Extraction de Connaissances à partir de Données
Fouille de Données
Apprentissage Artificiel
Classification - Clustering

Knowledge Discovery from Data
Data Mining
Machine Learning
Classification - Clustering

2018

Gérard Dray
gerard.dray@mines-ales.fr

Objectifs du cours

Knowledge Discovery from Data (KDD)
Extraction de Connaissances à partir de Données (ECD)
Classification - Clustering

Data Mining (DM)
Fouille de Données (FD)
Classification - Clustering

Mythes et réalités
Domaines d'application
Méthodes - modèles - outils
Mise en œuvre

Organisation du cours

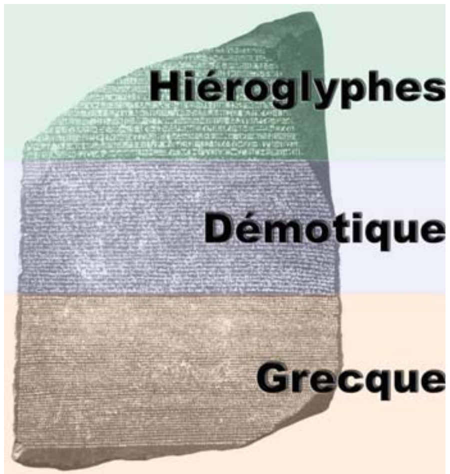
Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

Une histoire ancienne

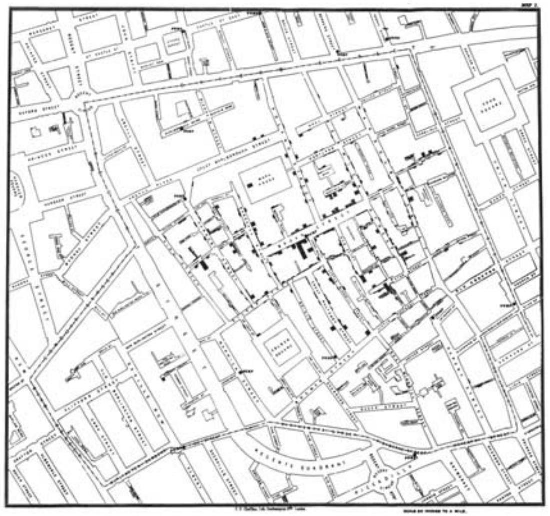
- Jean-François Champollion 1821



«Au cours du règne du jeune qui a succédé à son père, Seigneur des diadèmes, très glorieux, qui a établi l'Egypte et a été pieux envers les dieux, triomphant sur ses ennemis et ramenant la paix et la vie civilisée aux hommes, Seigneur des Cérémonies des Trente années, semblable à Ptah le Grand, un roi comme Ra, grand roi des pays Supérieur et Inférieur, progéniture des Dieux Philopatores, approuvé par Ptah, à qui Ra a donné la victoire, l'image vivante d'Amun, fils deRa, PTOLÉMÉE, VIVRA A JAMAIS, BIEN-AIMÉ DE PTAH, dans la neuvième année, quand le fils Aetos d'Aetos était prêtre d'Alexandrie, et les Dieux Soteres, et les Dieux Adelphoi, et les Dieux Euergetai, et les Dieux Philopatores et le Dieu Epiphanes Eucharistos; fille Pyrrha de Philinos qui est Athlophoros de Berenike Euergetis, fille Areia de Diogenes qui est Kanephoros d'Arsinoe Philadelphos; fille Irene de Ptolémée qui est Prêtresse d'Arsinoe Philopator; les quatrièmes du mois de Xandikos, d'après les Égyptiens les 18e Mekhir …»

