



From Data Science to Machine Learning

De l'analyse de données à l'apprentissage automatique

Jean-Yves Ramel

ramel@univ-tours.fr



Laboratoire d'Informatique Fondamentale
et Appliquée de Tours



Science / Analyse des données

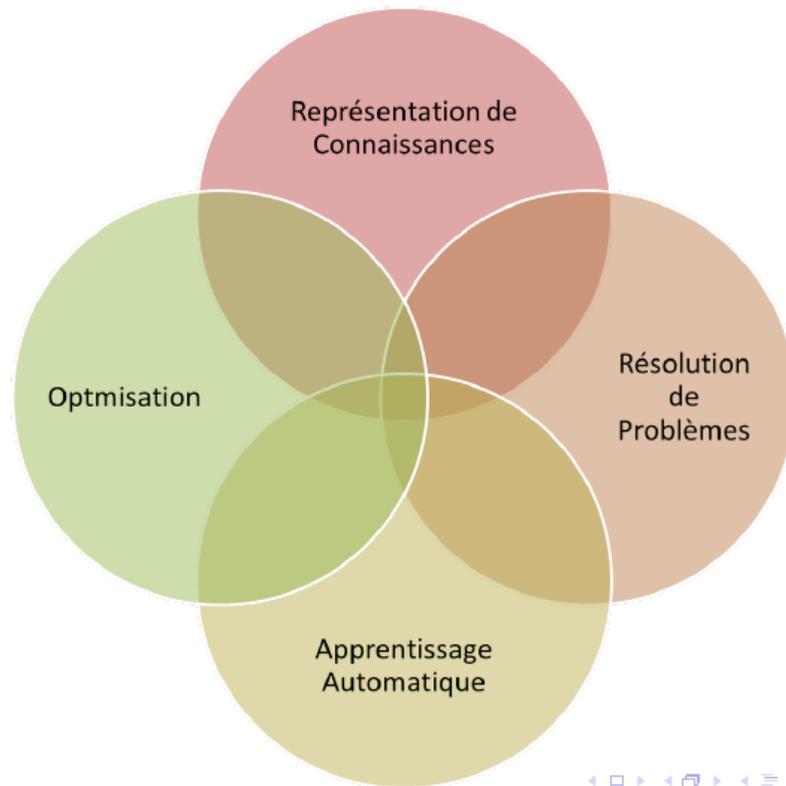
Data driven science : le 4e paradigme (Jim Gray - Prix Turing)

Extrait : "A l'heure actuelle, la science vit une révolution qui conduit à nouveau paradigme selon lequel 'la science est dans les données', autrement dit la connaissance émerge du traitement des données [...] **Le traitement de données et la gestion de connaissances représentent ainsi le quatrième pilier de la science après la théorie, l'expérimentation et la simulation.** L'extraction de connaissances à partir de grands volumes de données (en particulier quand le nombre de données est bien plus grand que la taille de l'échantillon) , l'apprentissage statistique, l'agrégation de données hétérogènes, la visualisation et la navigation dans de grands espaces de données et de connaissances sont autant d'instruments qui permettent d'observer des phénomènes, de valider des hypothèses, d'élaborer de nouveaux modèles ou de prendre des décisions en situation critique"

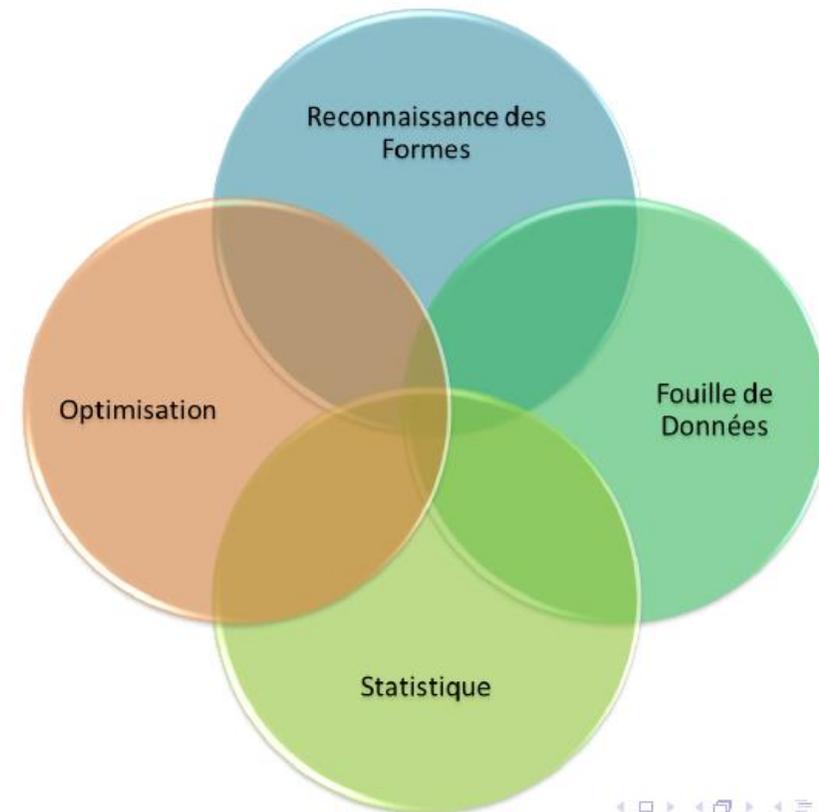
Science / Analyse des données

- L'apprentissage automatique fait largement appel à des outils et des concepts de la statistique, et fait partir de disciplines plus vastes appelées science des données et IA.

Intelligence Artificielle



Apprentissage Automatique



Apprentissage automatique

Définition initiale

- Learning is making useful changes in mind - [Marvin Minsky, 1985]
- Learning is any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population - [Herbert Simon, 1983]
- Learning is the organization of experience - [Scott, 1983]
- Learning is constructing or modifying representations of what is being experienced - [Ryszard Michalski, 1986]

Problématique

- Nous souhaitons avoir des ordinateurs / **agents intelligents**, adaptatifs, avec un comportement robuste
- Programmer de tels comportement est souvent impossible

Solution

- Faire un ordinateur capable de se programmer lui-meme
 - a partir d'exemples (apprentissage classique / par imitation)
 - a partir de son "expérience" (apprentissage par renforcement)

Apprentissage automatique

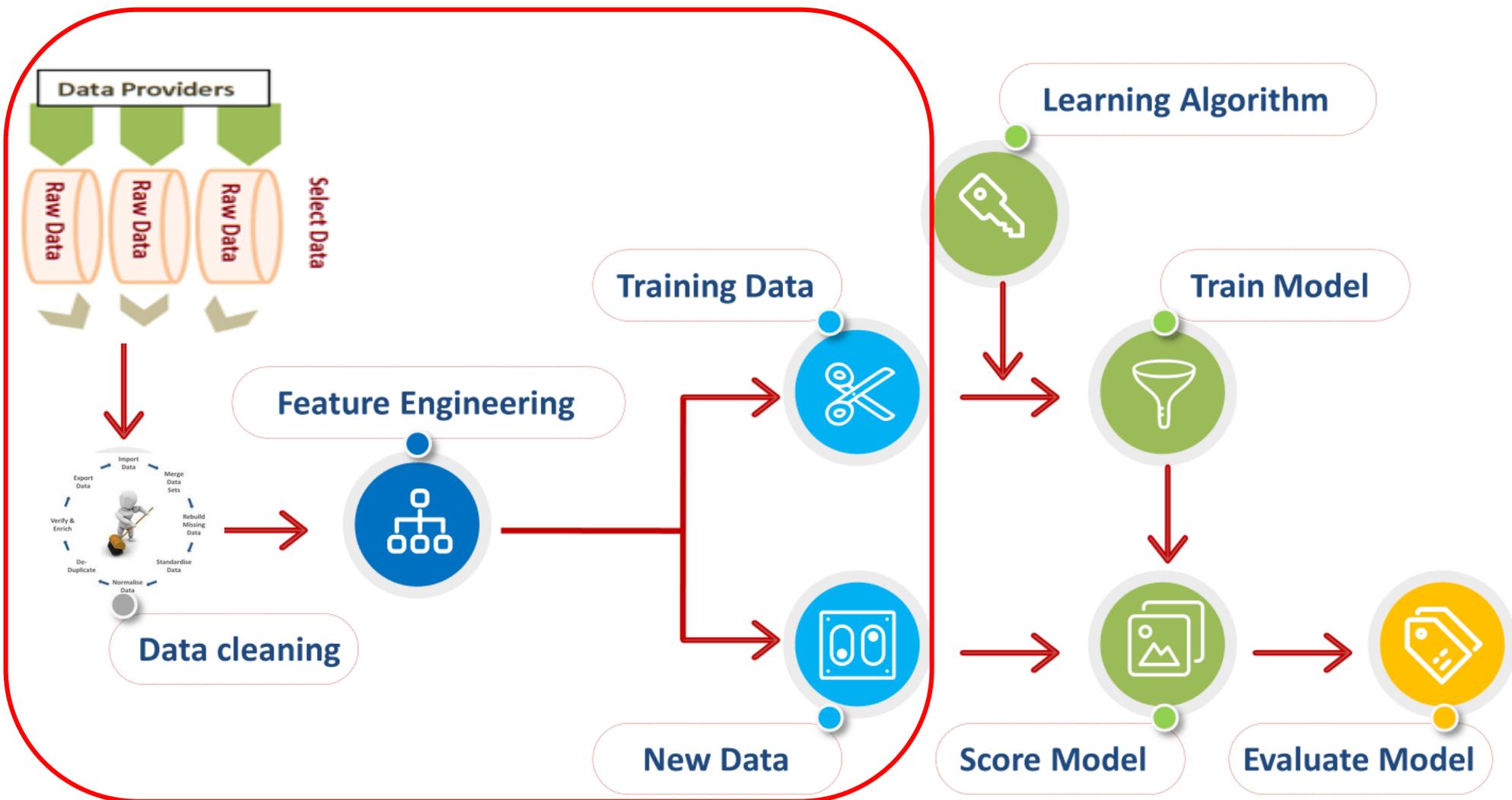
Autre définition

- Branche de l'IA qui concerne le développement d'algorithmes permettant de rendre une machine (un agent) capable d'accomplir des tâches complexes sans avoir été explicitement programmée dans ce but.

