

IA : Machine (Deep) Learning

Introduction

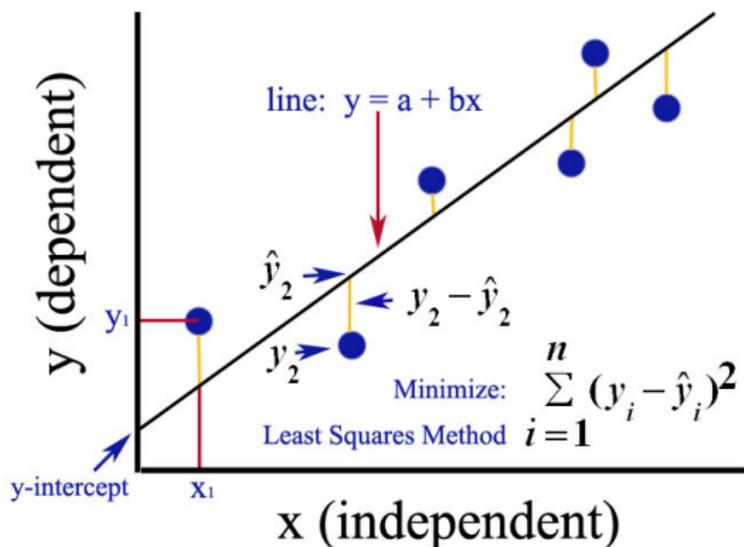
A. Zemmari

LaBRI - Université de Bordeaux

Jan. 09, 2018

Machine Learning ?

Pour demystifier la chose :

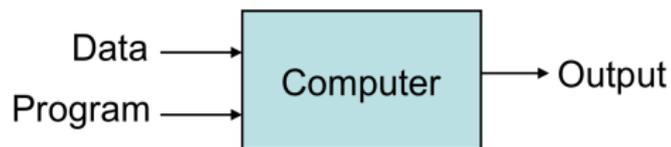


Machine Learning ?

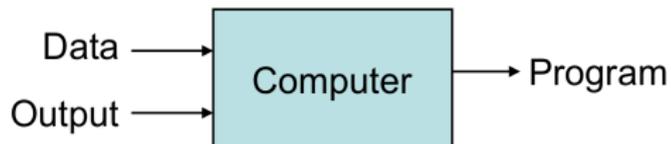
- ▶ "The goal of machine learning is to build computer systems that can adapt and learn from their experience." - Tom Dietterich
- ▶ To learn: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to learn by **experience E** with respect to class of **tasks T** **performance measure P**, if the **performance** at the task T, measured by performance P, **improves by experience E**.

Machine Learning ?

- ▶ Traditional Programming

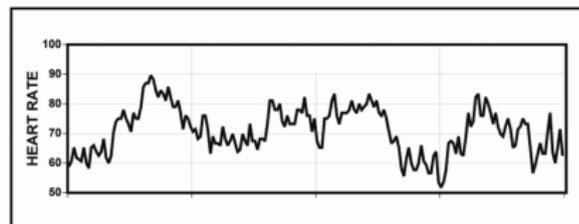


- ▶ Machine Learning



Quelques Exemples

- ▶ A partir des données médicales (sur 20 ans) sur un patient, peut-on dire s'il risque d'avoir une crise cardiaque dans les 2,3 prochaines années ?



Quelques Exemples

- ▶ A partir des données médicales (sur 20 ans) d'un patient, peut-on dire s'il risque d'avoir une crise cardiaque dans les 2,3 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?



125.74	+5.98%	▲	128.34	388,000
23.90	+12.3%	▲	543.23	120,000
15.89	+5.34%	▲	254.23	320,000
25.34	+1.23%	▲	351.54	430,000
7.34	+5.97%	▲	100.08	120,000
1.89	+2.13%	▲	564.23	300,000
1.45	+6.43%	▲	765.90	600,000
1.67	-11.6%	▼	120.34	380,000
1.64	+23.1%	▲	893.23	120,000
1.59	+5.56%	▲	128.98	320,000
1.58	-3.67%	▼	432.12	750,000
1.57	+11.3%	▲	765.23	150,000
1.56	+2.54%	▲	432.24	120,000

Quelques Exemples

- ▶ A partir des données médicales (sur 20 ans) sur un patient, peut-on dire s'il risque d'avoir une crise cardiaque dans les 2,3 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?
- ▶ Est-ce que le gribouillage suivant correspond au nombre 7 ?