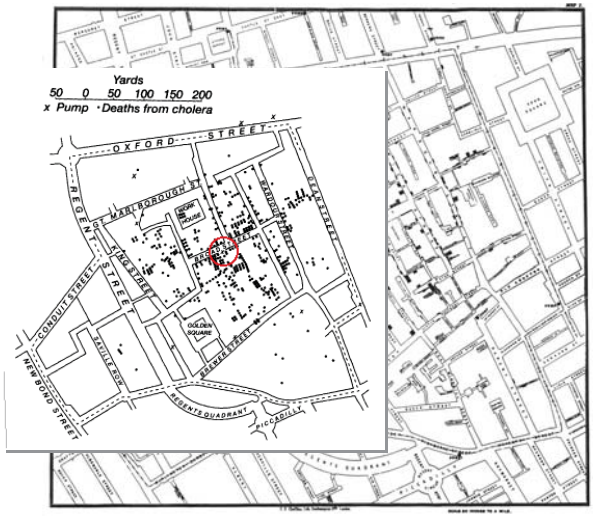
Une histoire ancienne

- John Snow 1854



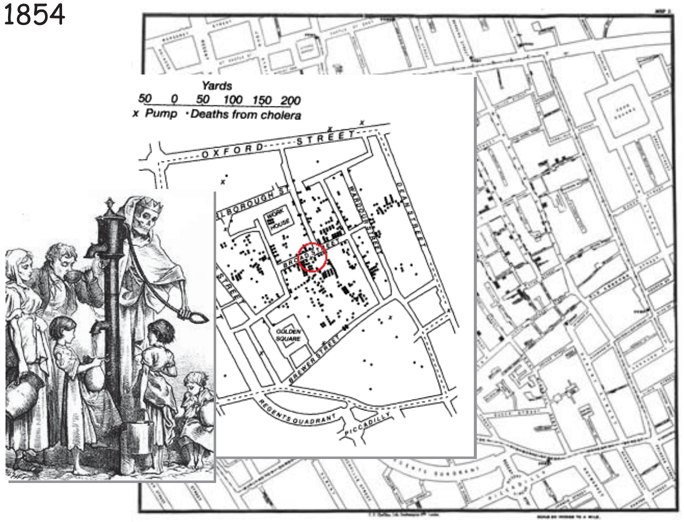
Une histoire ancienne

- John Snow 1854



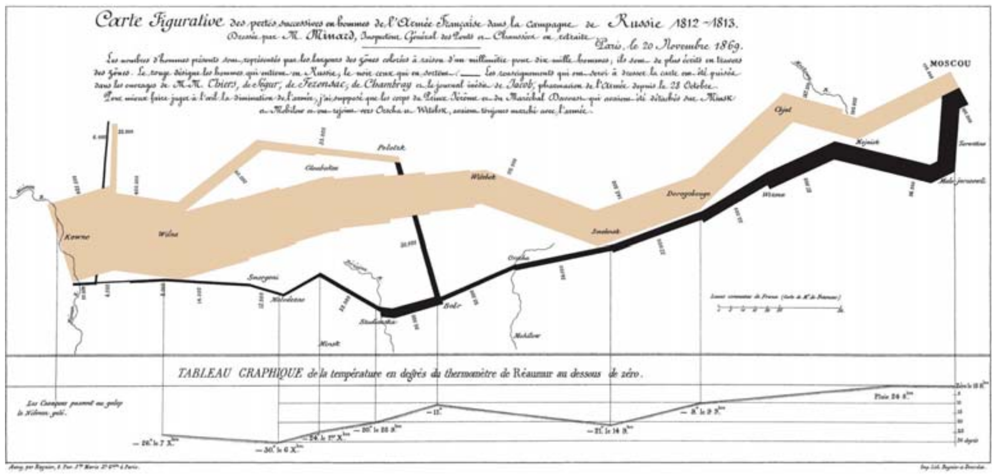
Une histoire ancienne

- John Snow 1854



Une histoire ancienne

- Charles Joseph Minard 1869



Une histoire ancienne

« Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's ? If this were then subjected to an appropriate course of education one would obtain the adult brain. »

Turing 1963

Data Mining

- Définition

« The non-trivial process of identifying valid, novel, potentially usefull and ultimately understandable patterns in data »

Fayyad, Shapiro et Smyth 1996

1998 : Une légende urbaine ?

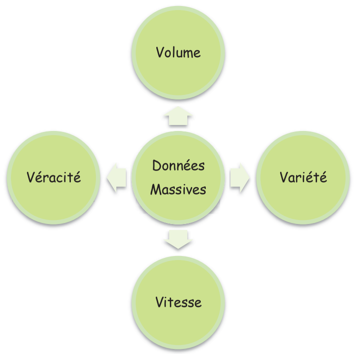
1998 : Une légende urbaine ?

1998 : Une légende urbaine ?

1998 : Une légende urbaine ?

En 2015

Le phénomène Big data (Données Massives) est considéré comme l'un des grands défis informatiques de la décennie 2010-2020



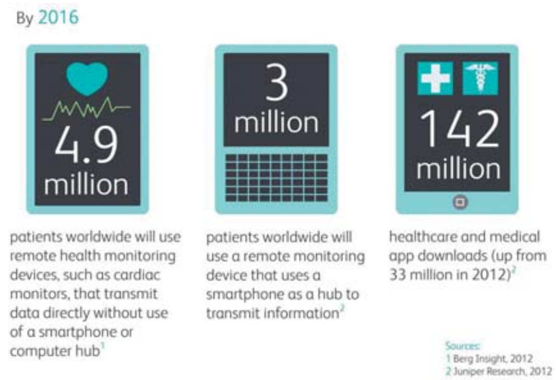
Volume = Quantité : de terabytes à zettabytes
Vitesse = Traitement par lots ou en flux temps réel
Variété = Structurées, Semi-structurées et non-structurées
Véracité = confiance en l'information

