



APPRENTISSAGE MACHINE & DEEP LEARNING

A. Boulch, A. Chan Hon Tong, S. Herbin, B. Le Saux



retour sur innovation

Apprentissage Machine et Deep Learning : Quelques repères

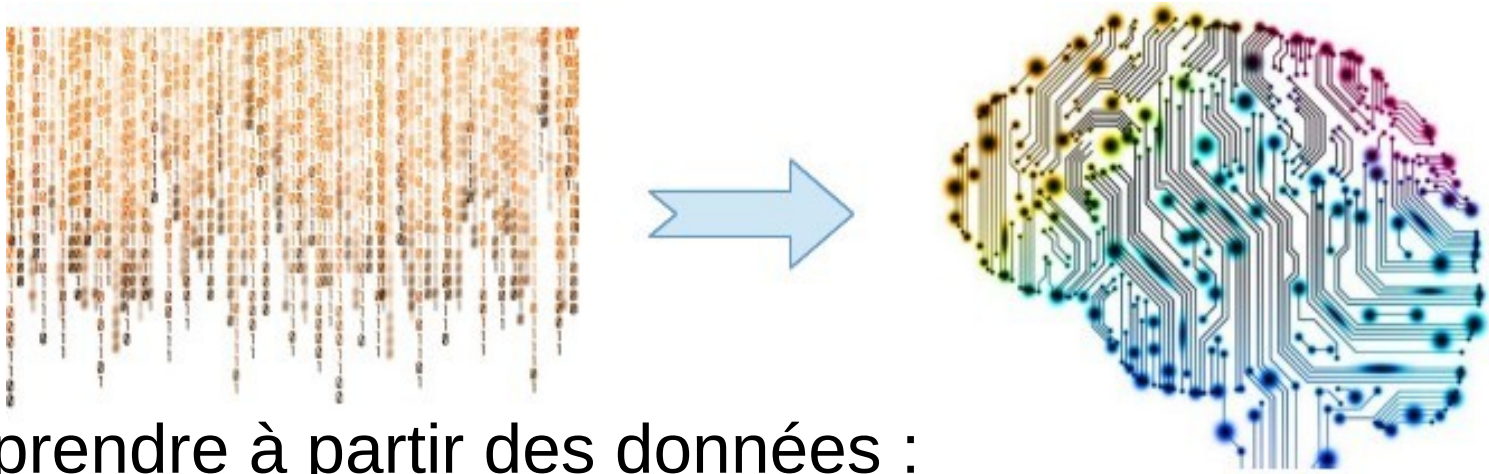
Apprentissage Machine & Deep Learning



Apprentissage machine :

Donner aux ordinateurs la capacité d'apprendre une tâche sans être explicitement programmés à l'exécuter
[Samuel, 59]

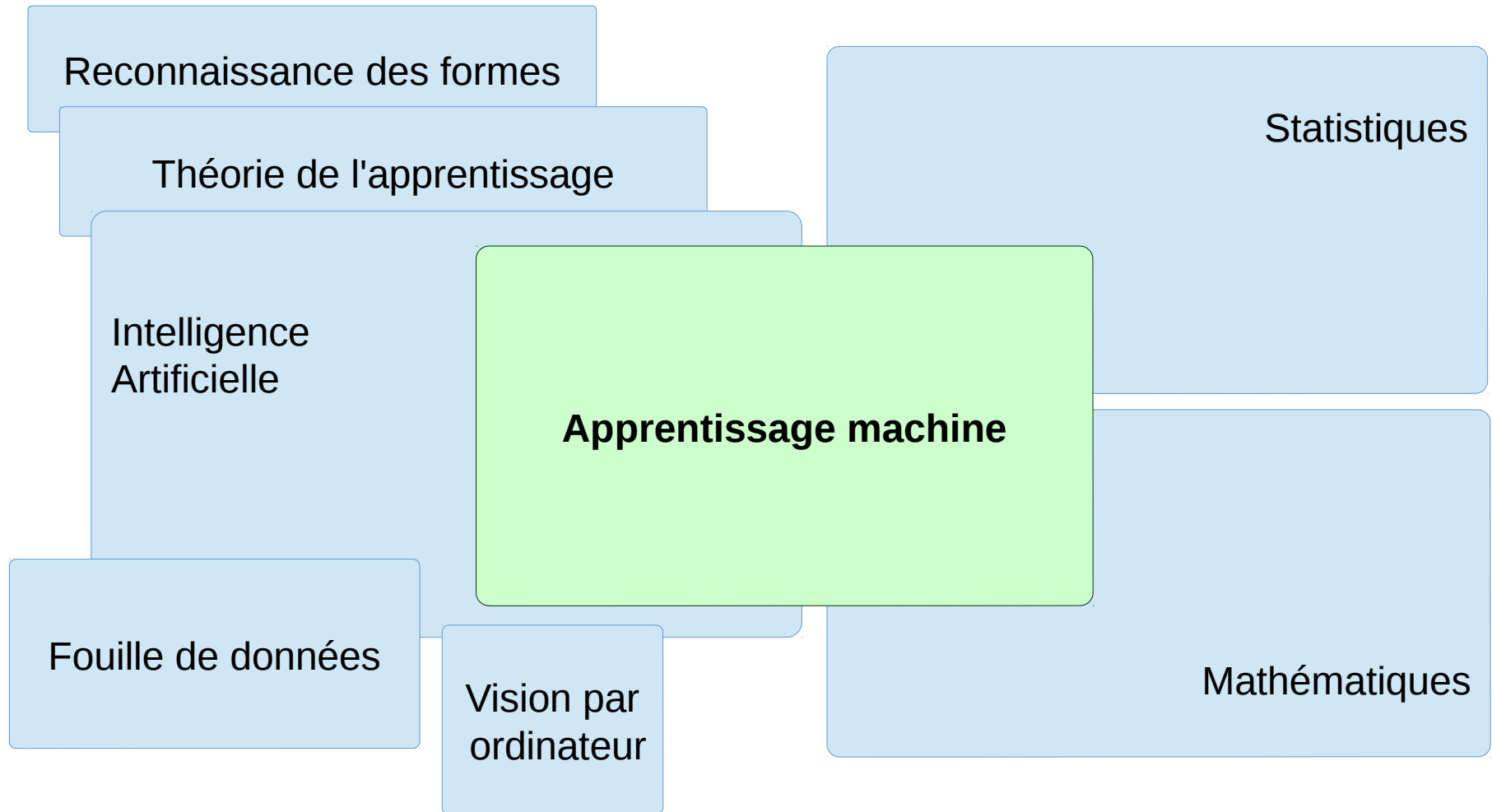
Apprentissage Machine & Deep Learning



Apprendre à partir des données :

- *Comment extraire l'information des données pour nous aider à prendre de meilleures décisions ?*
- *Comment prendre automatiquement des décisions en fonction des données ?*
- *Comment adapter un système à un environnement changeant ?*

Apprentissage Machine & Deep Learning



Apprentissage Machine & Deep Learning

Qq exemples de machine learning :

Anti-Spam (*Classifieur Bayésien*)

1997 : *DeepBlue bat Kasparov*

Tri postal automatique (*détection de chiffres manuscrits par réseaux de neurones*)

