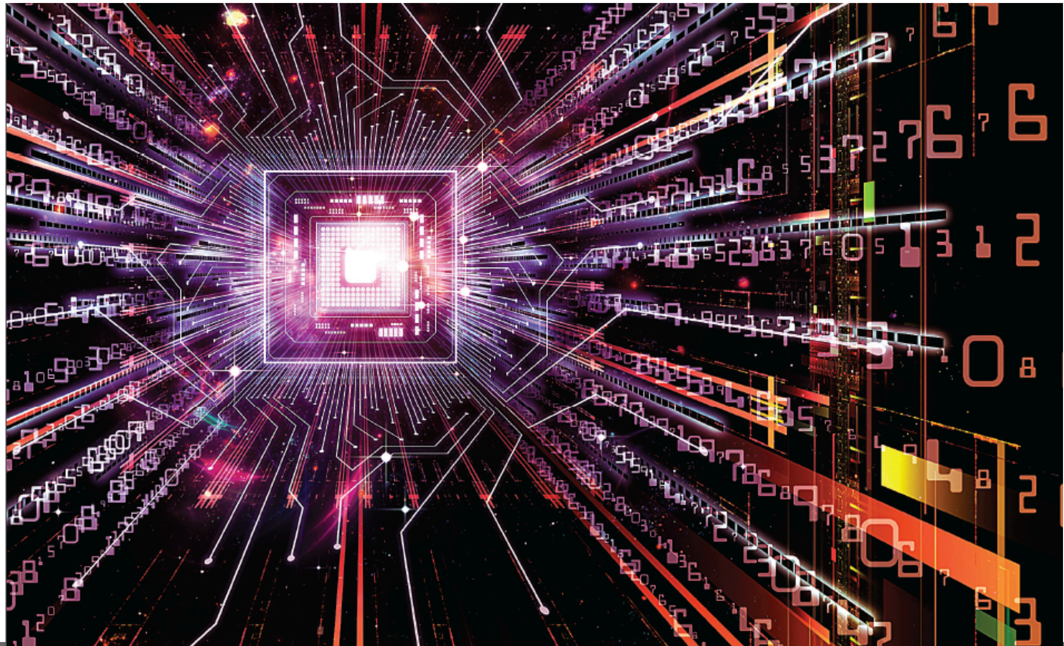
Autour de l'Apprentissage Artificiel

Enggelbert Mephu Nguifo

Usages ?



Apprentissage artificiel

Why ?

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, *A. Big data: The next frontier for innovation, competition, and productivity*. Technical report, **McKinsey Global Institute**, 2011.

"**Machine learning** (a.**k**.a. data mining or predictive analytics) will be the **driver of the next big wave of innovation**"

Apprentissage artificiel

Why ?

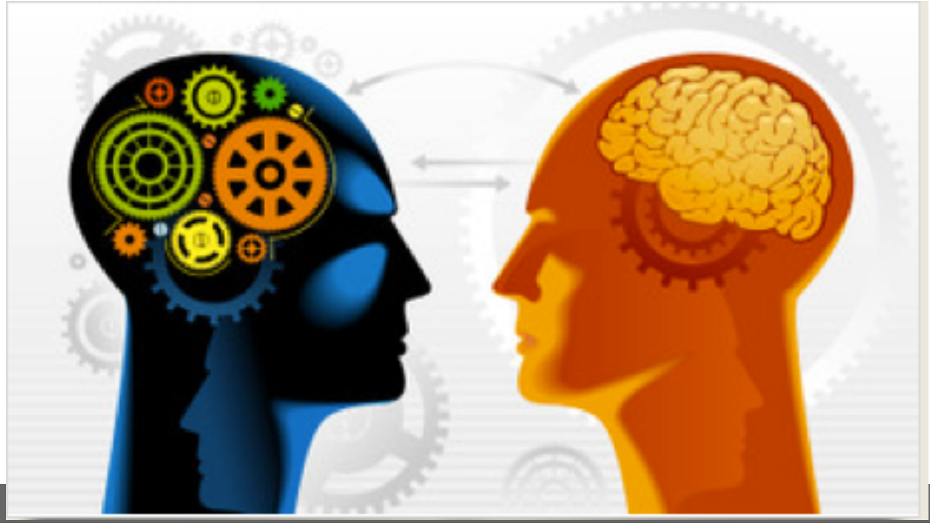
ML algorithms can figure out how to perform important tasks by **generalizing from examples**. This is often feasible and cost-effective where manual programming is not.

As **more data** becomes available, **more ambitious problems** can be tackled.

Apprentissage artificiel

Why ?

ML is widely used in computer science and other fields. However, developing **successful ML** applications requires a substantial amount of **“black art”** that is difficult to find in textbooks.



Apprentissage artificiel

What's in ?