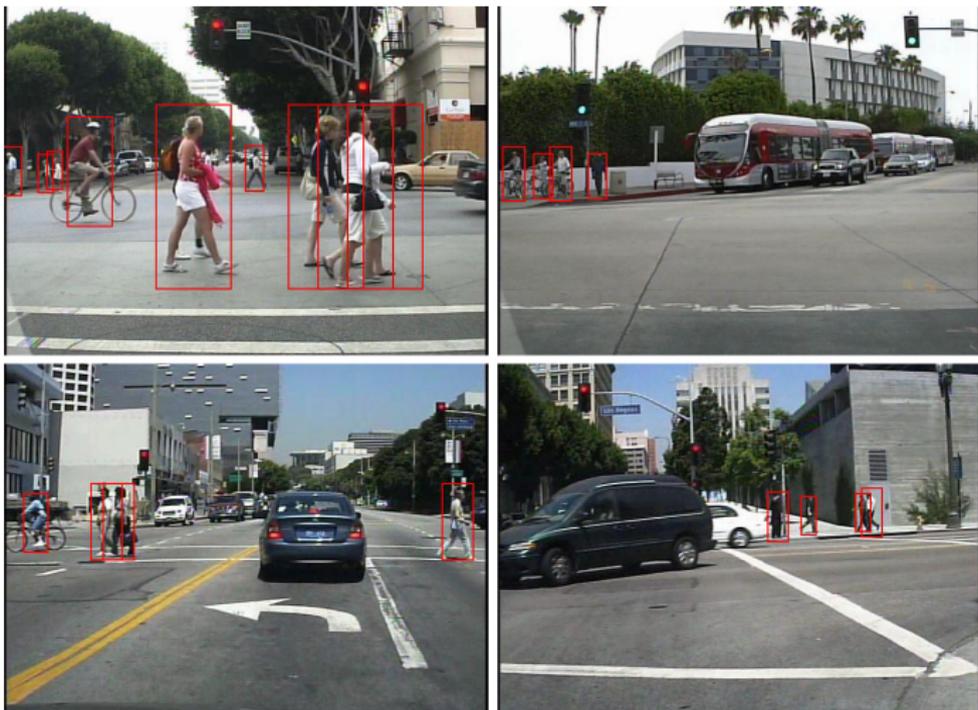
Quelques Exemples

- ▶ A partir des données médicales (sur 20 ans) sur un patient, peut-on dire s'il risque d'avoir une crise cardiaque dans les 2,3 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?
- ▶ Est-ce que le gribouillage suivant correspond au nombre 7 ?
- ▶ Est-ce que ce courriel que je viens de recevoir est un spam ?



Et plus récemment ... le *Deep* ...

Reconnaissance vidéo ¹



¹<https://3c1703fe8d.site.internapcdn.net/newman/gfx/news/hires/2016/newalgorit>

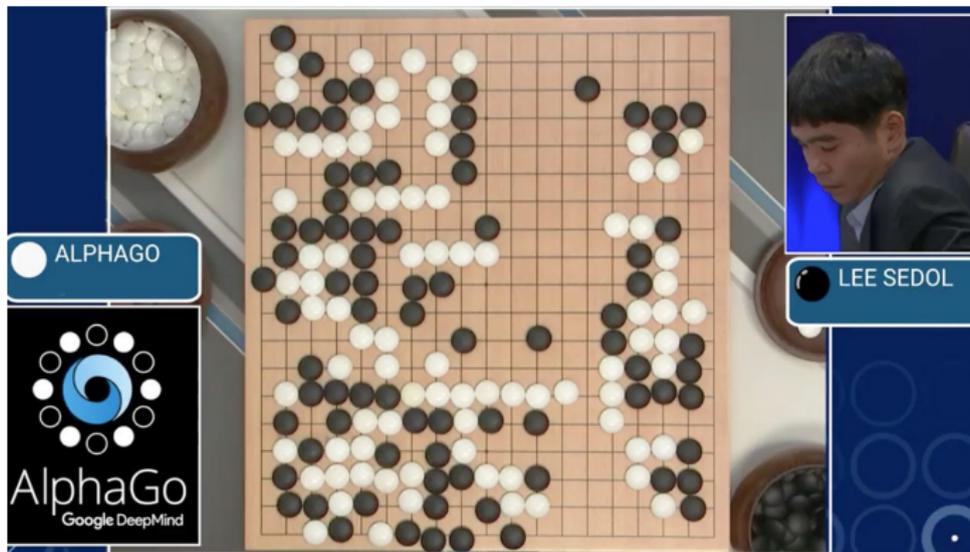
Et plus récemment ... le *Deep* ...