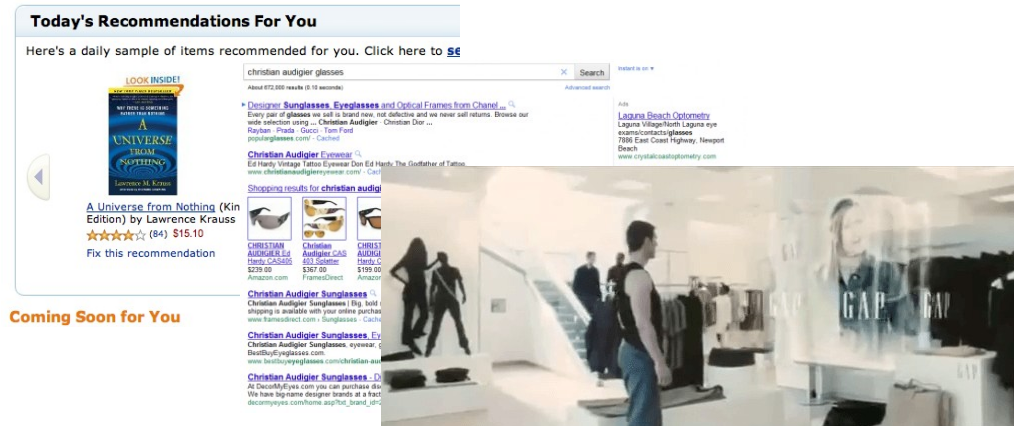


Apprentissage Machine & Deep Learning

Qq exemples de machine learning :

Recommandation ciblée
(*régression logistique*)

Michel, Welcome to Your Amazon.com (if you're not Michel.Trottier)



Appareil photo avec détection
de visages (*boosting*)



Apprentissage Machine & Deep Learning

Pouvons-nous imaginer l'algorithme capable d'exécuter ces tâches ?

- Au mieux immensément compliqué, souvent impossible
- Programmer un ordinateur à apprendre à partir d'exemples
- Fournir des données pour apprendre



Définition de l'apprentissage machine :

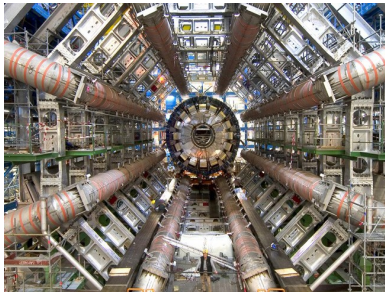
Des structures de données auto-configurables qui permettent à un ordinateur d'exécuter des tâches qui seraient dites “intelligentes” si un humain s'en chargeait.

[Chris Bishop]

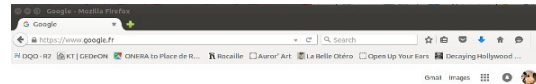
Apprentissage Machine & Deep Learning

Aujourd'hui, un monde de données :

CERN /
Large Hadron Collider
~70 Po/an



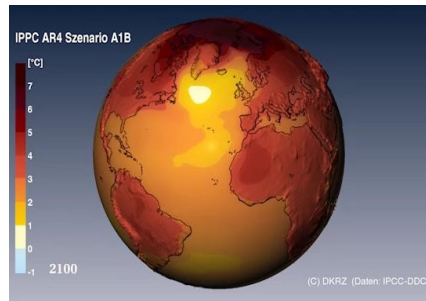
Google :
24 PetaOctets/jour



Copernicus :
> 1Po/an



DKRZ (Climat)
500 Po



Publicité Croquis À propos Confidentialité Conditions Paramètres

Square Kilometer Array
1376 Po/an (en 2024)



Apprentissage Machine & Deep Learning

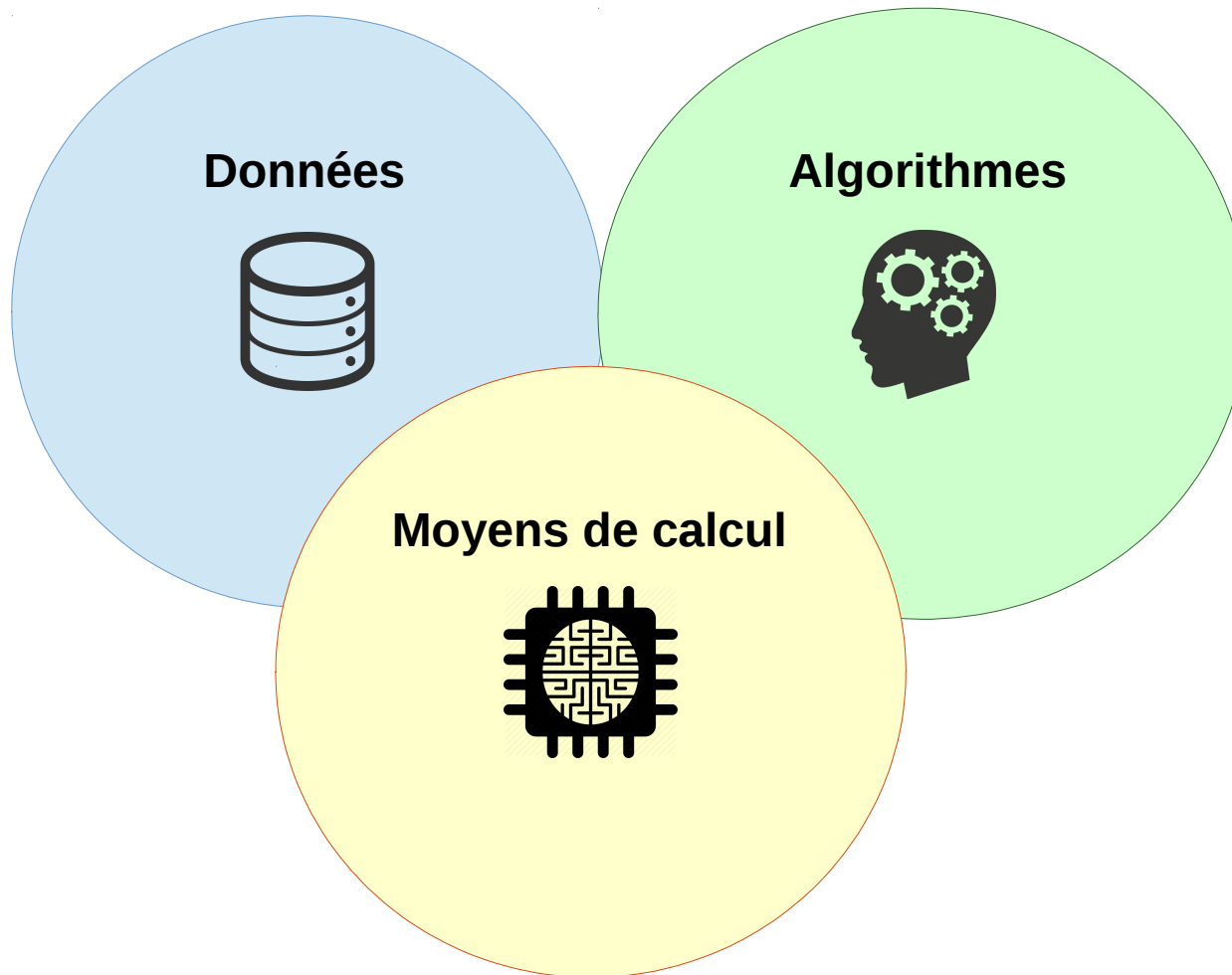
Deep learning:

Une rupture scientifique et technologique en apprentissage

Apprentissage Machine & Deep Learning

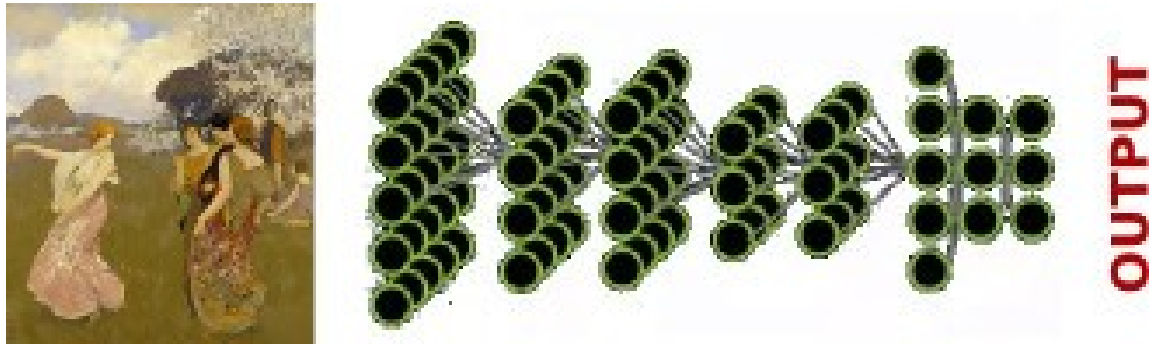
Deep learning:

Une rupture scientifique et technologique en apprentissage



Apprentissage Machine & Deep Learning

“Deep” = *des données aux résultats*

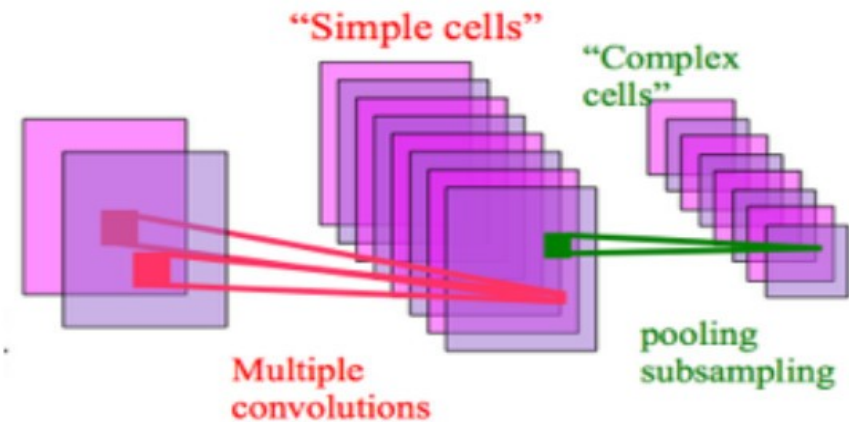


- Modèles hiérarchiques : cascade d'unités de calcul non-linéaire (les neurones)
- Capables d'extraire l'information des données
- Capables d'apprendre une hiérarchie de représentations qui correspondent à différents niveaux d'abstraction

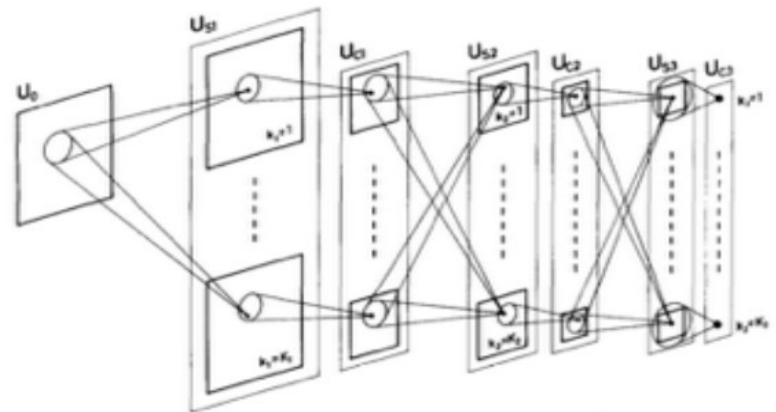
Apprentissage Machine & Deep Learning

Modèles de réseaux de neurones **hiérarchiques** :

Hubel & Wiesel [60s]



Fukushima's Neocognitron [70s]



Figures extraites de LeCun @ CVPR'2015

ONERA

THE FRENCH AEROSPACE LAB

Apprentissage Machine & Deep Learning

Modèles de réseaux de neurones hiérarchiques :

Réseaux convolutifs [80s]

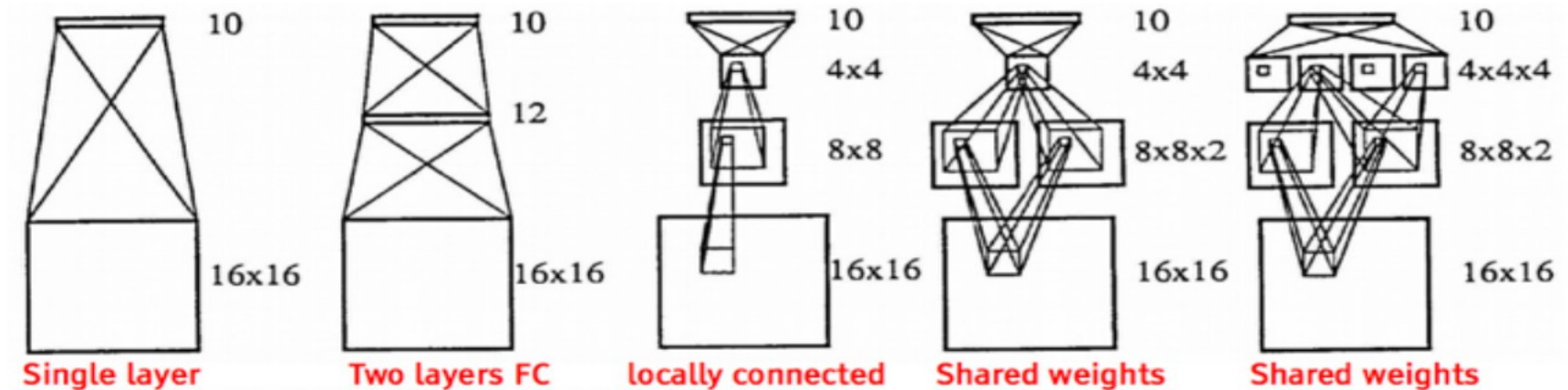


Figure extraite de LeCun @ CVPR'2015

ONERA

THE FRENCH AEROSPACE LAB

Apprentissage Machine & Deep Learning

Modèles de réseaux de neurones hiérarchiques :

Réseaux convolutifs pour la reconnaissance de caractères [90s]