Function approximation

Problem Setting:

- Set of possible instances X
- Unknown target function $f : X \rightarrow Y$
- Set of function hypotheses $H=\{ h \mid h : X \rightarrow Y \}$

superscript: i[th] training example

Input:

- Training examples $\{<x^{(i)},y^{(i)}>\}$ of unknown target function f

Output:

- Hypothesis $h \in H$ that best approximates target function f

http://www.cs.cmu.edu/~tom/10701_sp11/slides/DTreesAndOverfitting-1-11-2011_final.pdf

Apprentissage artificiel

What's in ?

ML = Representation + Evaluation + Optimization

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

Apprentissage artificiel

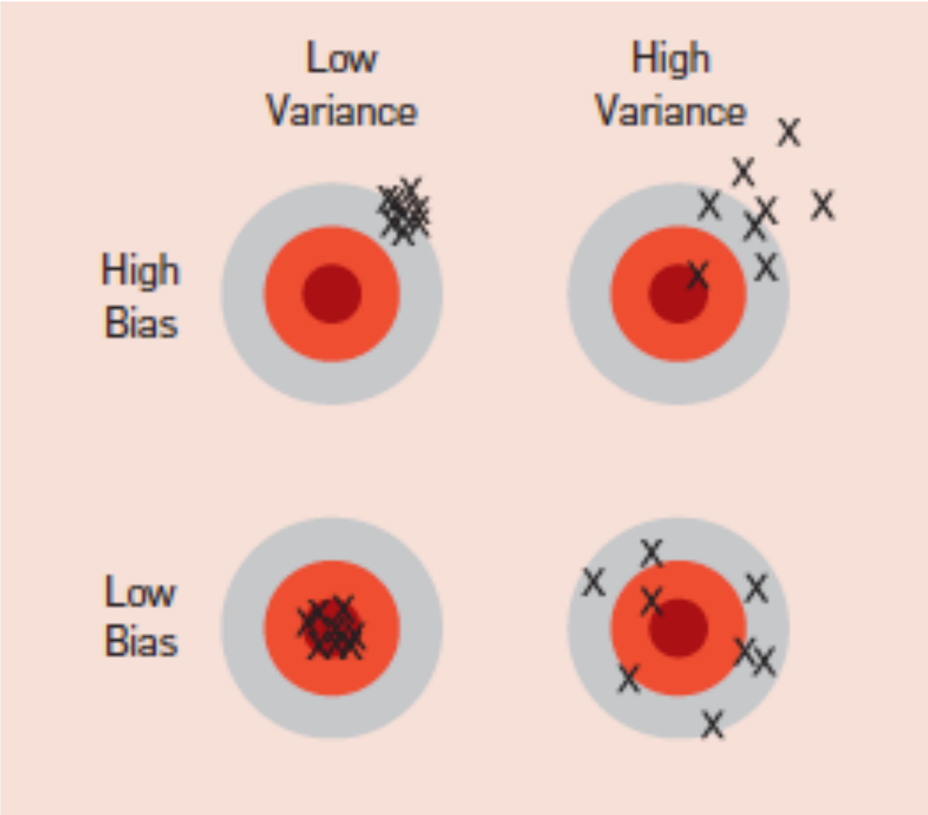
What’s up ?

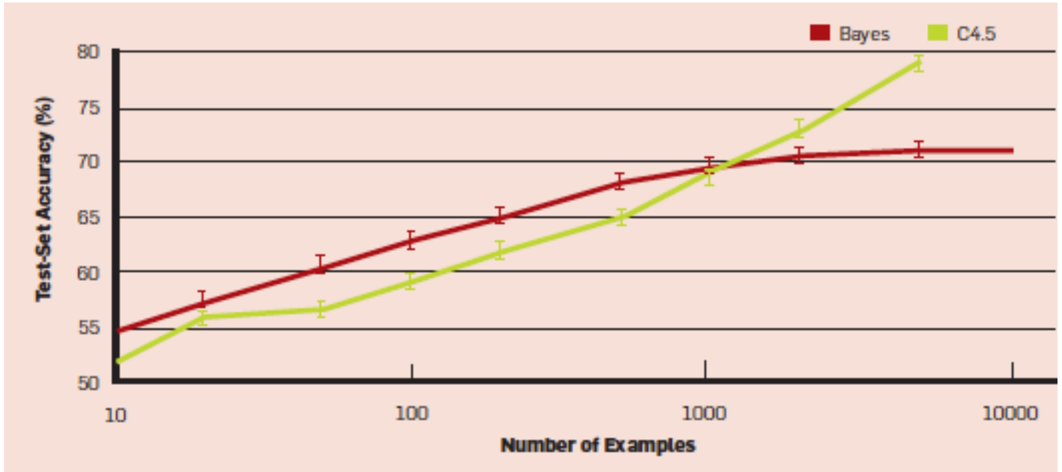
- It’s generalization that counts
- Data alone is not enough
- Overfitting has many faces

Apprentissage artificiel

What's up?