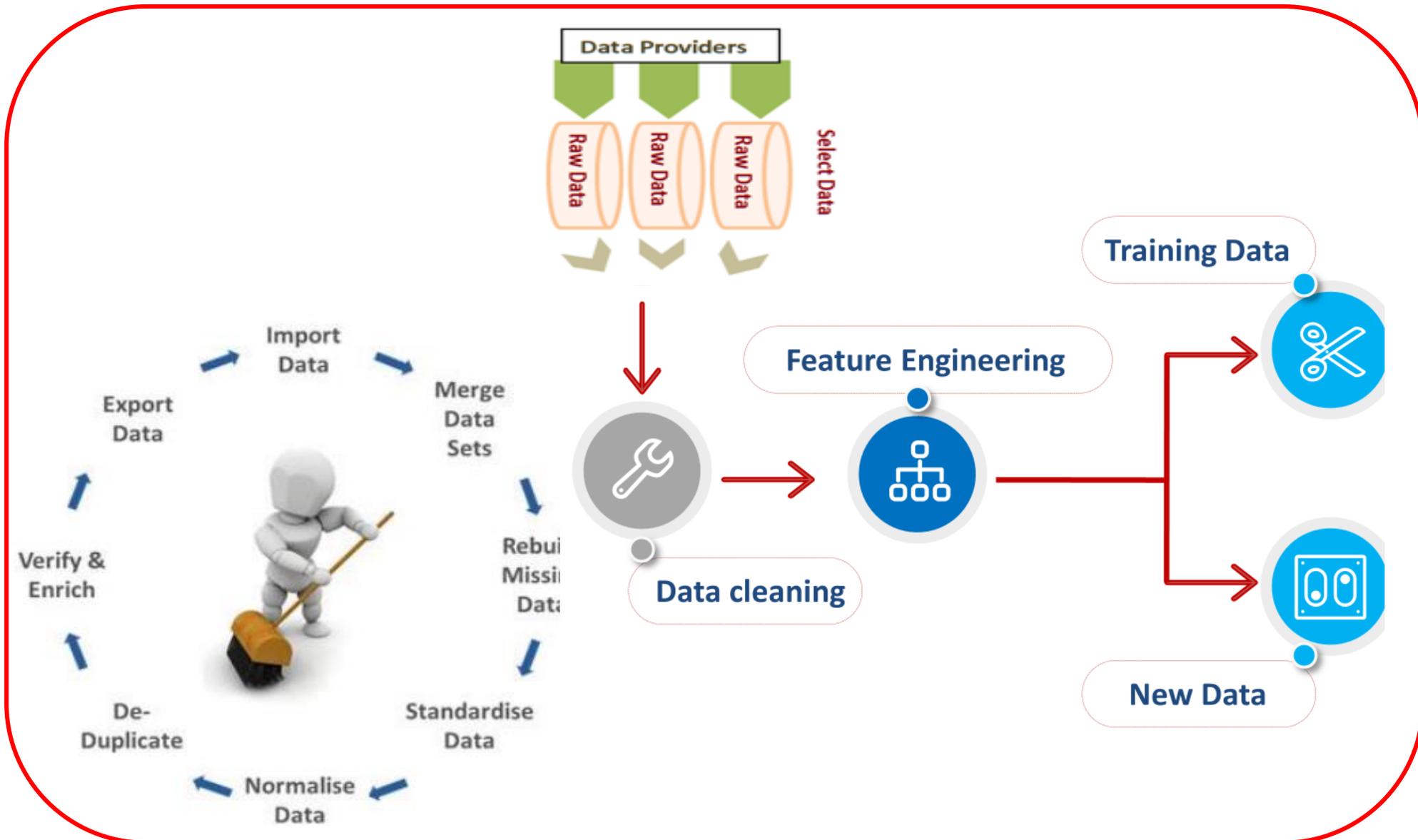


- Exemple : comment écrire un programme qui reconnaisse les caractères manuscrits ?
 - Entrer des règles manuellement (difficile et peu fiable)
 - Meilleure méthode : écrire un algorithme (générique) qui produit automatiquement un programme de reconnaissance de caractères à partir d'un grand nombre d'exemples.

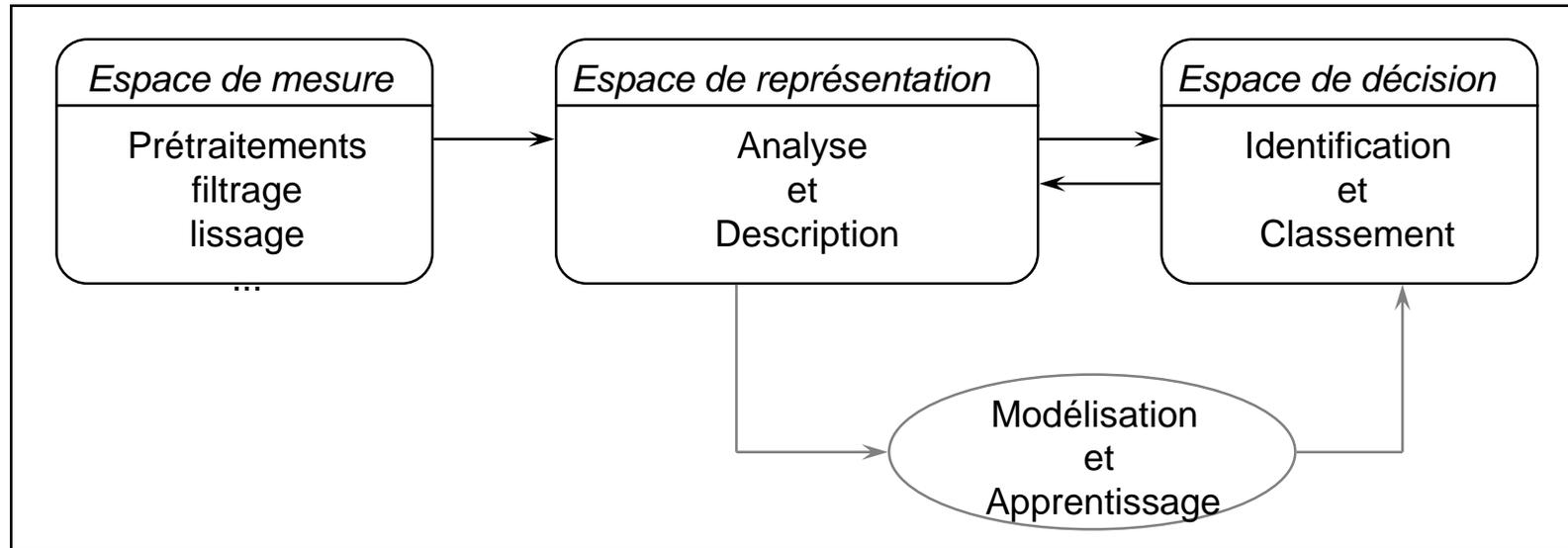
Modélisation possible d'un système d'IA



Zoom sur les données et leur préparation



Modélisation orientée « données »



Espace de mesure

- Représentation du monde réel
- Obtention des données à l'aide d'une méthode de perception : **le capteur**
- Un certain nombre de (pré)traitements peut être effectué dans l'espace de mesure (au plus proche du capteur – **Edge intelligence** [Plastiras2018])

Espace de représentation

- Espace dans lequel seront fait les analyses pour produire les décisions
- Par transformation : espace de représentation → espace de décision

Espace de décision

- doit aussi être réfléchi et **modéliser**

Zoom sur les données...

Les données peuvent être vues comme une collection d'objets décrits par des attributs

- Un attribut est une propriété, une caractéristique de l'objet
- Un ensemble d'attributs décrit un objet

Attribut - valeur

- La valeur d'un attribut est un nombre ou un symbole
- Ne pas confondre attribut et valeur

Types des valeurs

- Quantitative (numérique, exprime une quantité)
 - Discrète (ex : nb d'étudiants) ou continue (longueur, ...)
 - Echelle proportionnelle (chiffre d'affaires, taille), ou échelle d'intervalle (QI)
- Qualitative (ou symbolique)
 - Variable ordinale (classement à un concours, échelle de satisfaction client, ...)
 - Variable nominale (couleur de yeux, diplôme obtenu, sexe, ...)

Les modalités d'une variable sont l'ensemble des valeurs qu'elle peut prendre

Ex : les modalités de notes sont 0; 1; 2; ... ; 20 - les modalités de couleur sont bleu, vert, noir,...

Attributes

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Objects

Préparer les données

Etape 1 : Acquisition, fusion de données brutes

- Obtenir des données en accord avec les objectifs visés
 - **Obtenir un maximum de données** : Masse de données non étiquetées, non structurées (big data, IoT, ...)
 - **Représentativité** des données utilisées pour la conception du système
 - Nécessité (éventuelle) d'une expertise du domaine pour **labéliser les données**
 → produire le résultat que l'on désire faire apprendre à la machine (supervisé)
- Première tentative de structuration des données en champs typés

Client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bémol	Rue du moulin, Paris	7/10/2006	Voiture
23134	Bémol	Rue du moulin, Paris	12/5/2006	Musique
23134	Bémol	Rue du moulin, Paris	25/7/2005	BD
31435	Bodinoz	Rue Verte, Nancy	11/11/1111	BD
43342	Airinair	Rue de la source, Brest	30/05/2005	Sport
25312	Talonion	Rue du Marché, Paris	25/02/2007	NULL
43241	Manvussa	NULL	14/4/2006	Sport
23130	Bémolle	Rue du moulin, Paris	11/11/1111	Maison

Préparer les données

Etape 2 : Nettoyer, corriger les données

- Corrections des doublons, des erreurs de saisie
- Contrôle sur l'intégrité des domaines de valeurs : détection des valeurs aberrantes
- Détection des outliers
- Détection des informations manquantes
 - Cas où les champs ne contiennent aucune donnée
 - Parfois intéressant de conserver ces enregistrements car l'absence d'information peut être informative (e.g. fraude).

