least Angle Regression.*

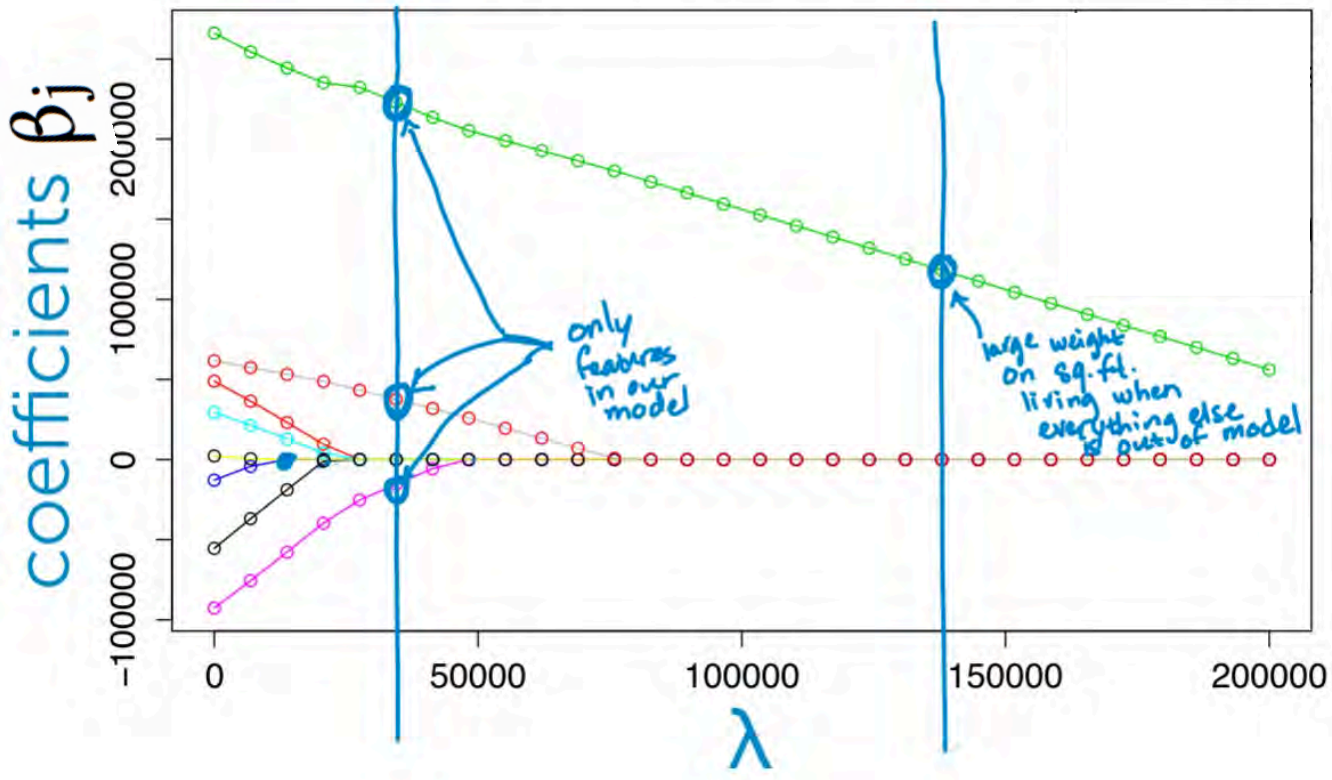
1. Standardize the predictors to have mean zero and unit norm. Start with the residual $\mathbf{r} = \mathbf{y} - \bar{\mathbf{y}}$, $\beta_1, \beta_2, \ldots, \beta_p = 0$.
2. Find the predictor $\mathbf{x}_j$ most correlated with $\mathbf{r}$.
3. Move β_j from 0 towards its least-squares coefficient $\langle \mathbf{x}_j, \mathbf{r} \rangle$, until some other competitor $\mathbf{x}_k$ has as much correlation with the current residual as does $\mathbf{x}_j$.
4. Move β_j and β_k in the direction defined by their joint least squares coefficient of the current residual on $(\mathbf{x}_j, \mathbf{x}_k)$, until some other competitor $\mathbf{x}_l$ has as much correlation with the current residual.
5. Continue in this way until all p predictors have been entered. After $\min(N-1, p)$ steps, we arrive at the full least-squares solution.

Algorithm 3.2a *Least Angle Regression: Lasso Modification.*

4a. If a non-zero coefficient hits zero, drop its variable from the active set of variables and recompute the current joint least squares direction.

Lasso: Least Absolute Shrinkage and

Coefficient path – lasso

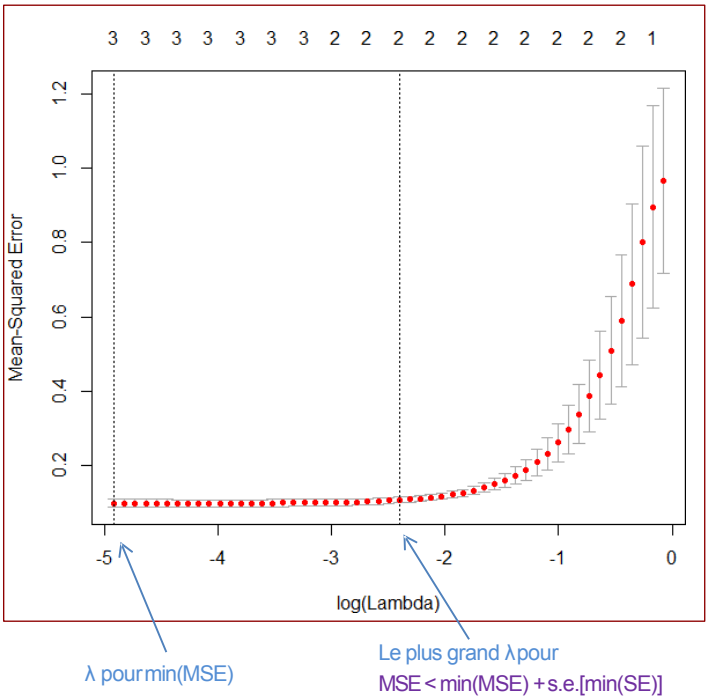


Choix du paramètre λ

Dans l'optique prédictive, la minimisation de l'erreur de prédiction en balayant les valeurs de λ reste la méthode la plus efficace. On procède souvent par validation croisée.

```
#package R
library(glmnet)

#3-fold cross-validation
cvfit <- cv.glmnet(as.matrix(S[1:3]),S$CONSO,nfolds=3)

#courbe log(Lambda) vs. MSE
plot(cvfit)

#valeur min de MSE(en validation croisée)
print(min(cvfit$cvm)) #0.0987

#lambda corresp.
print(cvfit$lambda.min) #0.007347

#lambda le plus élevé dont le MSE est inf.
#à la borne haute de l'intervalle de min(MSE)
cvfit$lambda.1se #0.09057664 (log[0.09057] = -2.4)
```