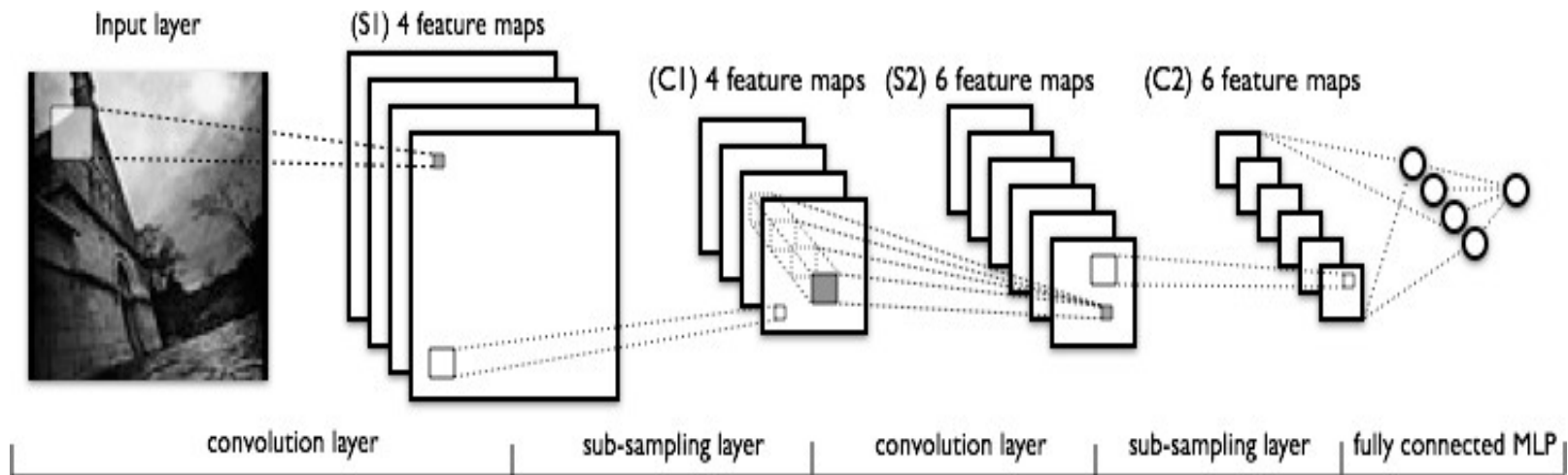


Figure extraite de LeCun IEEE'98

Apprentissage Machine & Deep Learning

Modèles de réseaux de neurones hiérarchiques :

Détecteur de visages convolutif
[2000s]

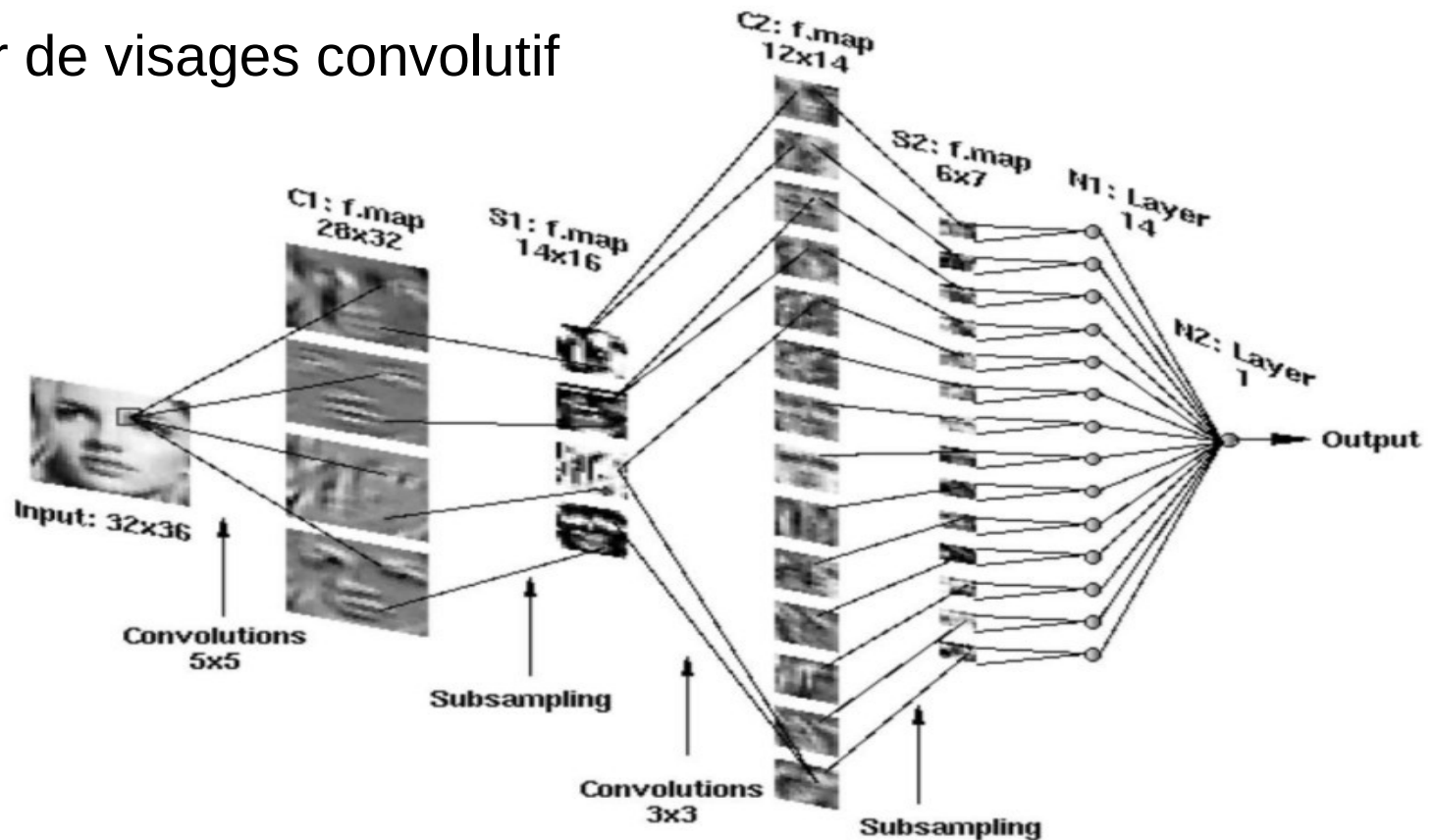


Figure extraite de Garcia PAMI'04

Apprentissage Machine & Deep Learning

Modèles de réseaux de neurones hiérarchiques :

L'age d'or [2010s] : classification Imagenet

2012 Teams	%error	2013 Teams	%error	2014 Teams	%error
Supervision (Toronto)	15.3	Clarifai (NYU spinoff)	11.7	GoogLeNet	6.6
ISI (Tokyo)	26.1	NUS (singapore)	12.9	VGG (Oxford)	7.3
VGG (Oxford)	26.9	Zeiler-Fergus (NYU)	13.5	MSRA	8.0
XRCE/INRIA	27.0	A. Howard	13.5	A. Howard	8.1
UvA (Amsterdam)	29.6	OverFeat (NYU)	14.1	DeeperVision	9.5
INRIA/LEAR	33.4	UvA (Amsterdam)	14.2	NUS-BST	9.7
		Adobe	15.2	TTIC-ECP	10.2
		VGG (Oxford)	15.2	XYZ	11.2
		VGG (Oxford)	23.0	UvA	12.1
ConvNet / non-ConvNet					

Figure extraite de LeCun @ CVPR'2015

→ *Performants pour une multitude de tâches (autres que vision), avec des jeux de données variés.*

Apprentissage Machine & Deep Learning

Deep learning, *de nouveaux algorithmes pour de vieux problèmes* :

Classification : réseaux de neurones, machines à vecteurs de support
→ *réseaux convolutifs*

Catégorisation : analyse en composantes principales, k-means
→ *auto-encoders, machines de Boltzmann*

Régression : linéaire, non-paramétrique
→ *mean-square loss nets*

Apprentissage Machine & Deep Learning

Objectifs de la journée :

- Maîtriser les notions sous-jacentes à l'analyse de donnée
- Comprendre dans le détail plusieurs algorithmes d'apprentissage
- Permettre concrètement la mise en oeuvre de ces approches

Apprentissage Machine & Deep Learning

Agenda de la matinée : cours / tutoriels

9h15 – 10h	Bertrand Le Saux : Principes de l'apprentissage machine (45mn)
10h – 10h45	Stéphane Herbin : Modèles d'apprentissage standard (45mn)

pause (15mn)