Client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bémol	Rue du moulin, Paris	7/10/2006	Voiture
23134	Bémol	Rue du moulin, Paris	12/5/2006	Musique
23134	Bémol	Rue du moulin, Paris	25/7/2005	BD

Client	Nom	Adresse	Date d'abonnement	Magazine
31435	Bodinoz	Rue Verte, Nancy		BD
43342	Airinair	Rue de la source, Brest	30/05/2005	Sport
25312	Talonion	Rue du Marché, Paris	25/02/2007	NULL
43241	Manvussa	NULL	14/4/2006	Sport
23130	Bémol	Rue du moulin, Paris		Maison

Client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bémol	Rue du moulin, Paris	7/10/2006	Voiture
23134	Bémol	Rue du moulin, Paris	12/5/2006	Musique
23134	Bémol	Rue du moulin, Paris	25/7/2005	BD
31435	Bodinoz	Rue Verte, Nancy	NULL	BD
43342	Airinair	Rue de la source, Brest	30/05/2005	Sport
25312	Talonion	Rue du Marché, Paris	25/02/2007	NULL
43241	Manvussa	NULL	14/4/2006	Sport
23130	Bémol	Rue du moulin, Paris	NULL	Maison

Préparer les données

Etape 3 : Enrichissement (facultative)

- Recours à d'autres bases de données souvent pour ajouter de nouveaux champs en conservant le même nombre d'enregistrements
 - Plusieurs difficultés : Relier les données, parfois hétérogènes, entre elles, introduction de nouvelles valeurs manquantes et/ou aberrantes, ...
- Génération de données synthétiques :
 - modèles de déformations appliquées aux données obtenues
 - Mélange d'enregistrements
 - Generative Adversarial Networks [Goodfellow2014]

Client	Date naissance	Revenus	Propriétaire	Voiture
Bémol	13/1/50	20 000	Oui	Oui
Bodinoz	21/5/70	12 000	Non	Oui
Airinair	15/06/63	9 000	Non	Non
Manvussa	27/03/47	15 000	Non	Oui

Encodage et normalisation des données

Etape 4 : Encodage et normalisation

- Etape très dépendante du choix de l'algorithme de fouilles de données utilisé
- Regroupements
 - Changements de types pour permettre certaines manipulations comme par exemple des calculs de distance, de moyenne (e.g. date de naissance)
 - Attributs discrets : Les attributs discrets prennent leurs valeurs (souvent textuelles) dans un ensemble fini donné (e.g. colonne magazine de l'exemple).
 - Deux représentations possibles : représentation verticale ou représentation horizontale ou éclatée (plus adaptée à la fouille de données)
 - A l'inverse : cas où les attributs prennent un très grand nombre de valeurs discrètes (e.g. adresses que l'on peut regrouper en 2 régions (Paris - Province))

- Uniformisation (cf plus loin dans le cours)

- Certains algorithmes sont basés sur des calculs de distance entre enregistrements. Les variations d'échelle selon les attributs peuvent perturber ces algorithmes

Client	Magazine
23134	Voiture
23134	Musique
23134	BD
31435	BD
43342	Sport
43241	Sport
23134	Maison

Client	Sport	BD	Voiture	Maison	Musique
23134	0	1	1	1	1
31435	0	1	0	0	0
43342	1	0	0	1	0
43241	1	0	0	1	0

Des données aux bases d'apprentissage

Individus, objets VS variables, descripteurs

Acquisition des données → Préparation → Espace de représentation

Dans l'espace de représentation (vectoriel), on parle de :

- **Population** : groupe ou ensemble d'individus que l'on analyse
- **Recensement** : étude de tous les individus d'une population donnée
- **Sondage** : étude d'une partie seulement d'une population appelée échantillon
- **Individus** : les enregistrements (ensemble d'attributs) deviennent des individus ou objets
- **Variables, descripteurs, caractéristiques** : les attributs, comme déjà indiqué, possiblement de différents types :
 - **Quantitatives** : nombres sur lesquels les opérations usuelles (somme, moyenne,...) ont un sens, peuvent être discrètes
 - Qualitatives : appartenance a une catégorie donnée, peuvent être nominales ou ordinales



Des données aux bases d'apprentissage

Description de données quantitatives

- **Définition**

On appelle variable un vecteur x de taille p
Chaque coordonnée x_i correspond à un individu
On s'intéresse ici à des valeurs numériques

- **Poids**

Chaque individu a éventuellement un poids p_i , tel que $p_1 + \dots + p_n = 1$
On a souvent $p = 1/n$



- **Indicateurs statistiques**

- on dispose d'une série d'indicateurs qui ne donne qu'une vue partielle des données : effectif, moyenne, médiane, variance, écart type, minimum, maximum, étendue, 1er quartile, 3eme quartile, ...
- Ces indicateurs mesurent principalement la tendance centrale et la dispersion

- **Représentations graphiques**

histogramme en découpant les valeurs de la variable en classes

- **Les méthodes d'analyse de données cherchent à proposer des alternatives plus performantes de fouille et visualisation des données**

Tableaux, vecteurs, nuages de points

Notation matricielle

- **Matrice**
tableau de données carre ou rectangulaire.

- **Vecteur**
matrice a une seule colonne.

- **Cas particuliers**

$$I = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} \quad 1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

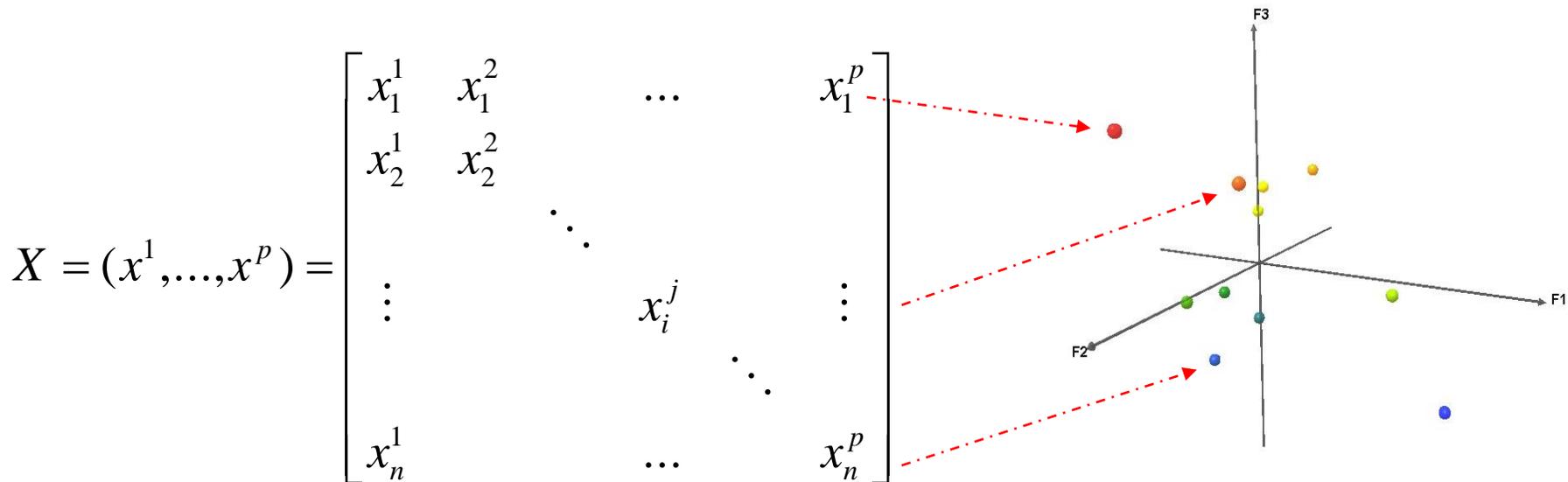
- **Transposition de matrice**
échange des lignes et des colonnes d'une matrice ; on note M' la transposée de M .

Tableaux, vecteurs, nuages de points

Tableau de données



- Pour n individus et p variables, on a le tableau X
- X est une matrice rectangulaire a n lignes et p colonnes
- X est aussi un nuage de n points dans \mathbb{R}^p



Tableaux, vecteurs, nuages de points

Vecteur variable et individu

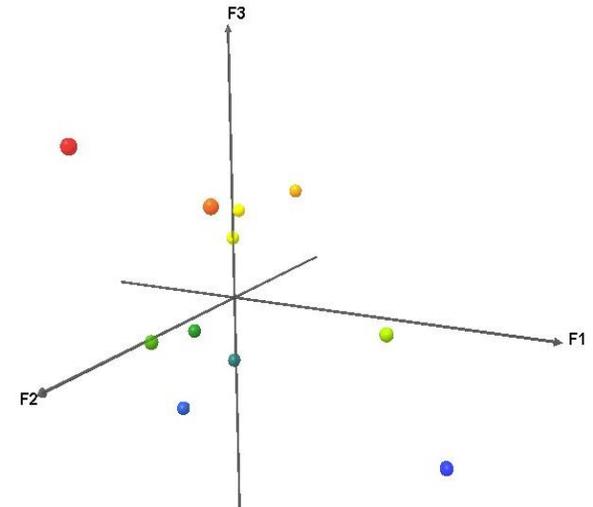
- **Variable**

Une colonne du tableau

$$x^j = \begin{bmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{bmatrix}$$

- **Individu**

Une ligne du tableau



$$e_i' = (x_i^1 \quad x_i^2 \quad \dots \quad x_i^p)$$

Pondération, corrélation et Inertie

La matrice des poids

- **Pourquoi**
utile quand les individus n'ont pas la même importance
- **Comment**
on associe aux individus un poids p_i tel que
$$p_1 + p_2 + \dots + p_n = 1$$
et on représente ces poids dans la matrice diagonale de taille n

$$D = \begin{bmatrix} p_1 & & \dots & 0 \\ & p_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & p_n \end{bmatrix}$$

- **Cas uniforme**
tous les individus ont le même poids $p_i = 1 / n$ et $D = I / n$

Tableaux, vecteurs, nuages de points

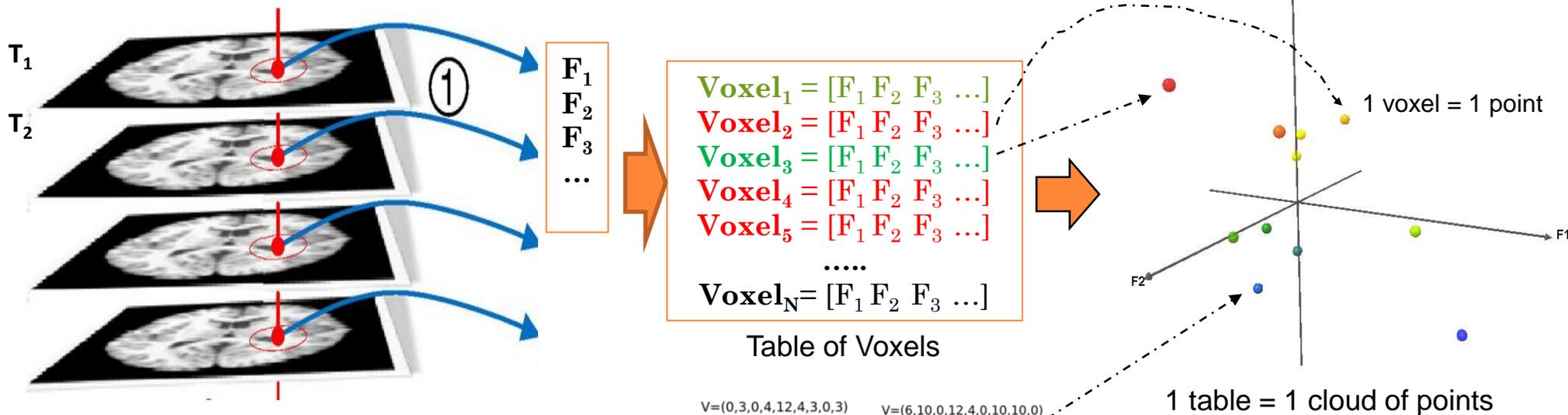
Principe : 1 objet \Rightarrow p caractéristiques \Rightarrow 1 Vecteur = 1 Point $\in \mathbb{R}^p$