Puisque c'est en validation croisée, on peut même avoir un écart-type et donc un intervalle de confiance [MSE + s.e.(MSE)]



Bilan LASSO

Avantage

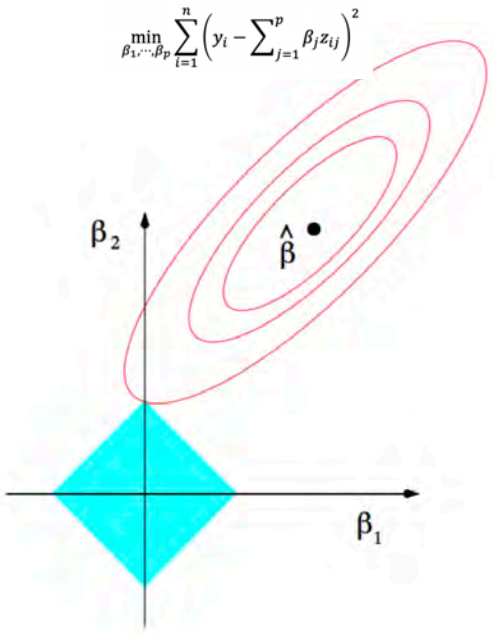
Capacité à sélectionner les variables en acceptant les **coefficients nuls**

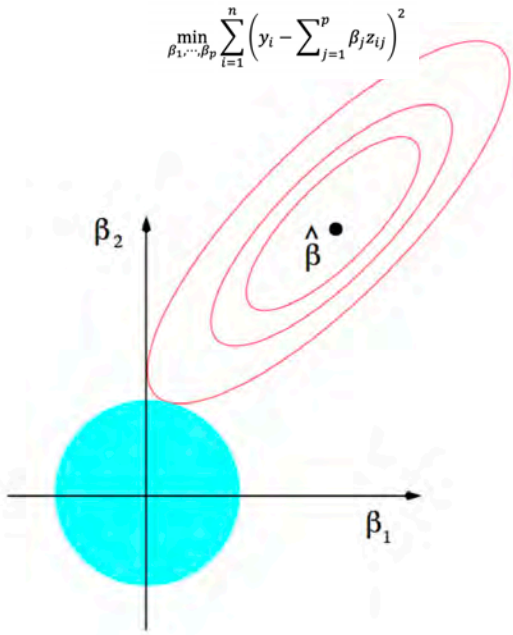
Inconvénients

Dans les problèmes à très grandes dimensions (p >> n), LASSO **ne sélectionne que n variables prédictives au maximum**, mécaniquement. C'est une vraie limitation de l'algorithme.

Parmi un groupe de variables corrélées, LASSO en choisit une, celle qui est la plus liée à la cible; souvent, masquant l'influence des autres. Cet inconvénient est inhérent aux techniques intégrant un mécanisme de sélection de variables (*ex. arbres de décision, quoique...*).

Interprétation géométrique





Norme L1 : $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$

Norme L2 $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$

Mix de Ridge et Lasso

ELASTICNET

Elasticnet : combiner les avantages de ridge et lasso

Formulation sous la forme d'un problème d'optimisation

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j z_{ij}\right)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2$$

λ_1 et λ_2 (≥ 0) sont des paramètres positifs à fixer. ($\lambda_1 = 0$ et $\lambda_2 > 0$) : Ridge ; ($\lambda_1 > 0$ et $\lambda_2 = 0$) : Lasso ; ($\lambda_1 = 0$ et $\lambda_2 = 0$) : MCO.

Formulation alternative de l'optimisation

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j z_{ij}\right)^2 + \lambda\left[\alpha \sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2\right]$$

R(β) = Fonction de pénalité de la régression elasticnet. ($\alpha = 0$) : Ridge ; ($\alpha = 1$) Lasso.

Autre formulation : optimisation sous contrainte

$$\min_{\beta_1,\cdots,\beta_p} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j z_{ij}\right)^2$$

$$\text{s.c.} \quad \alpha \sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2 \leq \tau$$

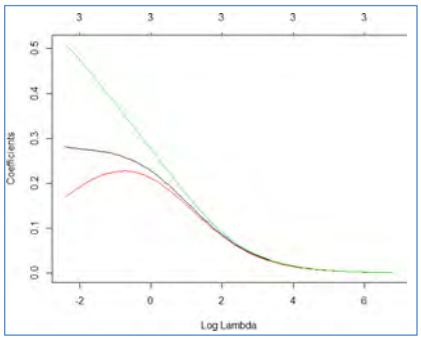
Avec toujours la relation inverse entre λ et τ

Quel intérêt ?

Double avantage :

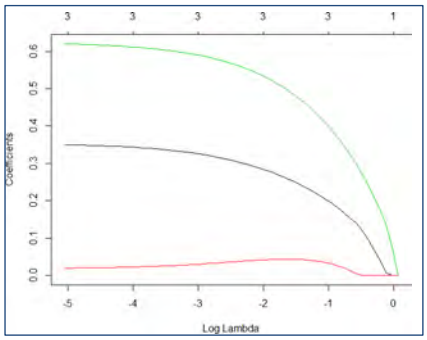
- Capacité de sélection de variables du LASSO conservée (coefficients nuls): exclusion des variables non pertinentes
- Groupe de variables prédictives corrélées, partage des poids (comme Ridge) et non plus sélection arbitraire

Le « coefficient-path » selon λ dépend de la valeur de α cette fois-ci.