11h – 11h45	Adrien Chan-Hon-Tong : Réseaux de neurones (architecture, optimisation) (45mn)
11h45 – 12h30	Alexandre Boulch : Deep Learning (45mn)

repas (1h)

Apprentissage Machine & Deep Learning

Agenda de l'après-midi : conférences invitées / table ronde

13h30 – 14h **Sidonie Lefèvre** (ONERA/DOTA) :

Exemples de mise en œuvre de méthodes d'apprentissage pour l'étude de la signature infrarouge d'aéronefs (30mn)

14h – 15h **Olivier Sigaud** (UPMC/ISIR) :

From Machine Learning to Deep Learning with a focus on regression and reinforcement learning (1h)

pause (15mn)

15h15 – 15h45 **Thibaut Castaings** (ONERA/DCPS) :

Exemples d'applications aux observations spatiales (30mn)

15h45 – 16h15 Table ronde et discussion de clôture (45mn)

Bases de l'apprentissage machine

Bases de l'apprentissage machine

Thèmes abordés :

Extraction de caractéristiques

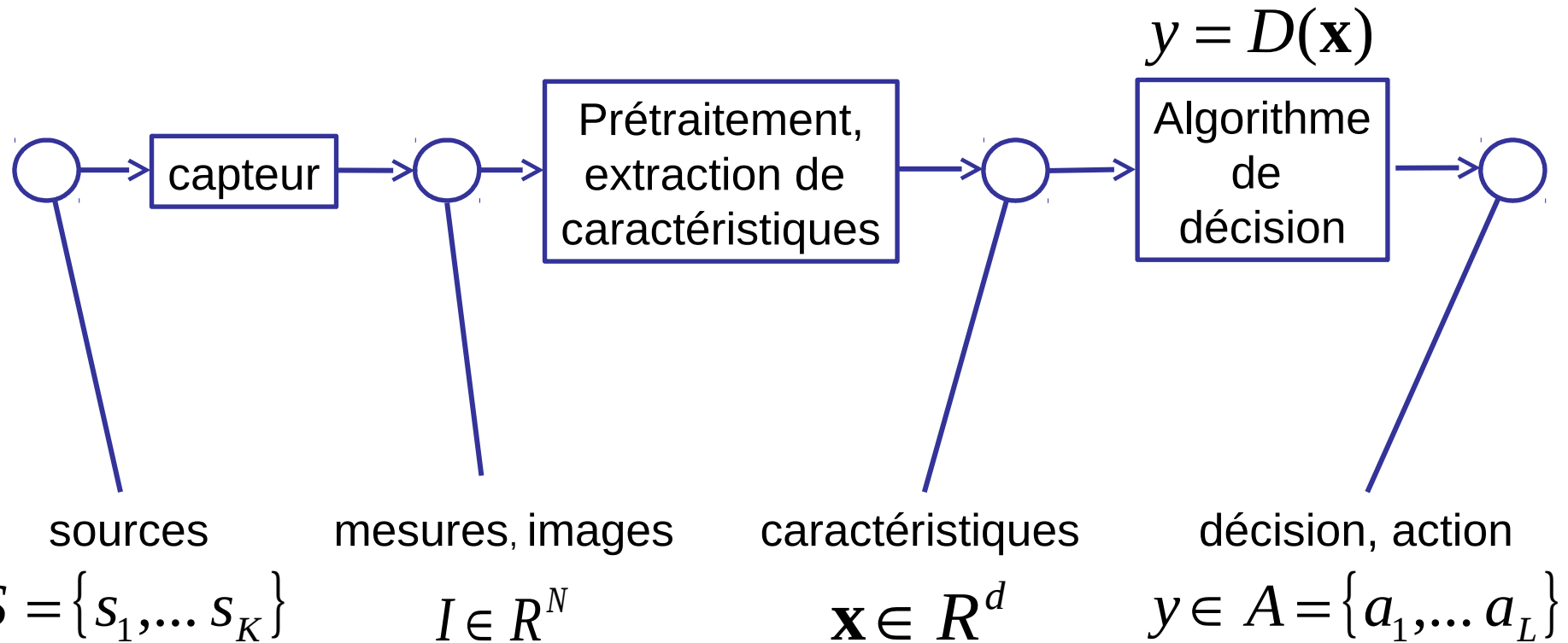
Sélection de caractéristiques et réduction de dimensionnalité

Catégorisation (non-supervisée)

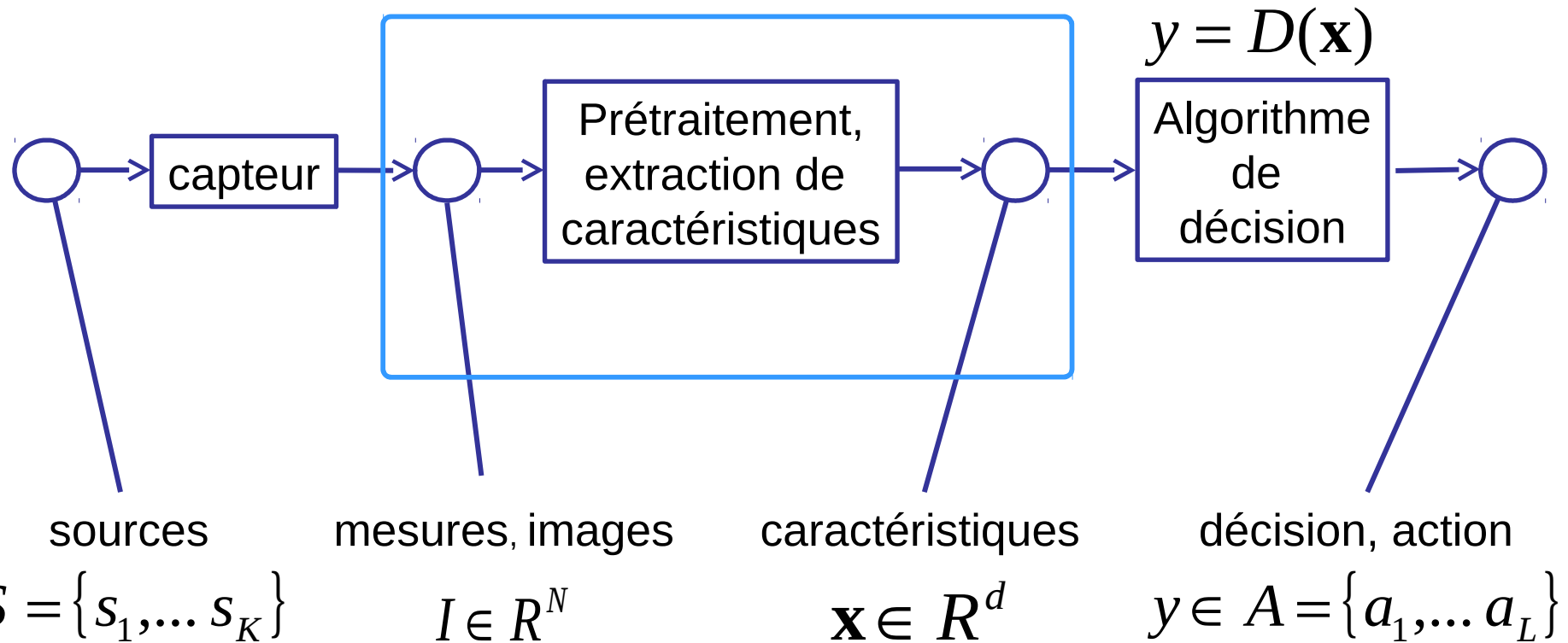
Classification supervisée : principes

Classification supervisée : Machine à Vecteurs de Support

La chaîne de reconnaissance des formes



La chaîne de reconnaissance des formes



Extraction de caractéristiques

Objectif : Passer d'un ensemble de mesures à une représentation concise des données