- 1 objet \Rightarrow p features \Rightarrow 1 Vector = 1 Point $\in \mathbb{R}^p$
- 1 tableau = n objets (individus)
- Statistiques, analyse de données, Machine Learning deviennent opérationnels

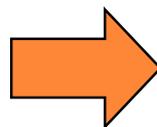


Exemple 1 : Images médicales, pixels/voxels

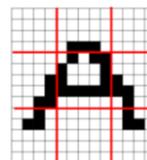
1 pixel / Voxel with several modalities



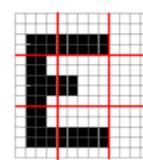
Exemple 2 : OCR



$V=(0,3,0,4,12,4,3,0,3)$



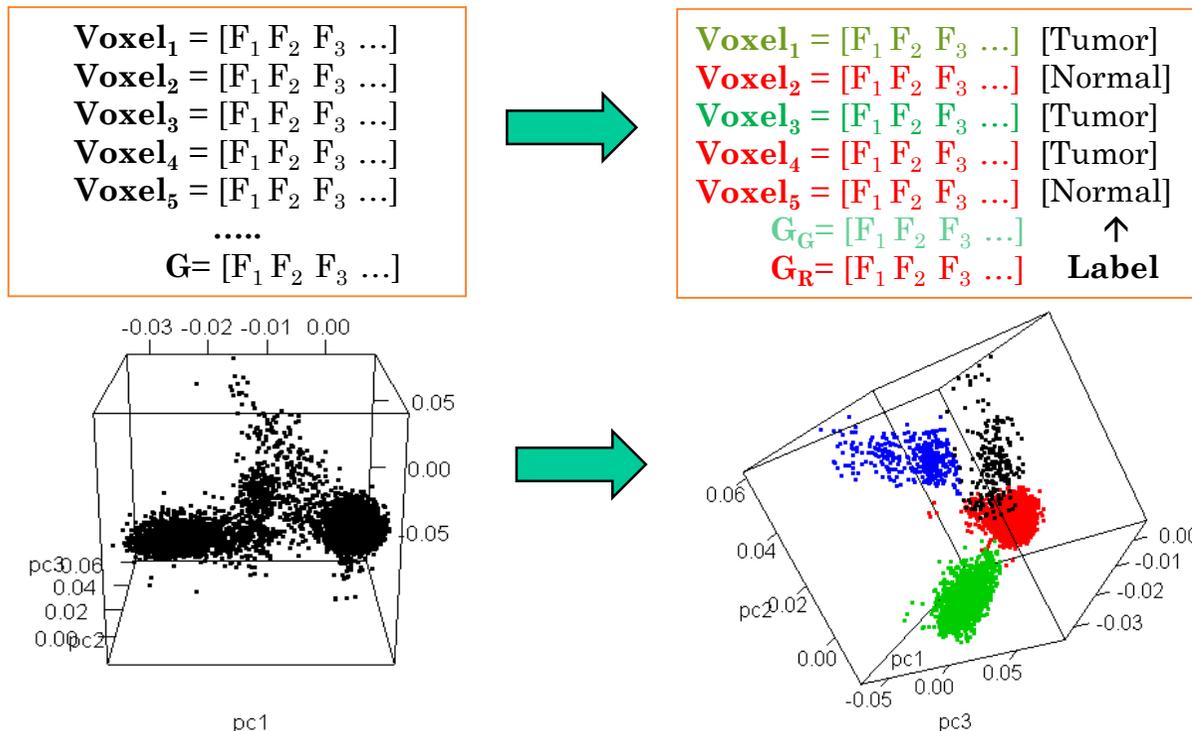
$V=(6,10,0,12,4,0,10,10,0)$



Bases d'apprentissage

Principe : 1 objet \Rightarrow p caractéristiques \Rightarrow 1 Vecteur = 1 Point $\in \mathbb{R}^p$

- Chaque individu/objet est décrit par des descripteurs numériques auxquels peuvent s'ajouter des **labels additionnels**
- 1 tableau = n objets (individus)
- 1 objet \Rightarrow p features \Rightarrow 1 Vecteur = 1 Point $\in \mathbb{R}^p$ **+ additional labels**
- **Apprentissage non-supervisé \rightarrow apprentissage supervisé**



Bases d'apprentissage

Les données (labélisées) deviennent le Graal → Big data / data science

- 1 object \Leftrightarrow p features \Leftrightarrow 1 Vector = 1 Point $\in \mathbb{R}^p$
- **Dataset** = Table of Vectors with or without a label (class)
- **La selection des caractéristiques** devient le problème crucial (expertise, ...)
 - Feature engineering → **Choix de l'espace de représentation**
 - Méthodes de sélection de caractéristiques
 - Analyse statistique : analyse des corrélations et des inerties
 - Analyse de données : ACP, AFD, ...
 - Analyses itératives [Kohavi1997] : SFS, F2FS, ...

	Pointure	Poids	Sexe	
Pierre	44	70	Homme	
Paul	42	70	Homme	
Marie	38	70	Femme	
Type de variable				

Pondération, corrélation et Inertie

Rappel Moyenne arithmétique

- **Définition**

On note

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

ou pour des données pondérées

$$\bar{x} = \sum_{i=1}^n p_i x_i$$

- **Propriétés**

la moyenne arithmétique est une mesure de tendance centrale qui dépend de toutes les observations et est sensible aux valeurs extrêmes. Elle est très utilisée à cause de ses bonnes propriétés mathématiques.

Pondération, corrélation et Inertie

Rappel Variance et écart-type

- **Définition**

la variance de x est définie par

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{ou} \quad s_x^2 = \sum_{i=1}^n p_i (x_i - \bar{x})^2$$

L'écart type s_x est la racine carrée de la variance.

- **Propriétés**

La variance satisfait la formule suivante

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n p_i x_i^2 - (\bar{x})^2$$

La variance est « la moyenne des carrés moins le carré de la moyenne ». L'écart-type, qui a la même unité que x , est une mesure de dispersion.

Pondération, corrélation et Inertie

Rappel Mesure de liaison entre deux variables

- Définitions la covariance observée entre deux variables x et y est

$$s_{xy} = \sum_{i=1}^n p_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n p_i x_i y_i - \bar{xy}$$

et le coefficient de corrélation est donnée par

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n p_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n p_i (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n p_i (y_i - \bar{y})^2}}$$

Pondération, corrélation et Inertie

Propriétés du coefficient de corrélation

- **Borne**

On a toujours (inégalité de Cauchy Schwarz)

$$-1 \leq r_{xy} \leq 1$$

- **Variables liées**

$$|r_{xy}| = 1 \Leftrightarrow (ax_i + by_i) = c \quad \forall 1 \leq i \leq n$$

$|r_{xy}| = 1$ si et seulement si x et y sont linéairement liées

En particulier, $r_{xx} = 1$.

- **Variables décorrélées**

si $r_{xy} = 0$, on dit que les variables sont décorrélées.

Cela ne veut pas dire qu'elles sont indépendantes !

Exercice – Statistiques de base

	Pointure	Poids	Sexe	
Pierre	44	70	Homme	
Paul	42	70	Homme	
Marie	38	70	Femme	
Type de variable				
$E(V_j)$				
Variance(V_j)				

Pondération, corrélation et Inertie

Point moyen et tableau centré

Point moyen (centre de gravité)

c'est le vecteur g des moyennes arithmétiques de chaque variable :

$$g' = (\bar{x}^1 \quad \dots \quad \bar{x}^p)$$

ou

$$\bar{x}^j = \sum_{i=1}^n p_i x_i^j$$

On peut aussi écrire $g = X' D 1$

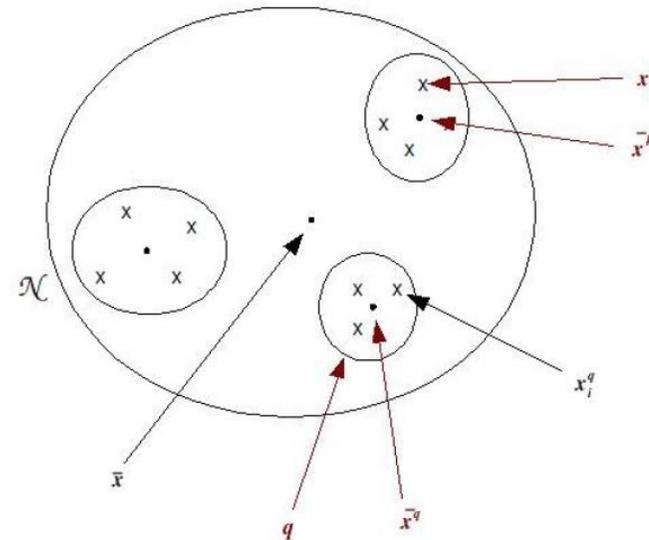
Tableau centré

il est obtenu en centrant les variables autour de leur moyenne

$$y_i^j = x_i^j - \bar{x}^j$$

ou, en notation matricielle,

$$Y = X - 1g' = (I - 11'D)X$$



Pondération, corrélation et Inertie

Matrice de variance covariance

- **Définition**
c'est une matrice carrée de dimension p

$$V = \begin{bmatrix} s_1^1 & s_1^2 & \dots & s_1^p \\ s_2^1 & s_2^2 & & \\ \vdots & & \ddots & \vdots \\ s_p^1 & & \dots & s_p^p \end{bmatrix}$$

ou s_{kl} est la covariance des variables x^k et x^l et s_j^2 est la variance de la variable x^j