Reconnaissance vocale ²



²<https://technologybux.com/urdu-speech-recognition-no-longer-dream/>

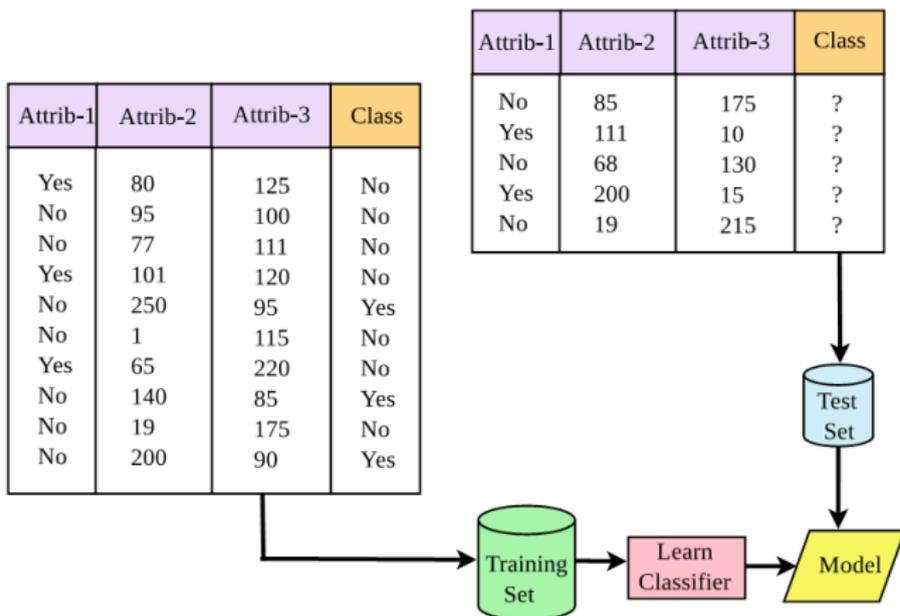
Et plus récemment ... le *Deep* ...



Modèles d'apprentissage

- ▶ **Apprentissage supervisé** :
 - ▶ **Supervision** : les données d'apprentissage (observations) sont accompagnés par les labels indiquant leurs classes
 - ▶ Les nouvelles données sont classifiées en se basant sur le training set

Apprentissage supervisé



Modèles d'apprentissage

- ▶ Apprentissage non supervisé :
 - ▶ Le label de classe des éléments observés (training set) **n'est pas connu**
 - ▶ Le but est de déceler l'existence de classes ou groupes dans les données