Propriétés des bonnes caractéristiques :

- Informatives et non-redondantes
- Faciliter la tâche suivante (par ex. apprentissage)
- Améliorer la capacité de généralisation et d'interprétabilité (en supprimant le bruit de mesure)

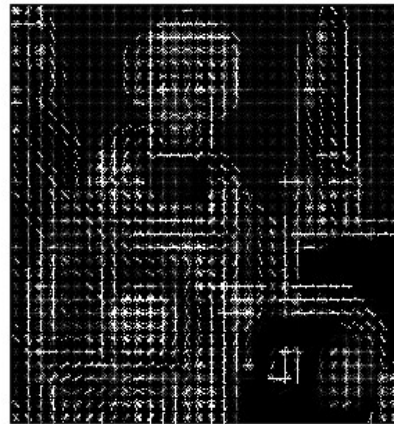
Extraction de caractéristiques

Caractéristiques des signaux images : points d'intérêt, contours, couleur, texture, flot optique...

Input image



Histogram of Oriented Gradients



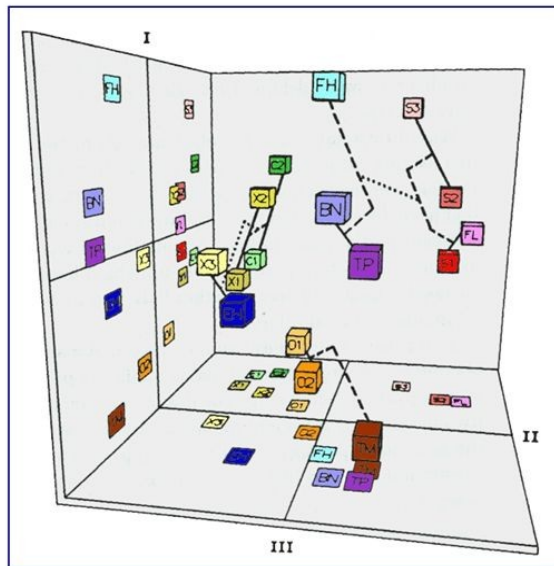
Histogramme d'orientation de gradients



SIFT (points & descripteurs locaux)

Extraction de caractéristiques

Caractéristiques des signaux audio : fréquence, pitch, bpm, timbre, spectrogrammes, transf. Fourier...



- BN - Bassoon
- C1 - E flat Clarinet
- C2 - B flat Bass Clarinet
- EH - English Horn
- FH - French Horn
- FL - Flute
- O1 - Oboe
- O2 - Oboe (different instrument and player)
- S1 - Cello, muted *sul ponticello*
- S2 - Cello
- S3 - Cello, muted *sul tasto*
- TM - Muted Trombone
- TP - B flat Trumpet
- X1 - Saxophone, played *mf*
- X2 - Saxophone, played *p*
- X3 - Soprano Saxophone

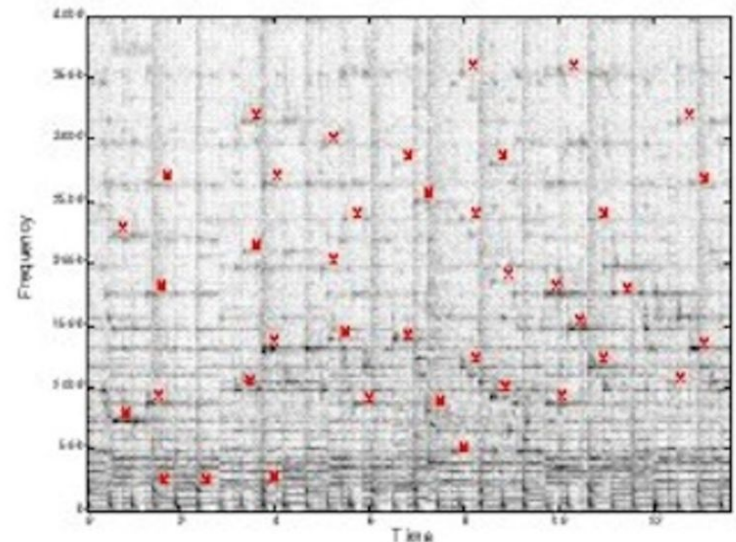
- Grey's Timbre cube
- Is it possible to classify timbres, for example recognition of instrument from audio?

- Dimension I: spectral energy distribution, from broad to narrow
- Dimension II: timing of the attack and decay, synchronous to asynchronous
- Dimension III: amount of inharmonic sound in the attack, from high to none

image by Hans-Christoph Steiner based on Grey, JM 1979, JASA, 61, 1270

Espace des timbres de Grey

Timbre
space



Shazam : constellation
de points d'intérêt du
spectrogramme

Extraction de caractéristiques

Construction experte de caractéristiques :

- Importance de l'expertise métier (*ex: connaissance des modèles physiques sous-jacents*)
- Connaissance et familiarité avec la donnée (*data science*)
- Adéquation avec le modèle / la tâche à exécuter
- Procédure d'essai / erreur itérative

Extraction de caractéristiques

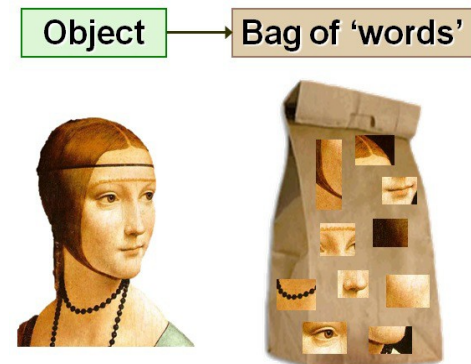
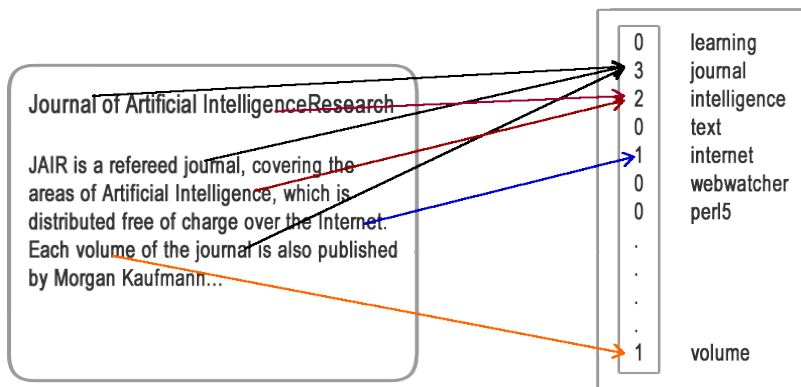
Apprentissage de caractéristiques :

- Données réelles complexes, redondantes
- Découvrir une structure de faible dimension inhérente aux données en grande dimension
- Supervisé : réseau de neurones multi-couche...
- Non-supervisé : K-moyennes, ACP

Extraction de caractéristiques

Apprentissage de dictionnaire :

- Estime un dictionnaire (= ensemble d'éléments de base) qui représente un ensemble de données
- Encode chaque donnée en fonction du dictionnaire (par une combinaison linéaire)
- Contrainte de parcimonie (*sparsity*): une donnée est représentée par seulement quelques éléments du dictionnaire.
- Approches non-supervisées (K-SVD) ou supervisées (SPAMS...)