- **Formule matricielle**

$$V = X'DX - gg' = Y'DY \quad Y \text{ tableau } X \text{ centré}$$

Pondération, corrélation et Inertie

Matrice de corrélation

- Définition**

Si l'on note

$$r_{kl} = \frac{s_{kl}}{s_k s_l}$$

$$R = \begin{bmatrix} 1 & r_1^2 & \dots & r_1^p \\ r_2^1 & 1 & & \\ \vdots & & \ddots & \vdots \\ s_p^1 & \dots & & 1 \end{bmatrix}$$

- Formule matricielle**

$$R = D_{1/s} V D_{1/s}$$

$$D_{1/s} = \begin{bmatrix} \frac{1}{s_1} & & & 0 \\ & \frac{1}{s_2} & & \vdots \\ & & \ddots & \\ 0 & & & \frac{1}{s_p} \end{bmatrix}$$

Pondération, corrélation et Inertie

Inertie d'un nuage de points



- **Définition**

l'inertie en un point a du nuage de points est

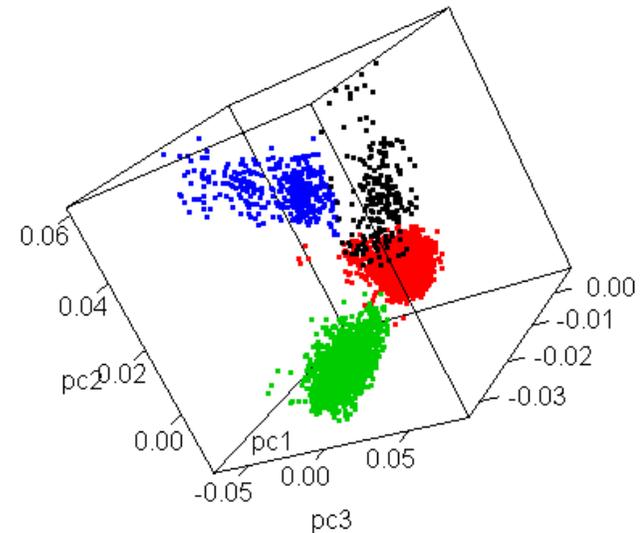
$$I_a = \sum_{i=1}^n p_i \|e_i - a\|_M^2 = \sum_{i=1}^n p_i (e_i - a)' M (e_i - a)$$

- L'inertie totale est aussi donnée par la trace de la matrice V (somme de ses éléments diagonaux).

$$D_k^2 = \frac{1}{n_k(n_k - 1)} \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} d_k^2(x_i, x_j)$$

- **Autres relations**

coefficient d'agrégation est la moyenne des carrés des distances entre les individus



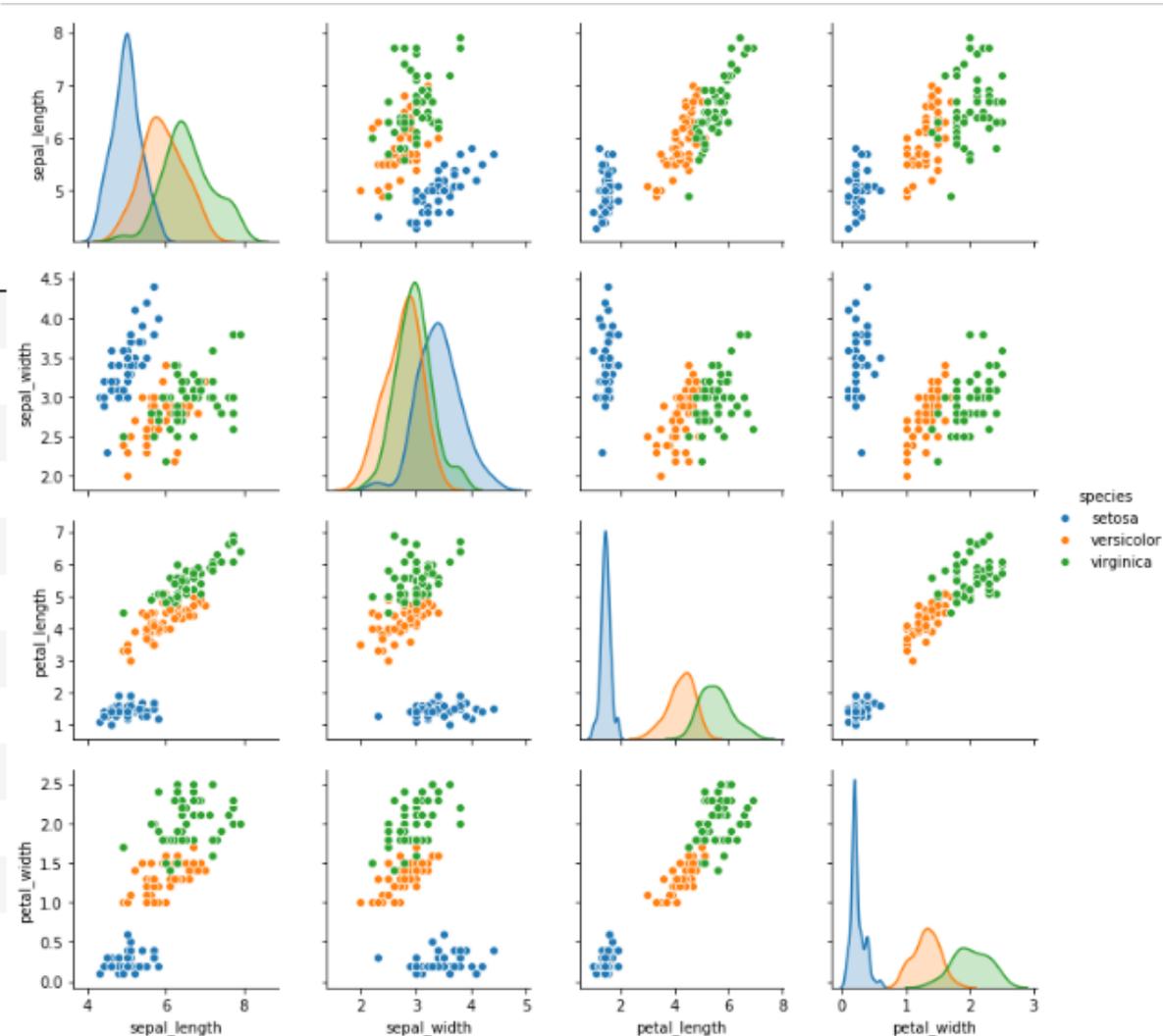
L'inertie est une mesure fondamentale pour évaluer la qualité d'un espace de représentation (inertie totale, inter-classe, intra-classe)

« Feature engineering », une étape clé

Bien choisir les descripteurs est crucial (expertise, ...)

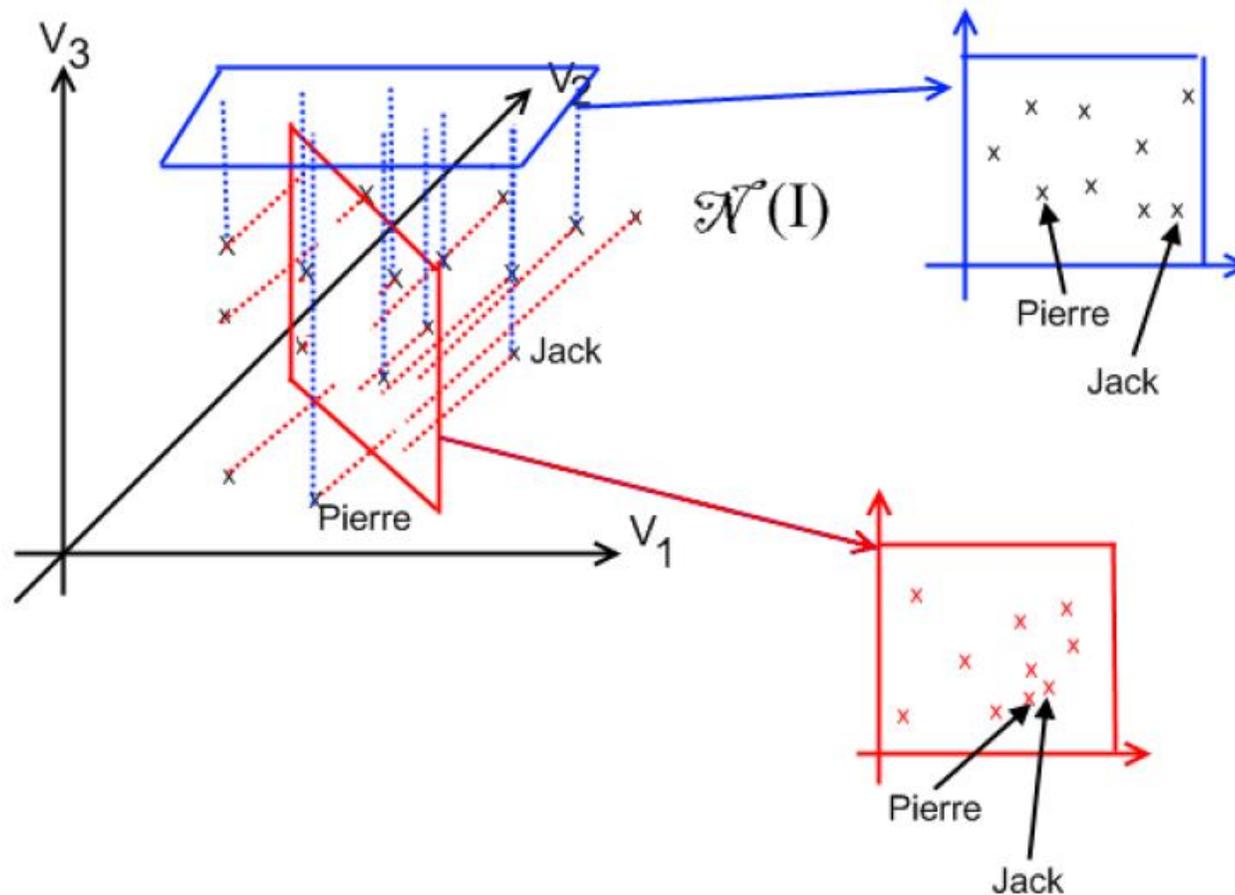
- Stables
- Discriminantes
- ...

	sepal_length	sepal_width	petal_length	petal_width	species
count	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	setosa
freq	NaN	NaN	NaN	NaN	50
mean	5.843333	3.057333	3.758000	1.199333	NaN
std	0.828066	0.435866	1.765298	0.762238	NaN
min	4.300000	2.000000	1.000000	0.100000	NaN
25%	5.100000	2.800000	1.600000	0.300000	NaN
50%	5.800000	3.000000	4.350000	1.300000	NaN
75%	6.400000	3.300000	5.100000	1.800000	NaN
max	7.900000	4.400000	6.900000	2.500000	NaN