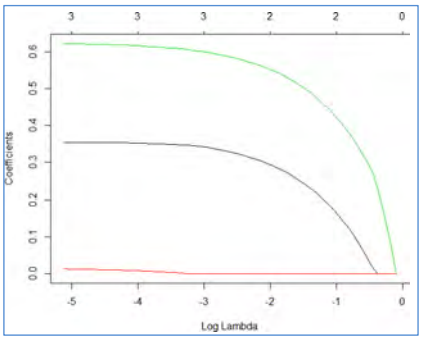
α = 0, Ridge

Pas de sélection de variables



α = 0.85

Plus de nuances dans la sélection



α = 1, Lasso

Sélection drastique

Conclusion

- La dimensionalité et la colinéarité sont des problèmes récurrents en régression.
  La régularisation via des fonctions de pénalités sur les coefficients est une solution viable.
- Notamment grâce à sa capacité à sélectionner les variables et à gérer les groupes de variables corrélées.
- Applicable à la régression linéaire et logistique, mais aussi à d'autres méthodes dès lors que l'on a des combinaisons linéaires de variables avec des coefficients à estimer (ex. réseaux de neurones, perceptron, SVM linéaire, etc.).
- Ne jamais oublier de ramener les explicatives sur la même échelle (centrer et réduire est le plus couramment utilisé) pour éviter les inégalités de traitements entre les coefficients.
- Ridge, Lasso et Elasticnet sont proposés dans de nombreux packages pour R et Python, dont certains ont servi dans ce support (MASS, glmnet, elasticnet, lars, penalizedSVM, scikit-learn… mais aussi ceux spécialisés dans les réseaux de neurones telles que tensorflow / keras).

Overfitting and regularization orange3

TP

https://blog.biolab.si/tag/regularization/

https://docs.biolab.si/2/_downloads/bio-tutorial.pdf

Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

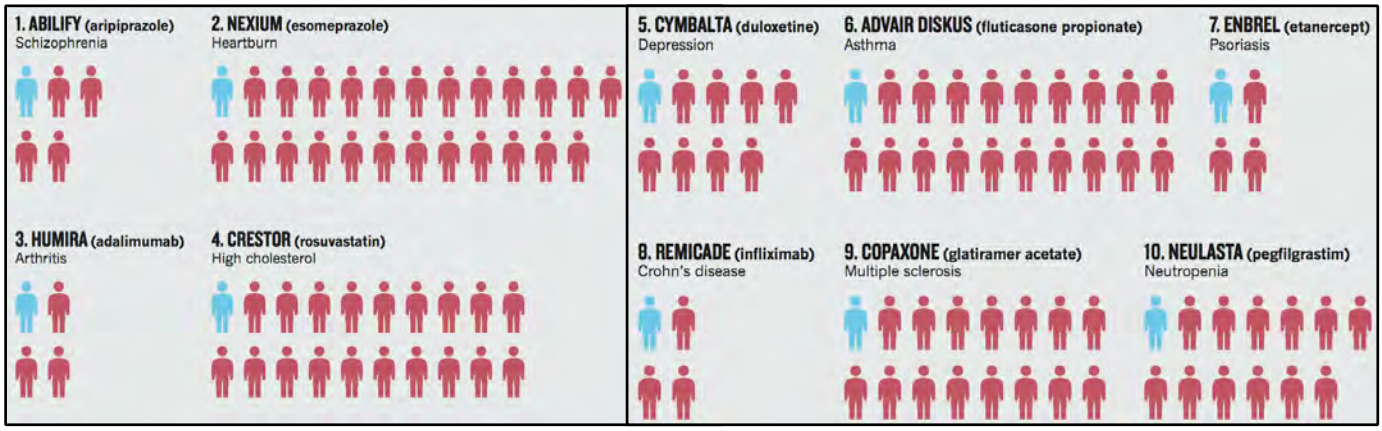
III. Exemple de la régression régularisée

IV. Médecine de précision et IA

V. Classeur avec rejet

Prédire la réponse aux traitements est un enjeu majeur

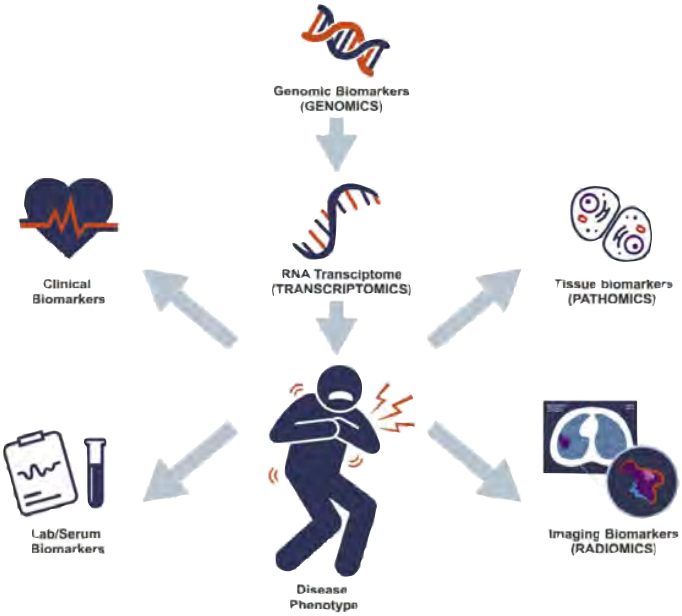
Pour **chaque** personne qu'ils aident (bleu), les dix médicaments les plus lucratifs aux États-Unis ne parviennent pas à améliorer les conditions d'entre **3** et **24** personnes (rouge).



Schork, N.J., 2015. Personalized medicine: time for one-person trials. Nature.