Caractérisation par “sac-de-mots”

Sélection de caractéristiques

Réduction de dimension

Problèmes de l'analyse de données :

- Données de grande dimension
 - Info redondante, non pertinente
 - Fléau de la dimension : en grande dimension les observations (mesures) sont des points isolés dans un espace vide
- Grand volume de données → temps de calculs longs

Objectif : sélectionner un ensemble réduit de variables de faible dimension utiles pour représenter les données

Sélection de caractéristiques

Réduction de dimension

Exemple : visualiser et analyser la structure de protéines

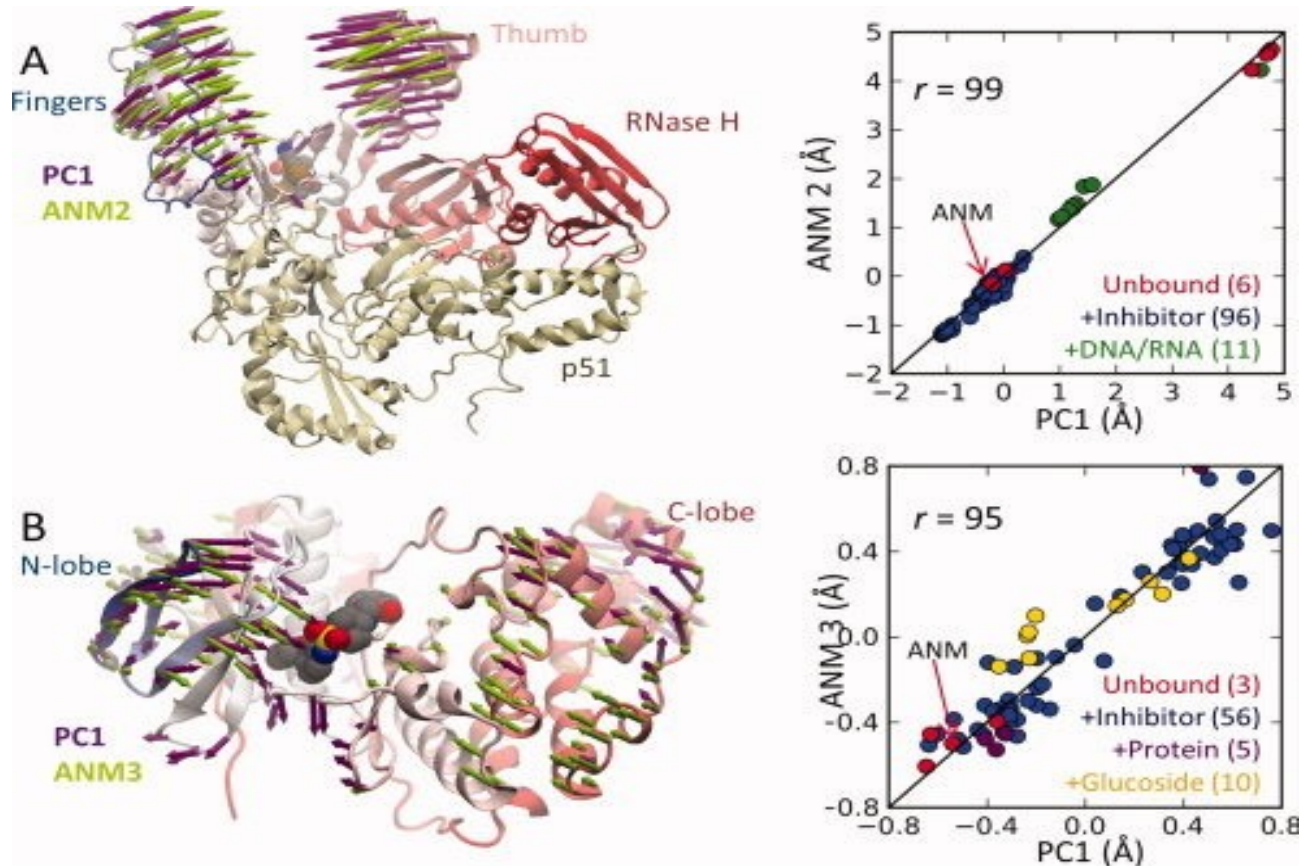


Figure extraite de Morelles et al., Science 2011

Sélection de caractéristiques

Réduction de dimension

Propriétés de caractéristiques bien sélectionnées :

- Avoir un modèle simple des données, humainement compréhensible
- Temps d'estimation et d'apprentissage réduits
- Éliminer l'information non pertinente pour permettre une bonne généralisation

→ motivations connexes à l'extraction de caractéristiques

Sélection de caractéristiques

Approche #1 : Analyse en Composantes Principales (ACP)

- Changement de base de l'espace de représentation de manière à maximiser la variance des données sur les 1ères dimensions
- Algorithme basé sur la diagonalisation de la matrice de variance-covariance centrée

Lien avec la théorie du signal :

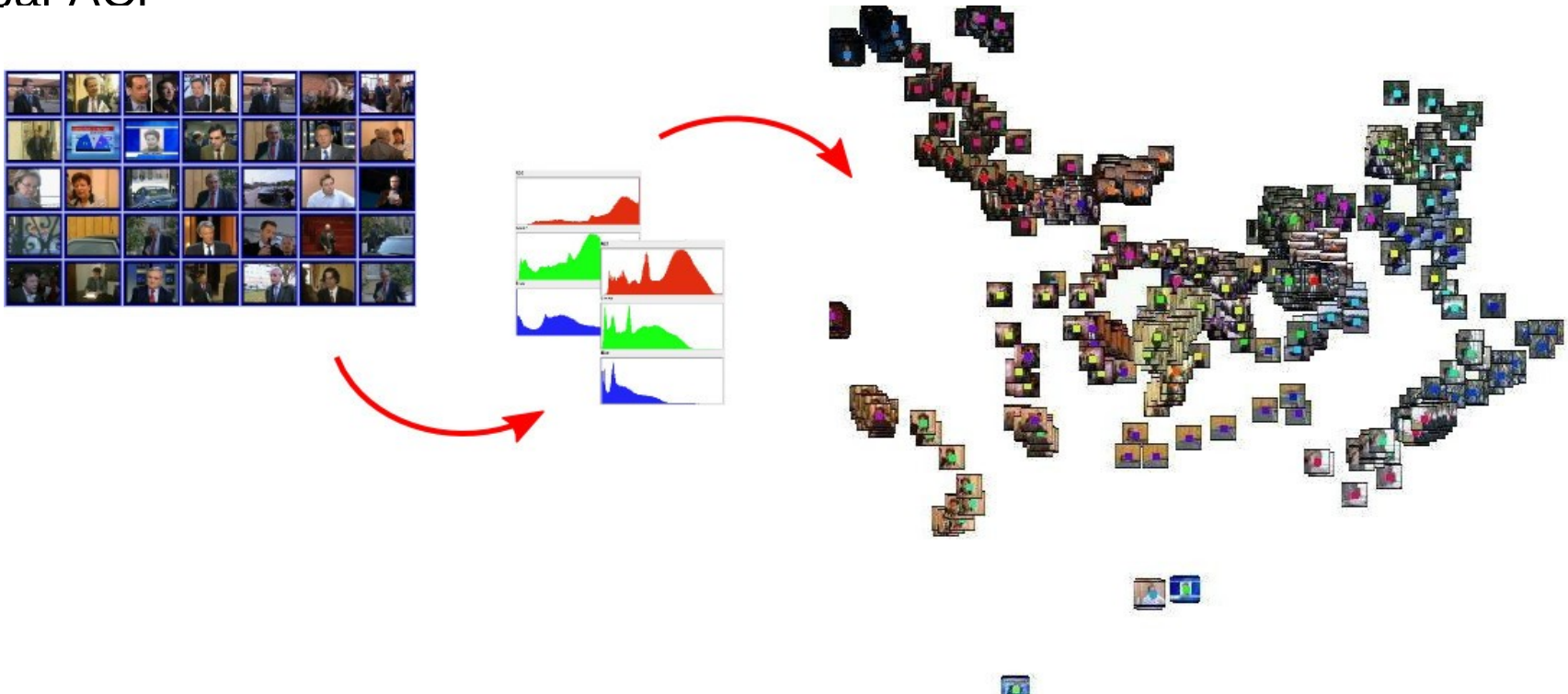
- Les composantes principales avec une large dynamique représentent le signal
- Celles avec une faible variance constituent le bruit