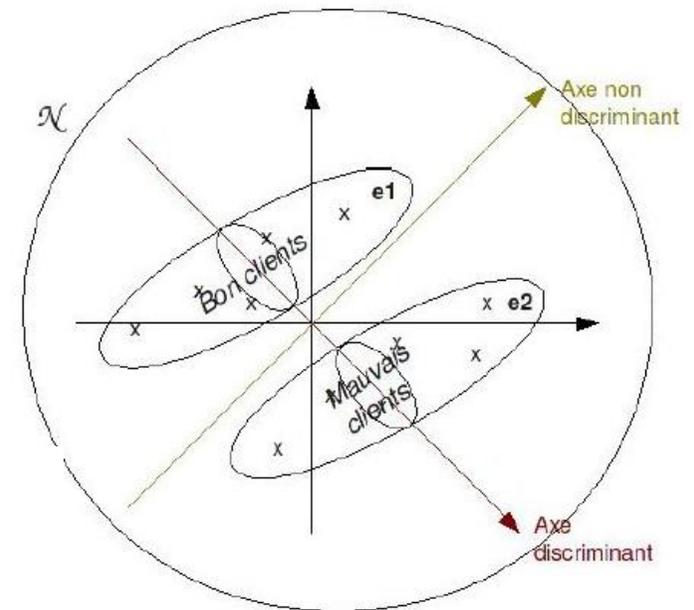
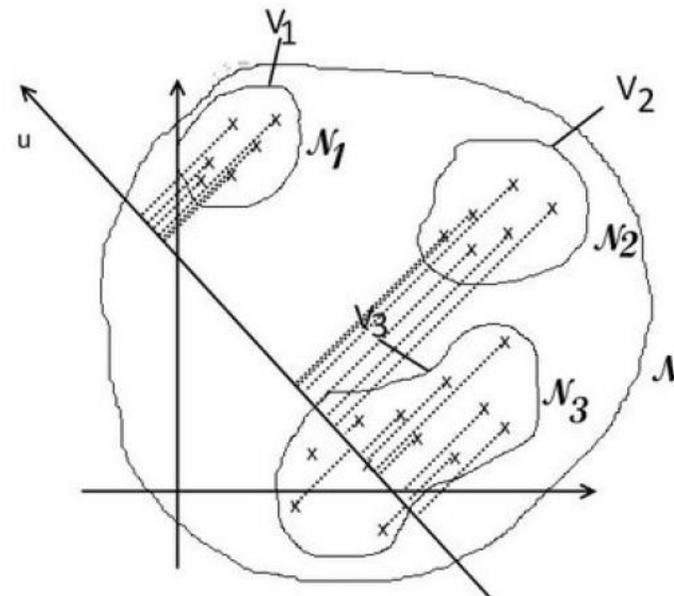
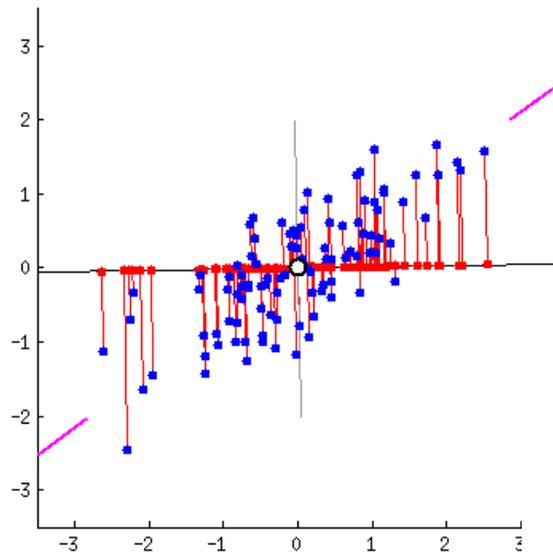
« Feature engineering », une étape clé

- Réduction de dimensionnalité \rightarrow ACP [Hotelling1933]
- Espace de représentation et projections



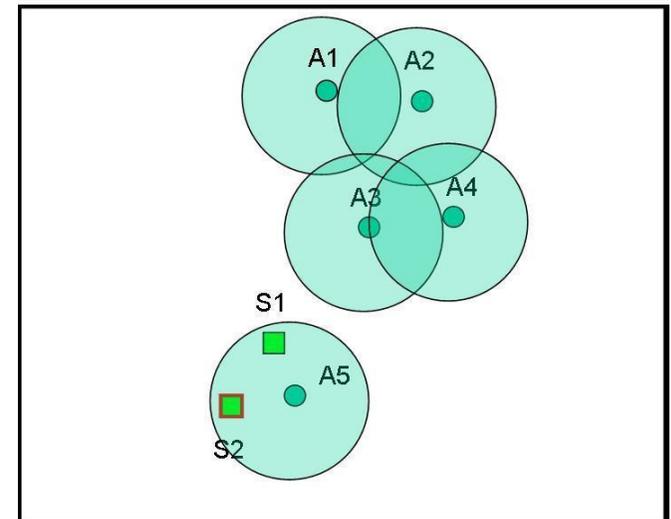
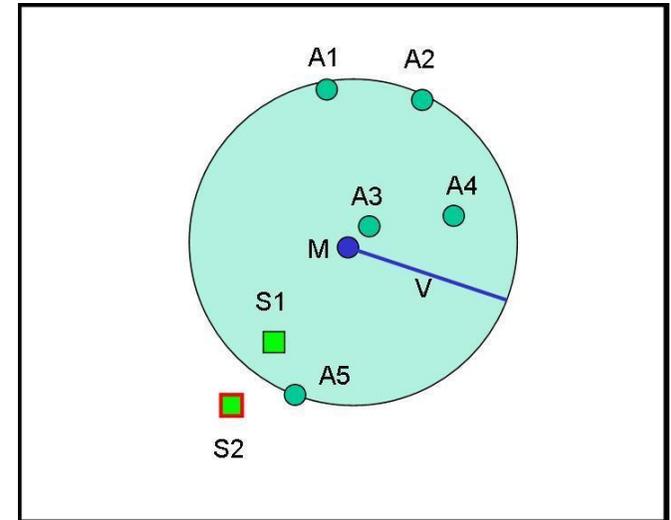
« Feature engineering », une étape clé

- Génération de variables discriminantes (AFD [Saporta2006])



Sélection des exemples d'apprentissage et choix (ou génération) des modèles, une étape clé

- Un modèle par classe
 - Un seul modèle par classe fonction de la moyenne et de l'écart-type de chacune des caractéristiques des exemples de l'ensemble d'apprentissage
→ Point moyen : G
- Plusieurs modèles par classe
 - Conservation de toute l'information, c'est-à-dire le vecteur de caractéristiques de chacun des exemples de l'ensemble d'apprentissage
 - Couverture, discrétisation de l'espace de représentation
 - Prendre en compte la variabilité
 - → KPPV

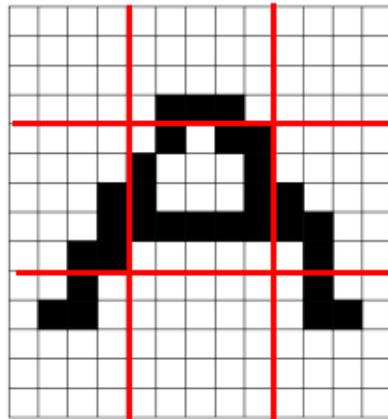


Ceci est une vision simpliste du problème pour illustrer l'importance du choix des exemples d'apprentissage et du modèle choisi ensuite

Exercice – Choix de caractéristiques ?

All the pixels become features or features by zones

- The image is splitted in n blocks
- For each block, some features are computed (number of black pixels)
- A new feature vector is obtained : $V = (Nb_1, Nb_2, \dots, Nb_n)$

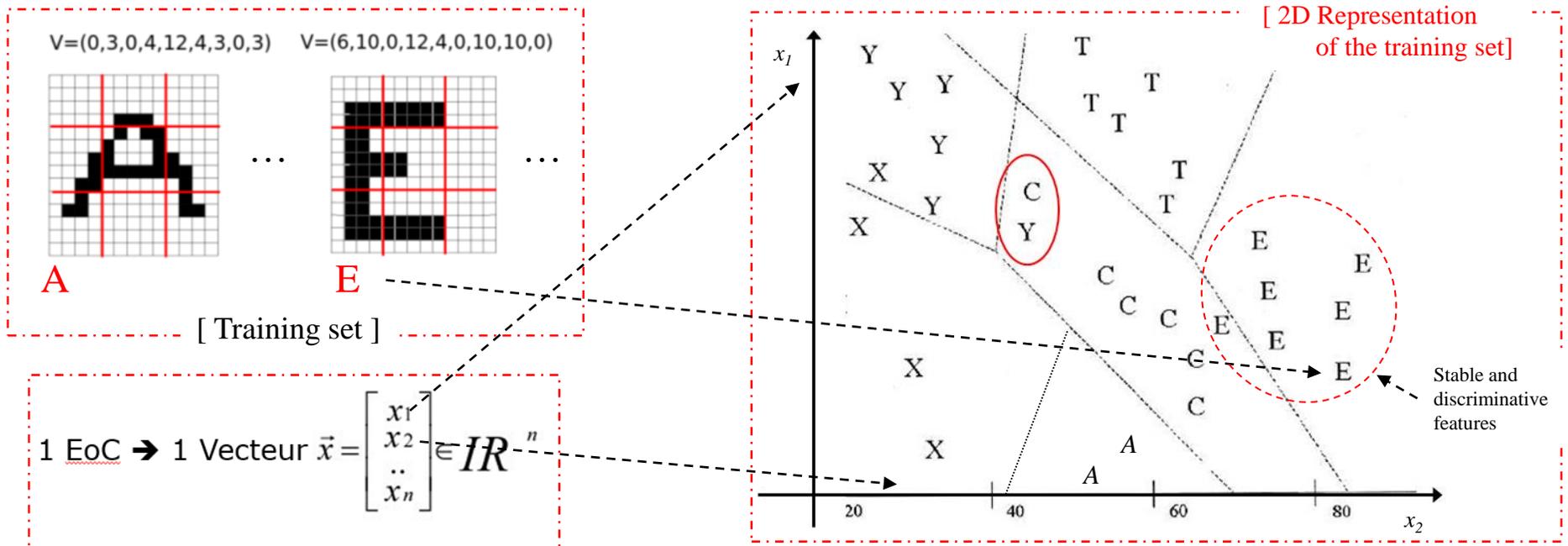


$$V = (0, 3, 0, 4, 12, 4, 3, 0, 3)$$

Exercice – Choix de caractéristiques ?