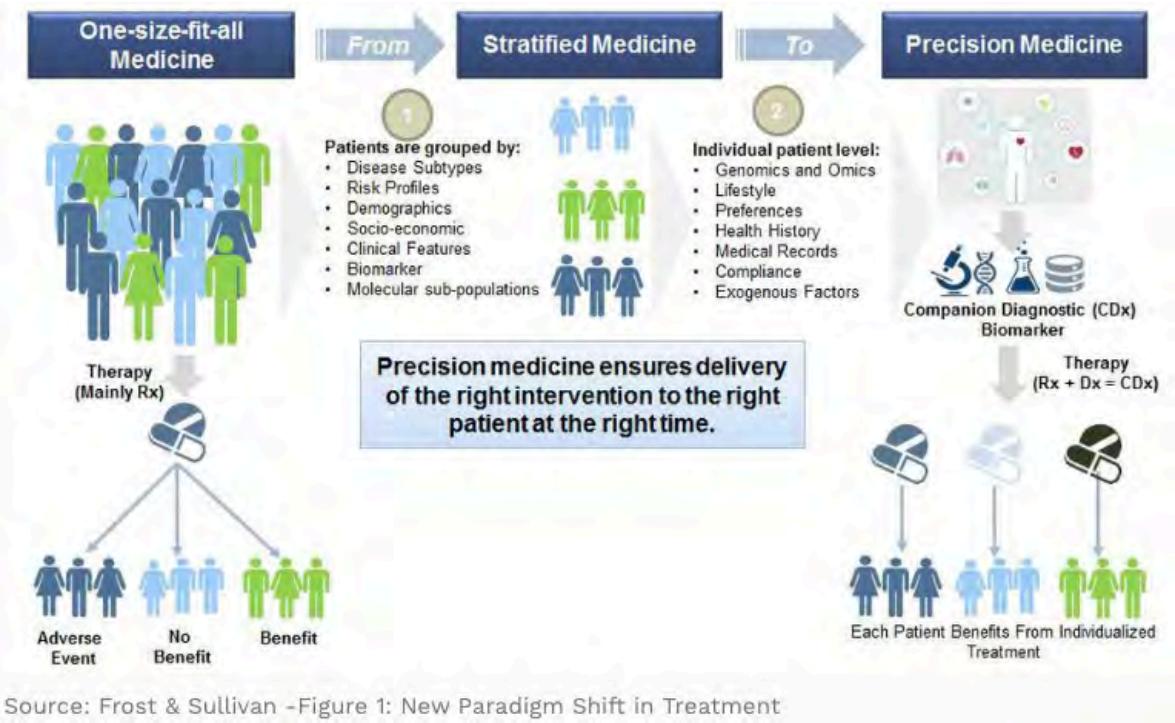
Emergence de la médecine de précision…

→ fournir les meilleurs soins disponibles à chaque patient, sur la base d'une stratification en sous-classes de maladies présentant une base biologique commune.



Shaikh et al., "Translational Radiomics: Defining the Strategy Pipeline and Considerations for Application-Part 1: From Methodology to Clinical Implementation," JACR, 2018.

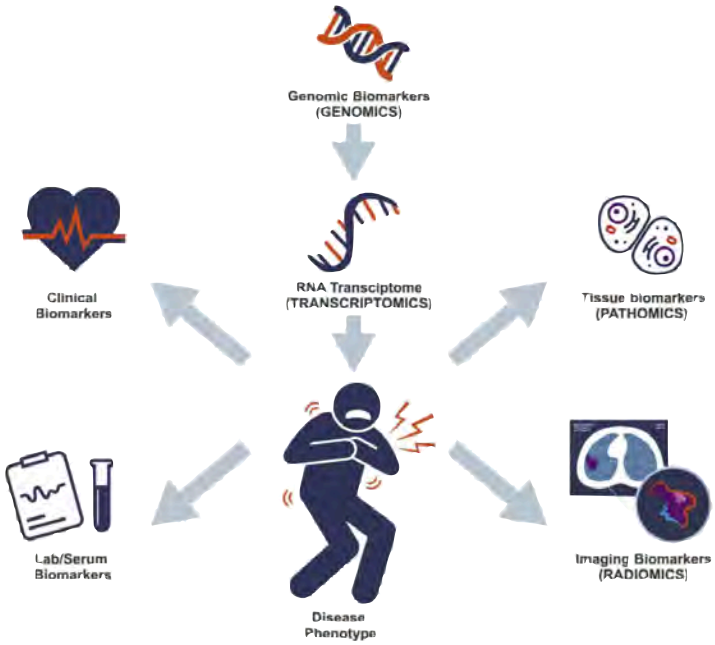
La fin de la médecine « one-size-fit all » annonce celle de la médecine de précision…



→ la bonne intervention au bon patient au bon moment.

Emergence de la médecine de précision...

→ fournir les meilleurs soins disponibles à chaque patient, sur la base d'une stratification en sous-classes de maladies présentant une base biologique commune.