Apprentissage non supervisé

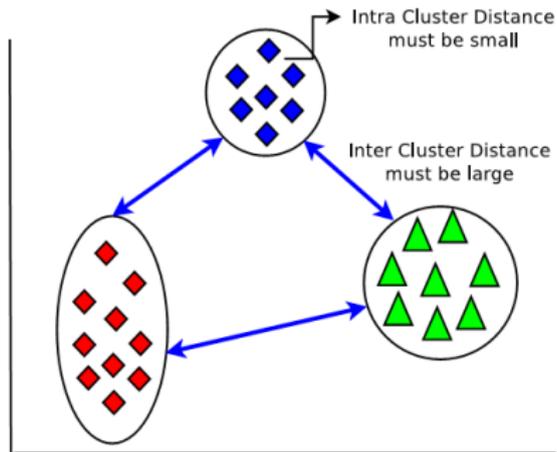


Figure: Unsupervised Learning

Apprentissage supervisé vs non supervisé

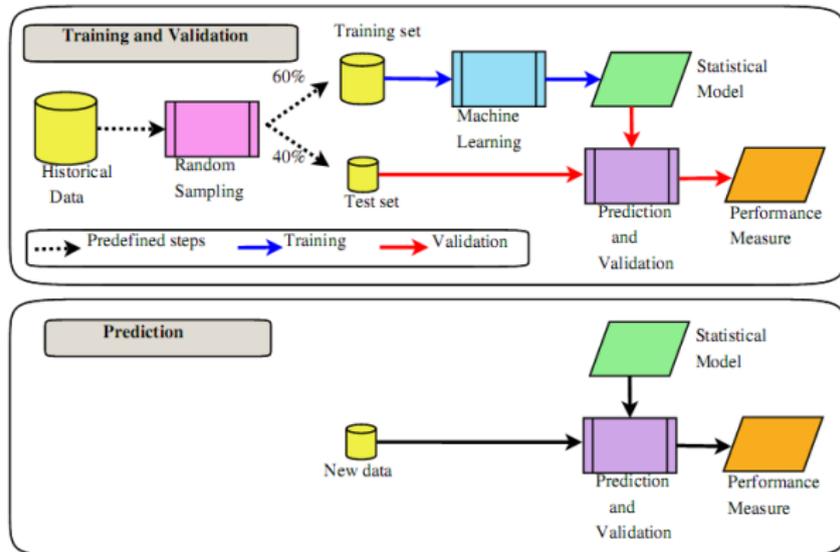
Plus formellement :

Task	Data: based on...	Target: learn...
▶ Supervized	$\mathcal{T} = \{(x_i, y_i)\}_{i=1..n}$	$f(x) = y$
▶ Unsupervised	$\mathcal{T} = \{x_i\}_{i=1..n}$	$x \in X_k$

On se limite ici (dans ce cours) à l'apprentissage supervisé.

Evaluation d'un modèle d'apprentissage (classifieur)

Méthode générale :



Evaluation d'un modèle d'apprentissage (classifieur)

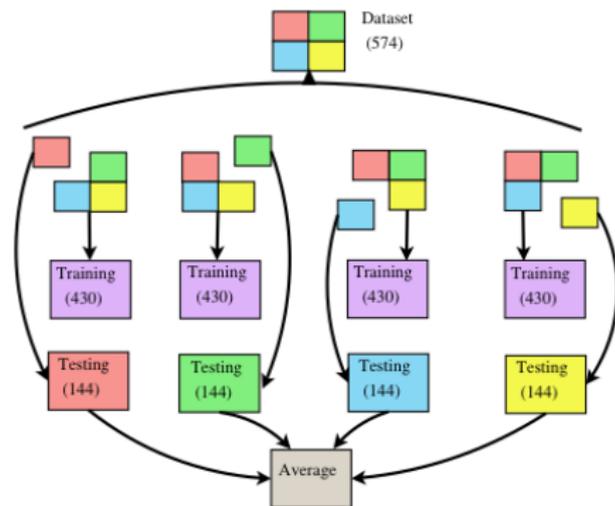
- ▶ Si l'ensemble de données est suffisamment grand :
Découper le dataset en deux sous-ensemble : **Training Set (60%)** et **Test Set (40%)**.
- ▶ Sinon, utiliser la **cross-validation**.

Evaluation d'un modèle d'apprentissage (classifieur)

- ▶ Si l'ensemble de données est suffisamment grand :
Découper le dataset en deux sous-ensemble : **Training Set (60%)** et **Test Set (40%)**.
- ▶ Sinon, utiliser la **cross-validation**.

Evaluation d'un modèle d'apprentissage (classifieur)

Cross-validation :



Evaluation d'un modèle d'apprentissage (classifieur)

Matrice de confusion :

	Actual class (Observation)	
Predicted class (Expectation)	TP	FP
	FN	TN

- ▶ TP : True Positive
- ▶ FP : False Positive
- ▶ FN : False Negative
- ▶ TN : True Negative

Table: Evaluation parameters

$$Precision(P) = \frac{TP}{TP + FP}$$

Ce que nous verrons dans ce cours

- ▶ Naive Bayes classification.
- ▶ Decision Trees and Random Forests.
- ▶ Support vector machines and kernel methods.
- ▶ (Deep) Neural networks.
- ▶ Deep Learning: Convolutional Neural Networks.
- ▶ ...

Ce que nous verrons dans ce cours

Et en pratique :

- ▶ Weka :
à télécharger depuis <https://www.cs.waikato.ac.nz/ml/weka/>
- ▶ Tensorflow
à télécharger et à installer depuis
<https://www.tensorflow.org/install/>