Overfitting has many faces





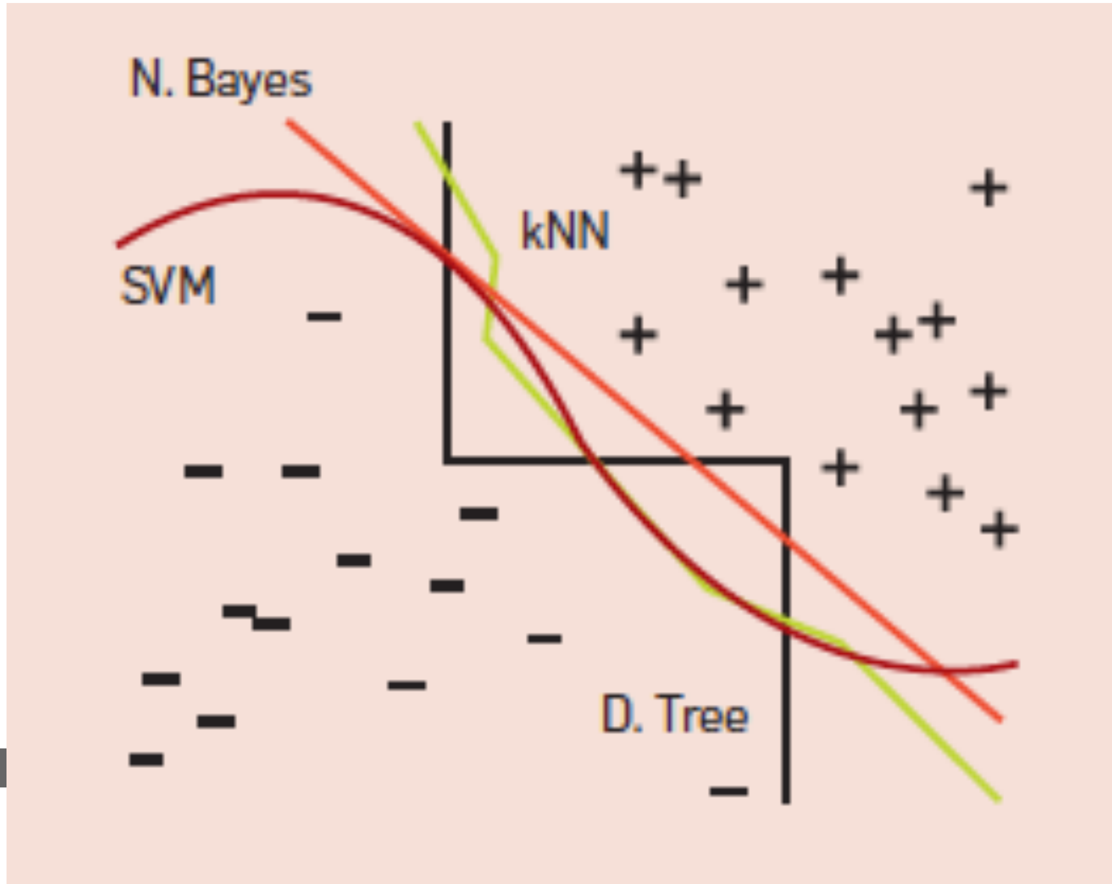
Apprentissage artificiel

- Intuition fails in high dimensions
- Theoretical guarantees are not what they seem
- Feature engineering is the key
- More data beats a clever algorithm

Apprentissage artificiel

What's up ?

- More data beats a clever algorithm



Apprentissage artificiel

- Learn many models, not just one
- Simplicity does not imply accuracy
- Representable does not mean learnable
- Correlation does not imply causation

Apprentissage artificiel

Where ?

- APPs :
  - Robot control
  - Computer vision
  - Speech recognition, Natural language processing
  - Medical outcomes analysis
  - …
- ML niche is growing :
  - Improved machine learning algorithms
  - Increased data capture, networking, new sensors
  - Software too complex to write by hand
  - Demand for self-customization to user, environment

Apprentissage artificiel

Where ?

Pedro DOMINGOS, A Few Useful Things to Know about Machine Learning. *Communications of the ACM*, 55 (10), 78-87, 2012.

Antoine CORNUÉJOLS - Laurent MICLET, "Apprentissage artificiel : Concepts et algorithmes (2ème éd.) », Eyrolles. Juin 2010. 830 pages. ISBN: 978-2-212-12471-2

Ressources

- http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml

- www.kdnuggets.com
- SIGKDD : www.sigkdd.org
- WEKA : www.cs.waikato.ac.nz/ml/**weka**/
- http://www.videolectures.net

- Nuage de mots
  - Outil : http://www.tagxedo.com/app.html