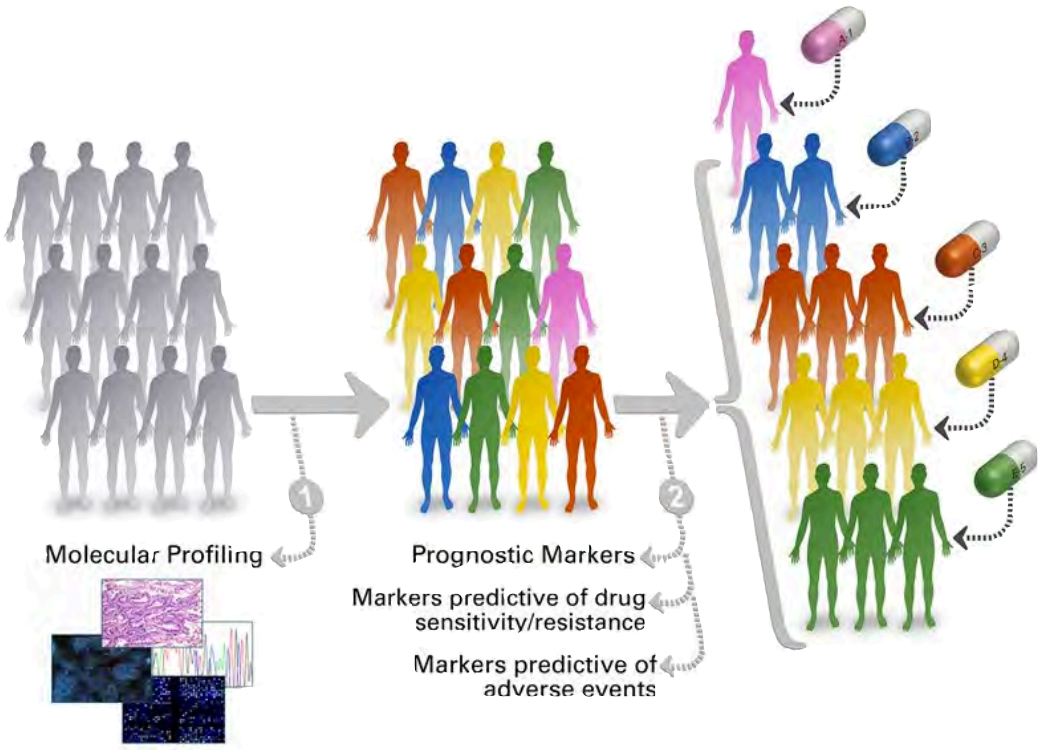


Grp2=Obese-Cardiometabolic Phenotype

Grp1=Natriuretic Peptide Deficiency Syndrome

Grp3=Right Ventricular Failure, Cardiorenal Phenotype

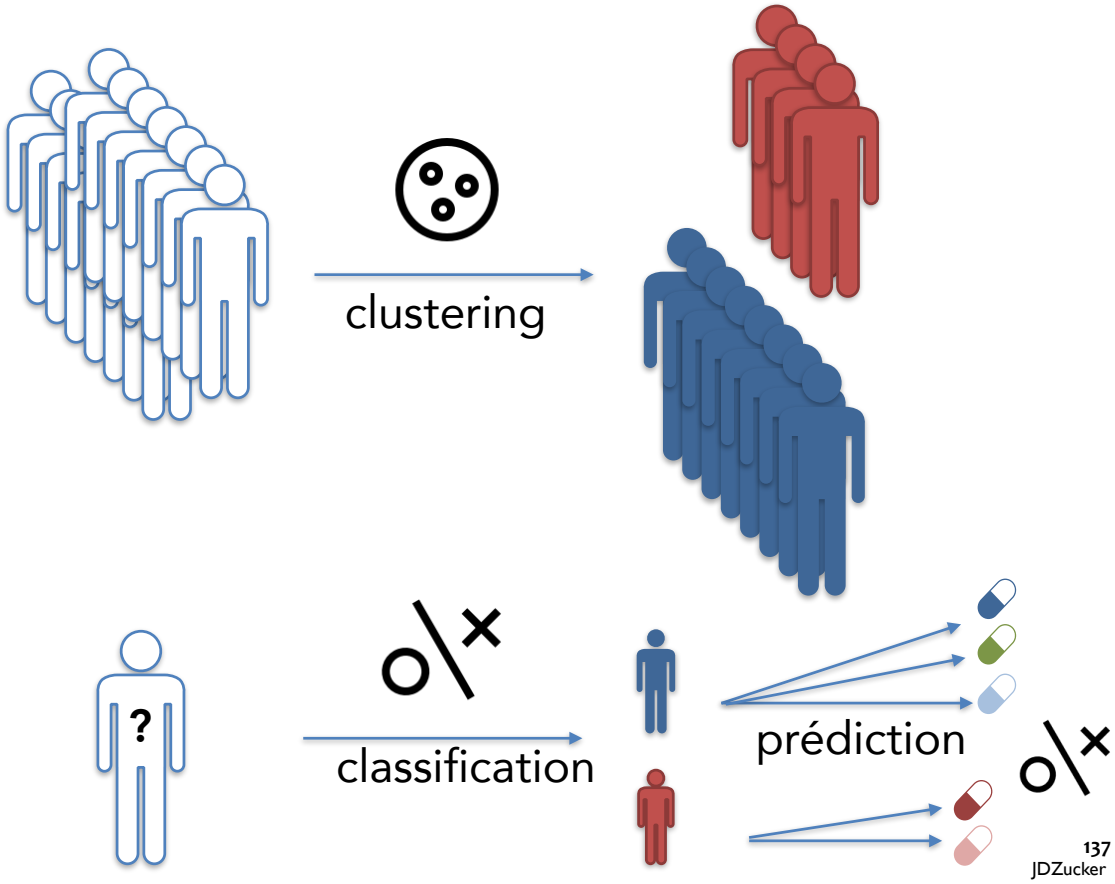
Médecine de précision pour le cancer



→ biomarqueurs tumoraux associés au pronostic du patient et / réponse tumorale au traitement.

https://pct.mdanderson.org/

Médecine de précision et apprentissage automatique:
Clustering pour stratifier et Classifier/prédire pour traiter



Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

III. Exemple de la régression régularisée

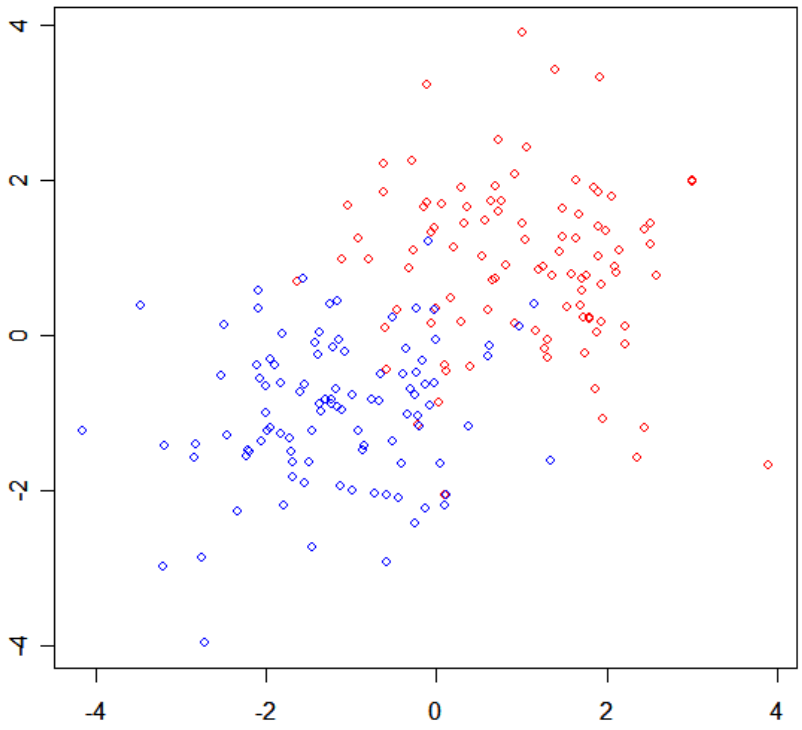
IV. Médecine de précision et IA

V. Classeur avec rejet

VI. Conclusion

Principle

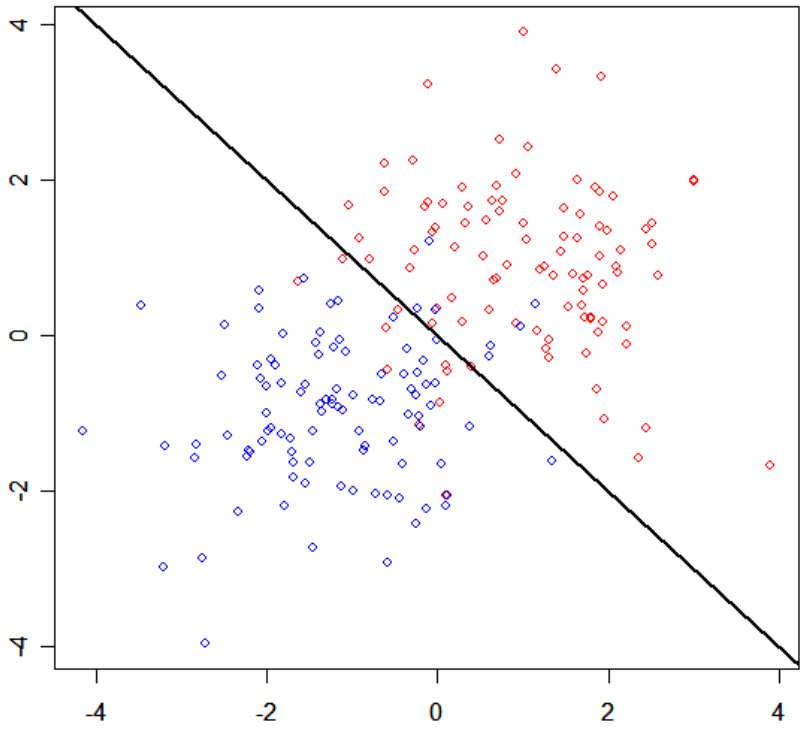
Crédit: Blaise Hanczar



2-class classification problem

Predicted class \ Actual class	P	N
P		
N		

Principle



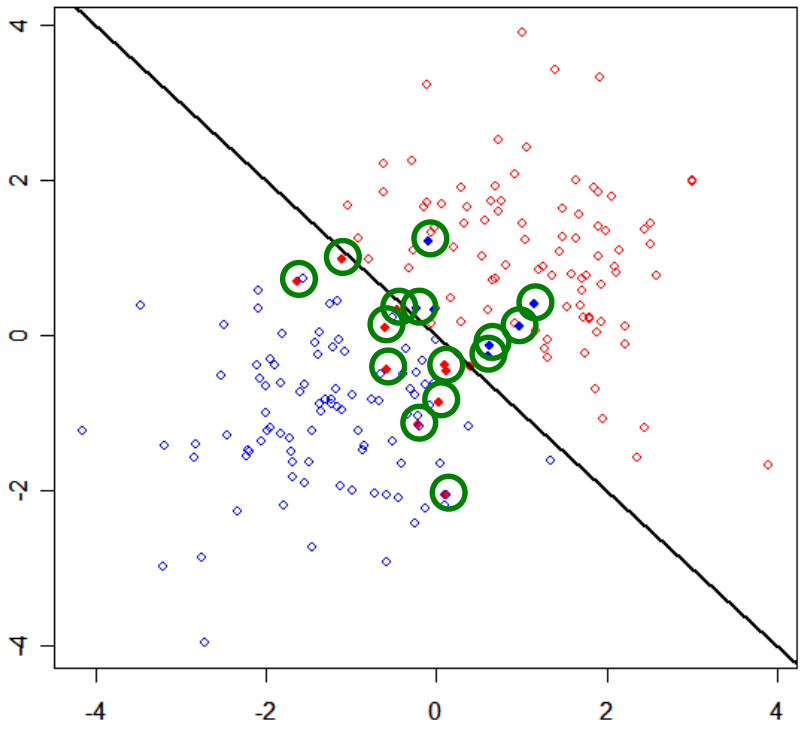
2-class classification problem

Predicted class \ Actual class	P	N
P		
N		

Classifier

$$\psi(x)\colon \mathbb{R}^d \rightarrow \{P, N\}$$

Principle



2-class classification problem

Predicted class \ Actual class	P	N
P		
N		

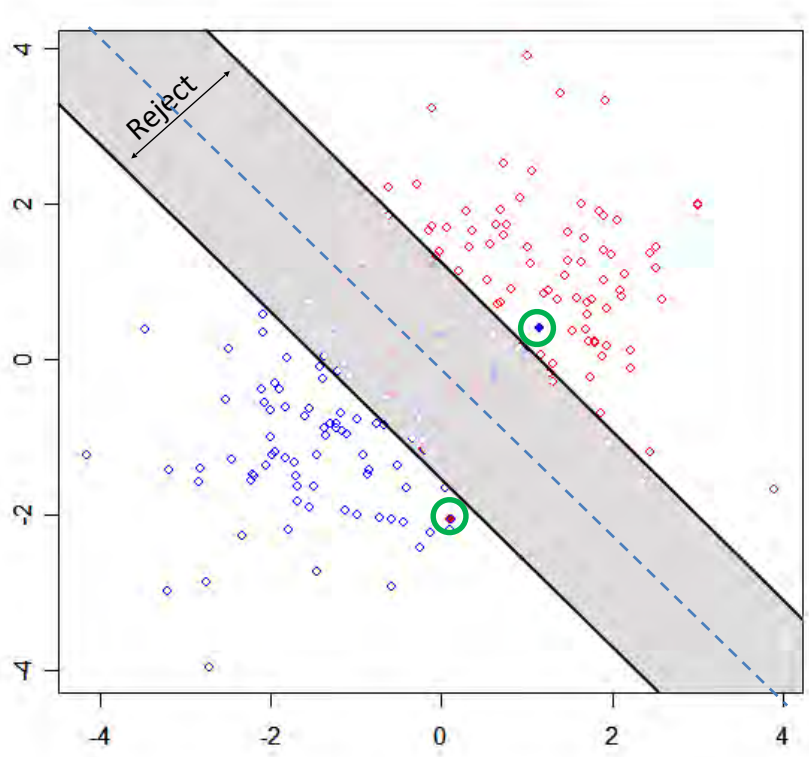
Classifier

$$\psi(x)\colon \mathbb{R}^d \rightarrow \{P, N\}$$

16 misclassifications

All errors are close to the separator

Principle



2-class classification problem

Predicted class \ Actual class	P	N
P		
N		
R		

Classifier with rejection option

$$\psi(x)\colon \mathbb{R}^d \rightarrow \{P, N, R\}$$

2 misclassifications
41 rejections

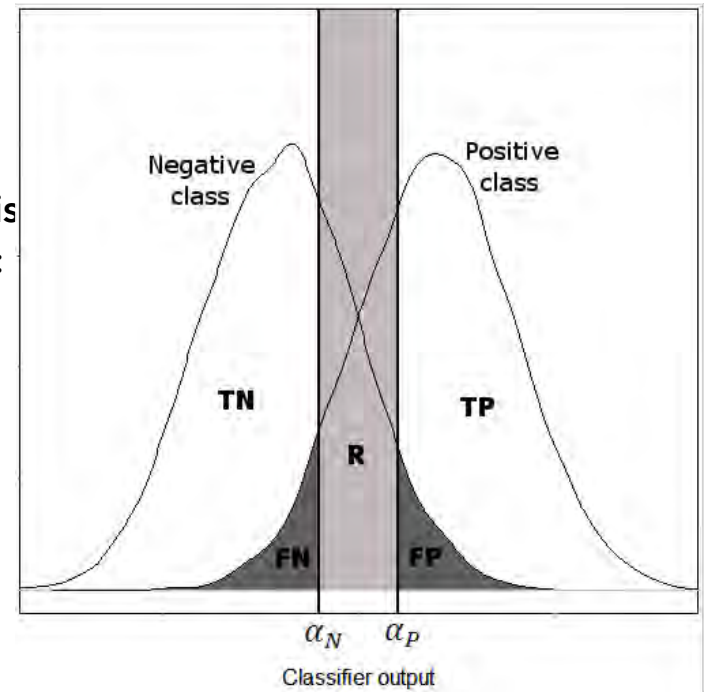
Plug-in methods

Classifier computes a score for each example :

$$\psi : \mathbb{R}^d \to \mathbb{R} \to \{P, N, R\}$$