Sélection de caractéristiques

Réduction de dimension

Approche #1 : Analyse en Composantes Principales

Ex: images caractérisées par des histogrammes de couleurs et visualisées par ACP



Sélection de caractéristiques

Réduction de dimension

Approche #2 : t-distributed stochastic neighbor embedding (t-SNE)

Approche non-linéaire de réduction de dimension
[van der Maaten & Hinton, 2008]

Objectif : *trouver un mapping en faible dimension qui reflète au mieux les **similarités** entre les observations*

Sélection de caractéristiques

Réduction de dimension

Approche #2 : t-distributed stochastic neighbor embedding (t-SNE)

Soient N observations x_1, \dots, x_N en haute dimension :

Calcul des probabilités de similarité des obs. :

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}, \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

Définition des probabilités de similarité des projections y_1, \dots, y_N :

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq m} (1 + \|\mathbf{y}_k - \mathbf{y}_m\|^2)^{-1}}$$

Minimisation de la divergence de Kullback-Leibler de la distribution Q par rapport à P :

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Optimisation par descente de gradient (facilement dérivable)

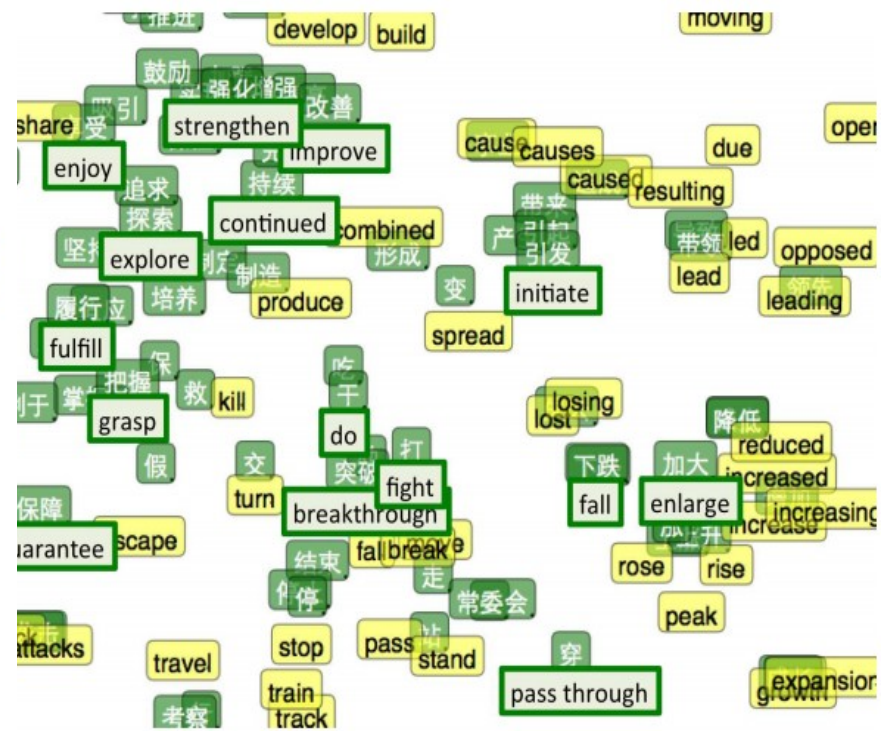
Sélection de caractéristiques

Réduction de dimension

Approche #2 : t-distributed stochastic neighbor embedding (t-SNE)



t-SNE sur base d'images



t-SNE sur mots (chinois +anglais)

Catégorisation

Synonymes: partitionnement, classification non-supervisée, clustering

Objectifs :

- trouver les *catégories* d'objets
- principe :
 - grande similarité intraclasse
(*qui se ressemble s'assemble*)
 - grande dissimilarité interclasse (catégories distinctes)

Catégorisation

Catégories ?

Ensembles d'objets proches (notion de distance → algorithme k-means, **approches basées centroïde**)

Ensembles d'objets générés par différentes distributions statistiques (ex : estimation des paramètres d'un mélange de Gaussiennes, **approches basées distribution**)

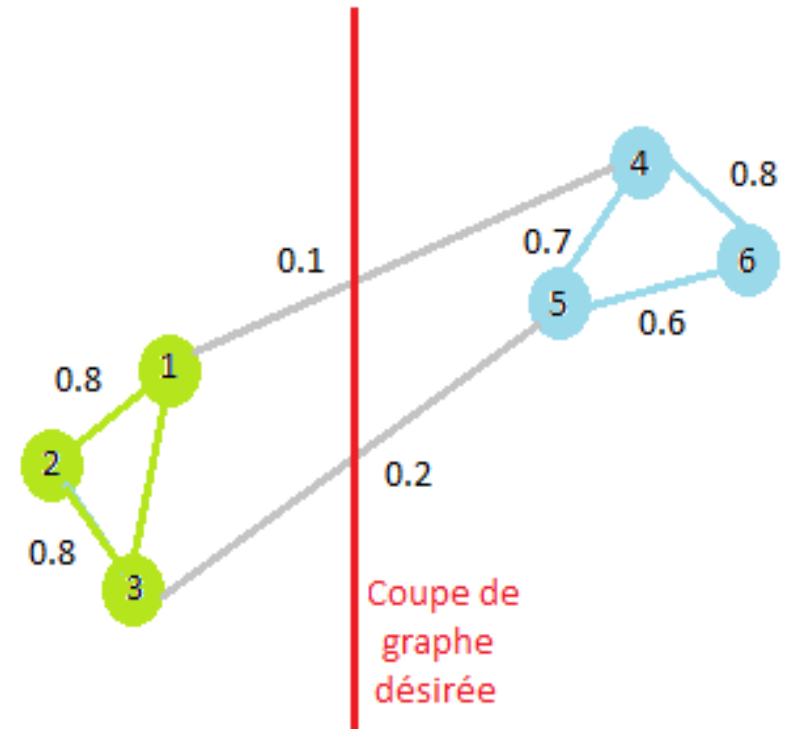
Ensembles d'objets similaires et nombreux (notion de densité → mean-shift, DBSCAN, **approches basées densité**)

Catégorisation

Partitionnement spectral :

- Calcul de la matrice de similarité
- Réduction de dimension par ACP (1ères dimensions conservées)
- K-means sur les projections

Peut s'appliquer aux objets complexes, non-vectoriels



Caractérisation des données : conclusion

*i.e.: Extraction de caractéristiques
Réduction de dimension
Catégorisation*

Objectifs communs :

- Projection dans un **espace de représentation** de plus faible dimension
- Concision et **parcimonie** de la représentation
- Représentation conservant la notion de **similarité** (parfois sémantique)