Representation space visualization and analysis

- We need a **large set of (labelled) examples** similar to the patterns to be recognized → **a training set**
- We need a list of **stable and discriminative features** (shape, color, size,...) used to describe the patterns (labelled ones and unknown one)
- What is the feature space here? Is it a good one?



Similarités, dissimilarités, distances

Distance entre individus

- **Motivation**

afin de pouvoir considérer la structure du nuage des individus, il faut définir une distance, qui induira une géométrie.

- **Distance euclidienne classique**

la distance la plus simple entre deux points de \mathbb{R}^p est définie par

$$d^2(u, v) = \sum_{j=0}^p (u_j - v_j)^2 = \|u - v\|^2$$

- **Généralisation simple**

on multiplie la variable j par $\sqrt{a_j}$

$$d^2(u, v) = \sum_{j=0}^p a_j (u_j - v_j)^2$$

$\alpha_j =$ Poids associé aux variables

Similarités, dissimilarités, distances

Métrie

- **Matrice définie positive M**
c'est une matrice symétrique telle que, pour tout u non nul, $u'Mu > 0$.

- **Définition**

soit $M = (m_{jk})$ définie positive de dimension p . On pose

$$\|u\|_M^2 = u'Mu = \sum_{j=0}^p \sum_{k=1}^p m_{jk} u_j u_k \quad \text{et} \quad d_M^2(u, v) = \|u - v\|_M^2$$

- **Espace métrique**

il est défini par le produit scalaire

$$\langle u, v \rangle_M = u'Mv = \sum_{j=0}^p \sum_{k=1}^p m_{jk} u_j v_k$$

On dit que u et v sont orthogonaux si $\langle u, v \rangle_M = 0$

Similarités, dissimilarités, distances

Métriques particulières

- **Métrique usuelle**

- $M = I$ correspond au produit scalaire usuel et à une distance classique
- **Problèmes**
- la distance entre individus dépend de l'unité de mesure.
- la distance privilégie les variables les plus dispersées.

- **Métrique réduite**

- c'est la plus courante, on prend la matrice diagonale des inverses des variances (cf Mahalanobis)

$$M = D_{1/s^2} = \begin{bmatrix} \frac{1}{s_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_p^2} \end{bmatrix}$$

- **Metric Learning** [Bellet2015]

- Il est aussi possible d'essayer d'adapter la métrique aux données et objectifs (de classification) d'un problème particulier

Similarités, dissimilarités, distances

Métriques et tableaux transformés

- Utiliser la métrique $M = T'T$ sur le tableau X est équivalent à travailler avec la métrique classique I sur le tableau transformé XT' .

- **Tableau transformé**

Si on travaille sur le tableau transformé XT' (changement de variables) au lieu de X , alors les nouveaux individus seront de la forme Te_i et

$$\langle Te_{i_1}, Te_{i_2} \rangle = (Te_{i_1})'(Te_{i_2}) = e_{i_1}'T'Te_{i_2} = e_{i_1}'Me_{i_2} = \langle e_{i_1}, e_{i_2} \rangle_M$$

- **Réciproque**

pour toute matrice symétrique positive M , il existe une matrice T (racine carrée de M) telle que

$$M = T'T$$

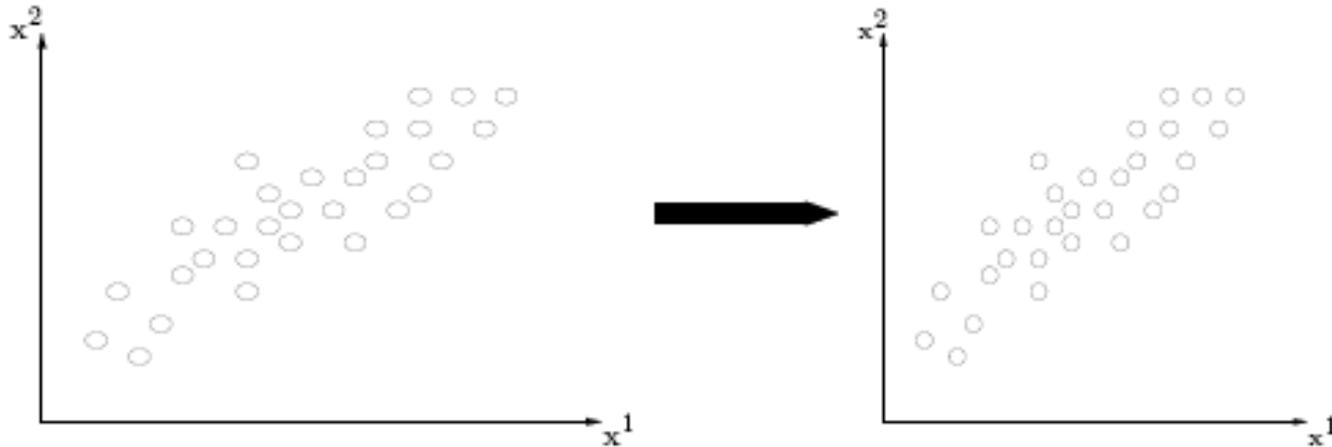
et donc on peut ramener l'utilisation de la métrique à un changement de variables.

- CF notion de **noyau** déjà mentionné dans les cours précédents

Similarités, dissimilarités, distances

Métriques et tableaux transformés (suite)

- Utiliser une métrique est donc équivalent à tordre les données pour les rendre comparables



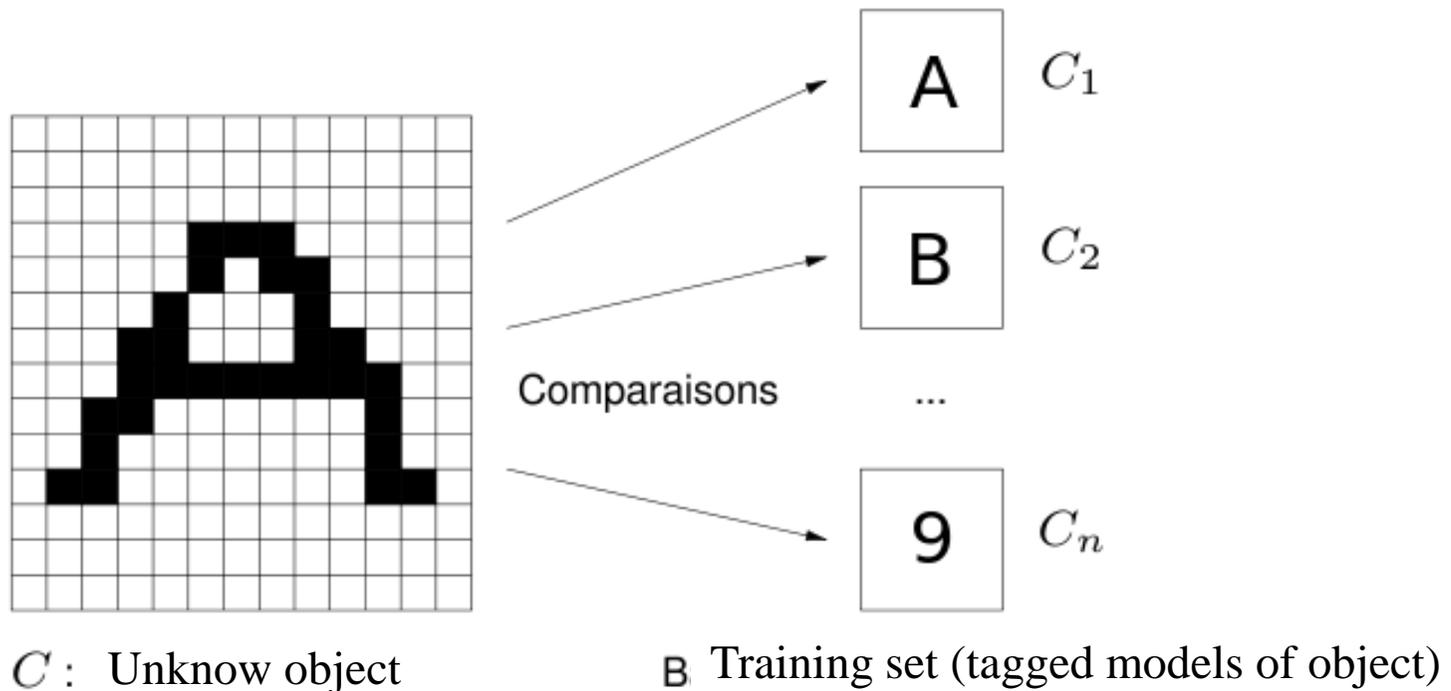
- Exemple utiliser la métrique réduite est équivalent à travailler sur les données centrées réduites $Z = YD_{1/s}$.

Exercice – Calcul de dissimilarité

Object comparison

- Pixels (or zoning) can be considered as features
- A metric can be defined:

$$d^2(C_i, C_{i'}) = \sum_j w_j \times |P(i, j) - P(i', j)|^2$$



Exercice – Calcul de dissimilarité

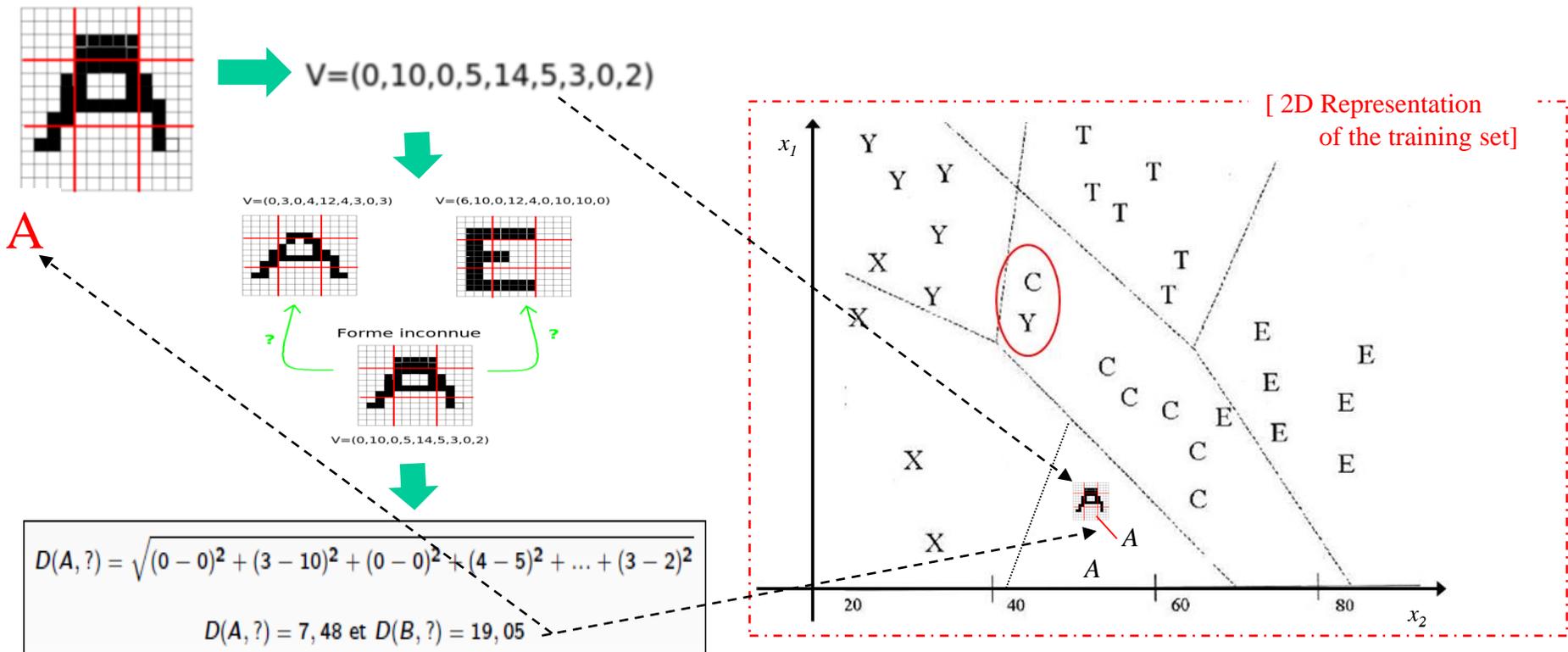
Object comparison

Pixels (or zoning) can be considered as features

A metric can be defined:

$w_j = 1 \forall j \rightarrow$ distance euclidienne classique

Unknown object



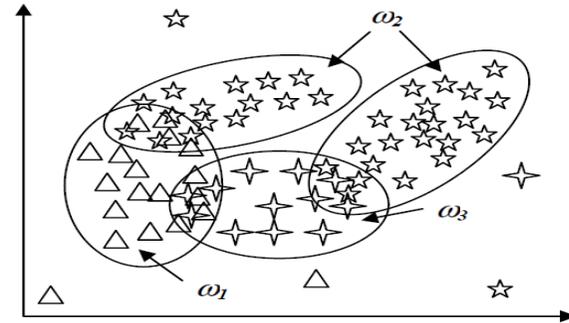
Et si le futur était ailleurs ???

Ne peut on pas aller plus loin que de simples descriptions statistiques ?

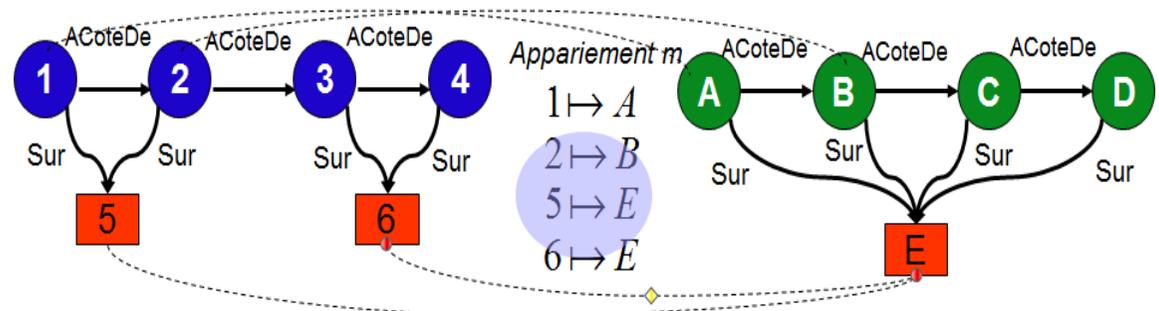
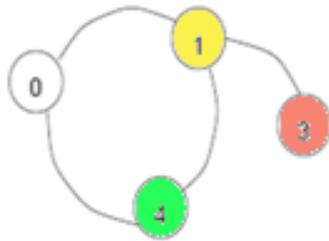
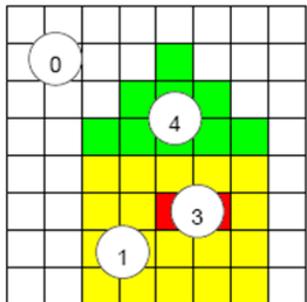
- Pendant des décennies et encore très majoritairement aujourd'hui, ML égal
- 1 Objet \rightarrow 1 Vecteur



$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$



- Mais une autre approche existe (et commence enfin à susciter de l'intérêt) ...
- Descriptions structurales et graph matching [Luqmann2013]
- 1 Objet \rightarrow 1 Graphe



Méthodes statistiques vs structurelles ?

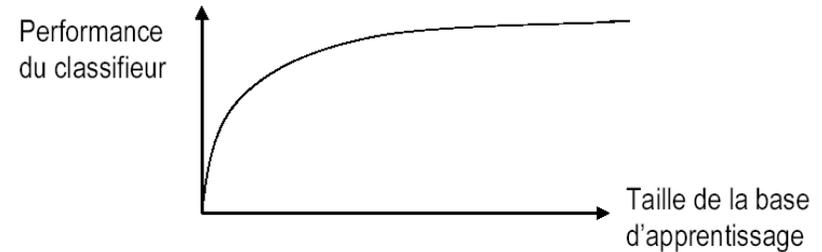
Méthodes statistiques

- Multiplicité des descripteurs bas niveau,
- Boîtes noires qui analysent la séparabilité des classes et permet de conclure sur la pertinence des descripteurs choisis
- Rapide, performant sur des données connues
Multitude de classifieurs
- Apprentissage (base, supervisé ou non)

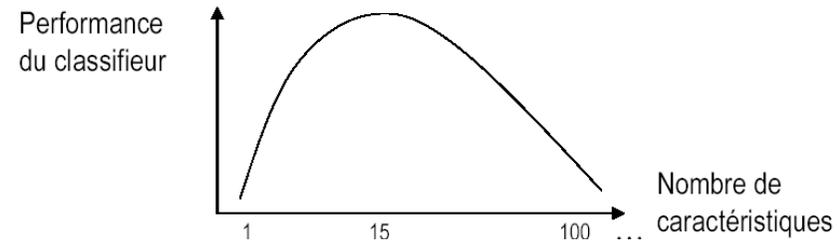
Méthodes structurelles

- Transparence, interopérabilité
- Dimensionnalité adaptative
- Prise en compte du contexte
- Reconnaissance partielle, locale et incrémentale
- Choix des modèles, des caractéristiques
- Temps de calcul, complexité ?
- Apprentissage ?

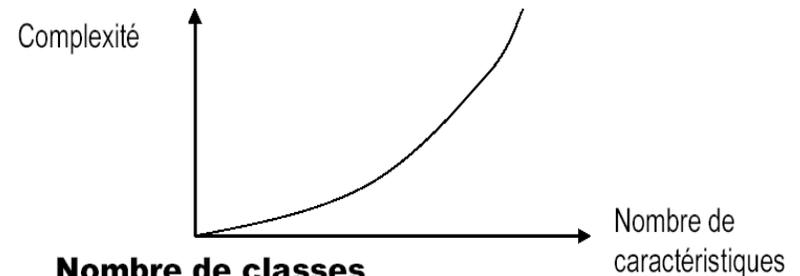
Statistique suffisante



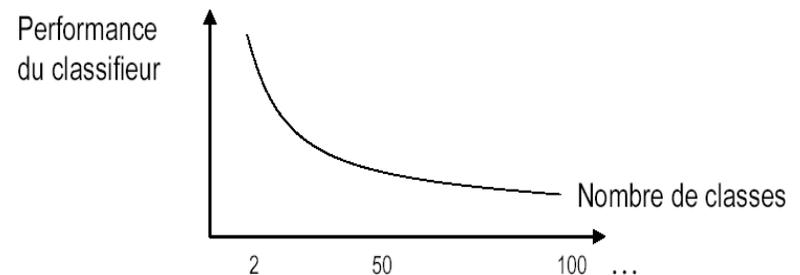
Malédiction de la dimensionalité



Complexité vs nombre de caractéristiques



Nombre de classes



Dans tous les cas...

- Quelques évidences bonnes à rappeler
 - Évidence 1 : sans de bonnes caractéristiques, aucun classifieur ne donnera de bonnes performances
 - Évidence 2 : Avec des caractéristiques raisonnablement pertinentes, tous les classifieurs ne donnent pas les mêmes résultats
 - Évidence 3 : Les données observées (base d'apprentissage) conditionnent complètement les performances
 - Évidence 4 : En cas d'échec, il faut remettre en question les descripteurs, les données observées et enfin le type de classifieur (ou son implémentation)

Références

- Meduim 2018. Data Preprocessing With its implementation in Python, Afroz Chakure. <https://towardsdatascience.com/data-preprocessing-3cd01eefd438>
- George Plastiras ; Maria Terzi ; Christos Kyrkou ; Theocharis Theocharidcs Edge Intelligence: Challenges and Opportunities of Near-Sensor Machine Learning Applications. 2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville et Yoshua Bengio, « Generative Adversarial Networks », Advances in Neural Information Processing Systems 27, 2014
- Harold Hotelling. « Analysis of a Complex of Statistical Variables with Principal Components », 1933, Journal of Educational Psychology.
- Gilbert Saporta, Probabilités, Analyse des données et Statistiques, Paris, Éditions Technip, 2006,
- Kohavi, R., John, G. H. (1997), Wrappers for Feature Subset Selection, Artificial Intelligence, Volume 97, Issue 1-2, Special issue on relevance, p273 – 324.
- Aurelien Bellet, Amaury Habrard, Marc Sebban, Metric Learning (livre). Morgan & Claypool Publishers. 2015
- Muzzamil Luqman, Jean-Yves Ramel, Josep Lladós, Thierry Brouard: Fuzzy multilevel graph embedding. Pattern Recognition 46(2): 551-565 (2013)

Licence

- Cette présentation est distribuée sous licence Creative Commons
- Attribution-ShareAlike 4.0 International