$$x \mapsto \omega(x) \mapsto \psi(x)$$

Two thresholds are applied on this score to do the class assignation :

$$\psi_{\alpha_P,\alpha_N}(x) = \begin{cases} P & if \quad \omega(x) \geq \alpha_P \\ N & if \quad \omega(x) \leq \alpha_N \\ R & if \ \alpha_N < \omega(x) < \alpha_P \end{cases}$$



The best classifier minimizes :

$$L\left(\psi_{\alpha_P,\alpha_N}\right) = \pi_P\left(\lambda_{TP}TPR + \lambda_{FN}FNR + \lambda_R RPR\right) + \pi_N\left(\lambda_{TN}TNR + \lambda_{FN}FPR + \lambda_R RNR\right)$$

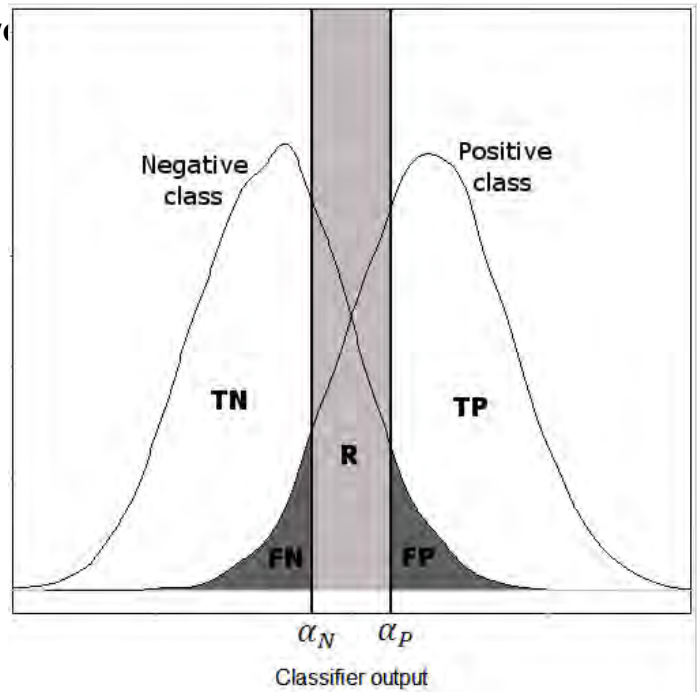
Plug-in methods

Classifier with reject option can also be viewed as a combination of two no-reject classifiers :

ψ_{α_P} is focused on Positive class

ψ_{α_N} is focused on Negative class

$$\psi_{\alpha_P,\alpha_N}(x) = \begin{cases} P \ if \ \psi_{\alpha_P}(x) = \psi_{\alpha_N}(x) = P \\ N \ if \ \psi_{\alpha_P}(x) = \psi_{\alpha_N}(x) = N \\ R \ if \quad \psi_{\alpha_P}(x) \neq \psi_{\alpha_N}(x) \end{cases}$$



The best classifier minimizes :

$$L\left(\psi_{\alpha_P,\alpha_N}\right) = \pi_P\left(\lambda_{TP}TPR + \lambda_{FN}FNR + \lambda_R RPR\right) + \pi_N\left(\lambda_{TN}TNR + \lambda_{FN}FPR + \lambda_R RNR\right)$$

Plan

I. Santé et IA: type de problèmes, type de donnée et types d'apprentissage

II. Des modèles interprétables pour la santé

III. Exemple de la régression régularisée

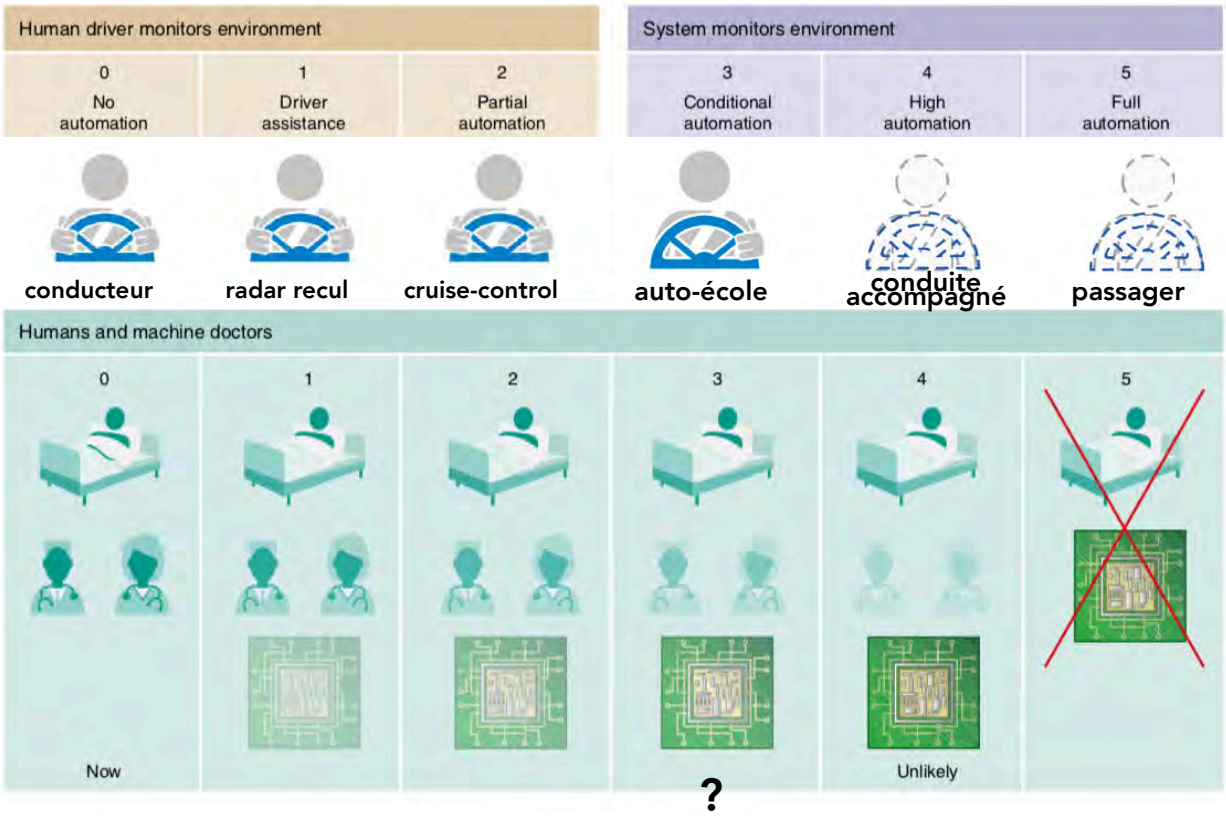
IV. Médecine de précision et IA

V. Classeur avec rejet

VI. Conclusion

Jusqu'à quel point l'IA décidera de notre santé ?

Le but est une synergie, afin que le clinicien reste au commande....



E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat Med, pp. 1–13, Jan. 2019.

Conclusion

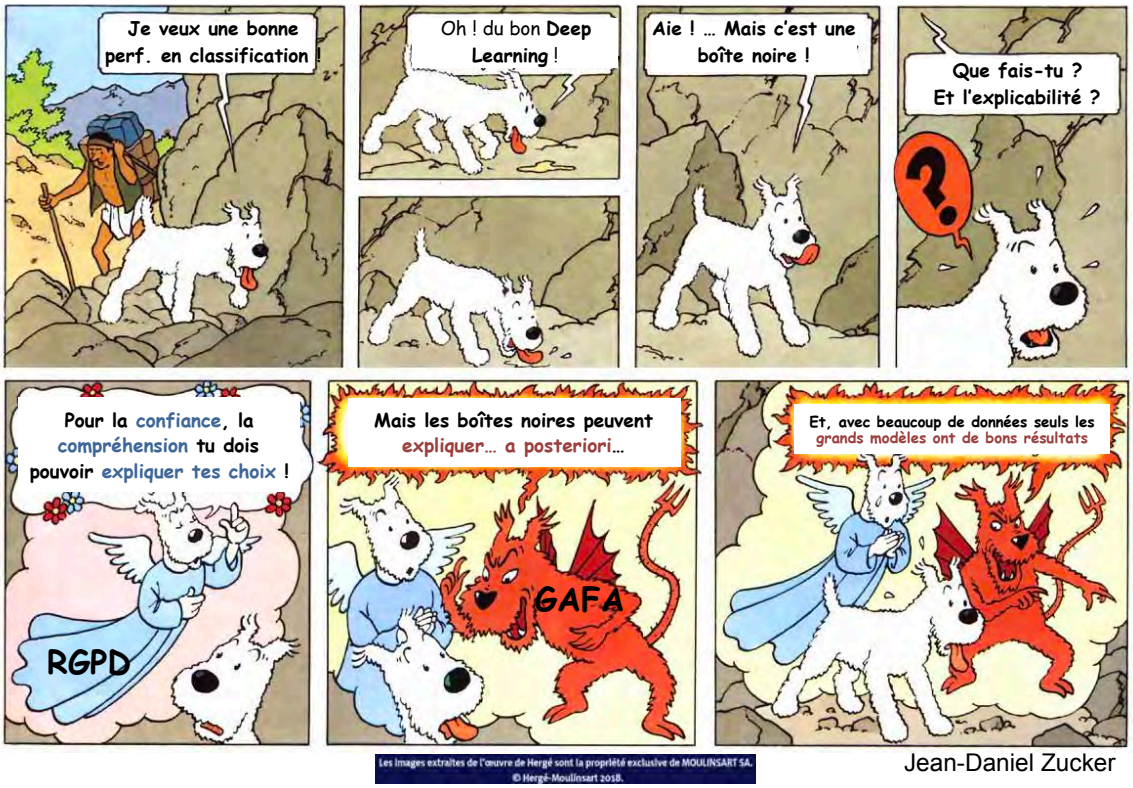
- **La médecine de précision** annonce un bouleversement dans la prise en charge des patients, leur parcours de soin et leur suivi grâce à l'IA.

- N**ouveaux diagnostics moléculaires** (omiques) et **d'imagerie**
  - **stratification** des maladies → meilleurs diagnostic,
  - aide au **prognostic** → meilleurs choix des traitements,
  - **désert médicaux** → tri des patients les plus à risques**.**

- Progrès de l'IA et du **Deep Learning** posent des **questions éthiques** sur son adoption en médecine : équité/confiance/transparence/**interprétabilité**

- **L'IA doit aider les cliniciens** (pas se substituer) à être plus efficace mais **l'interpretabilité** est utile pour éviter les erreurs et contribuer à la recherche de l'étiologie …

Milou, l'interpretabilité et sa consommation de Deep Learning



Jean-Daniel Zucker

Merci

Bibliographie