Le mouvement de l'open data considère les données publiques comme un bien commun

En 2015

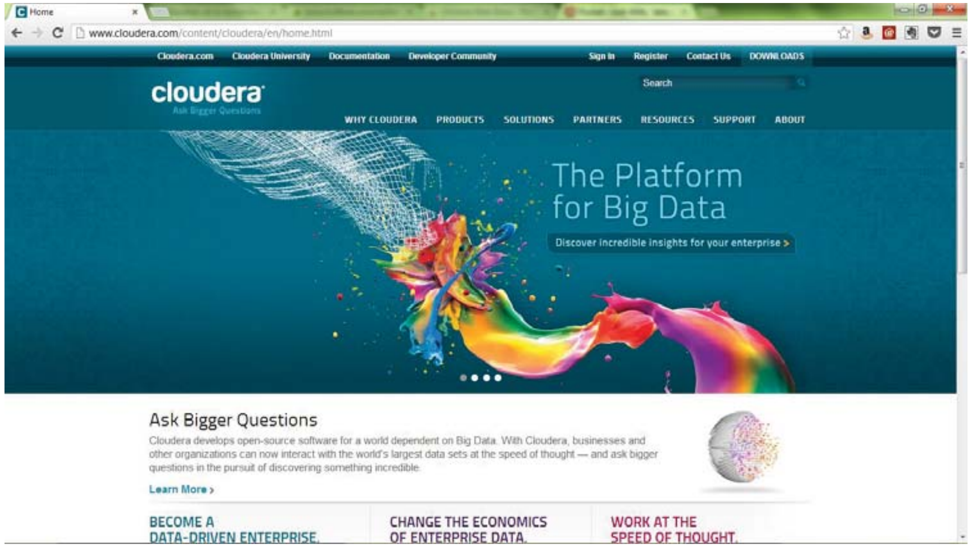
Internet des objets

Quantified Self



Caisse nationale de l'assurance maladie :
plus d'un milliard de demandes de remboursement par an

En 2015



En 2015

Josh Wills, 'data scientist' chez Cloudera

Qu'est-ce qu'un data scientist ?

« *Je suis d'abord un 'nerd des maths' qui trouve que la visualisation c'est cool !* », nous a affirmé Josh Wills. Pour lui, un data scientist doit d'abord passer beaucoup de temps à nettoyer la donnée. « *Plus propre elle sera, plus efficace sera l'analytique. Nous devons penser à résoudre des problèmes et à basculer les données dans un environnement opérationnel. Je passe mon temps à essayer de multiples idées, à paralléliser tout ce que je fais, à trouver des solutions en 6 mois contre plusieurs années auparavant, et à réaliser des recherches reproductibles.* »

Source : http://www.silicon.fr/josh-wills-data-scientist-chez-cloudera-82387.html 4/1/2013

En 2015

Comme personne ne peut maitriser tous les processus technologiques, la première étape d'un projet analytique sera de créer une équipe. « *Informaticiens et statisticiens ne parlent pas le même langage. Ils devront pourtant créer un modèle, tout mesurer, expérimenter, tester encore et toujours. Puis trouver de nouveaux modèles. Avec le risque de créer un modèle complexe que personne ne comprendra !* »

La complexité tient souvent dans l'étendue des volumes de données à traiter. C'est pourquoi, pour amortir les coûts, les acteurs de l'analytique doivent créer de l'automatisation. Un point de vue auquel adhère Josh Wills, qui se veut cependant prudent : « *Optimiser un modèle ne se traduit pas toujours par l'optimisation du business. Nous ne croyons que dans la production, mais il existe un gap entre le business model et machine model* »

Source : http://www.silicon.fr/josh-wills-data-scientist-chez-cloudera-82387.html 4/1/2013

En 2015

Un métier en devenir

Selon l'étude McKinsey « *Global Institute Big Data Report* », de 140.000 à 190.000 postes de data scientist devraient être créés aux États-Unis, principalement dans la santé. « *Les outils sont là, mais les gens ne savent pas les utiliser, ni établir les passerelles pour cela. Toutes les universités dans le monde devraient avoir un cursus data scientist* ».

Quant aux difficultés qu'il rencontre dans l'exercice de son métier, Josh Wills les exprime sans ambages : « *Le volume est un problème, le rythme de changement l'est également. Tout le monde a des problèmes d'ETL. Et nous n'avons pas besoin de programmeurs Java…* », probablement un retour d'expérience malheureux… Et comment démarrer un projet ? « *La recherche est le premier 'use case' d'Hadoop, car toute information a un document.* »

Source : http://www.silicon.fr/josh-wills-data-scientist-chez-cloudera-82387.html 4/1/2013

En 2015

Un métier qui n'a pas de prix… pour le moment !

Notre dernière question portera sur le prix d'un data scientist sur le marché ? *« Le prix dépend de l'activité. Par exemple dans la pub c'est très cher, mais ce n'est pas défini».* Et de nous rappeler qu'une rémunération peut être indexée à un résultat, comme par exemple à un chiffre d'affaires réalisé à la suite d'une analyse…

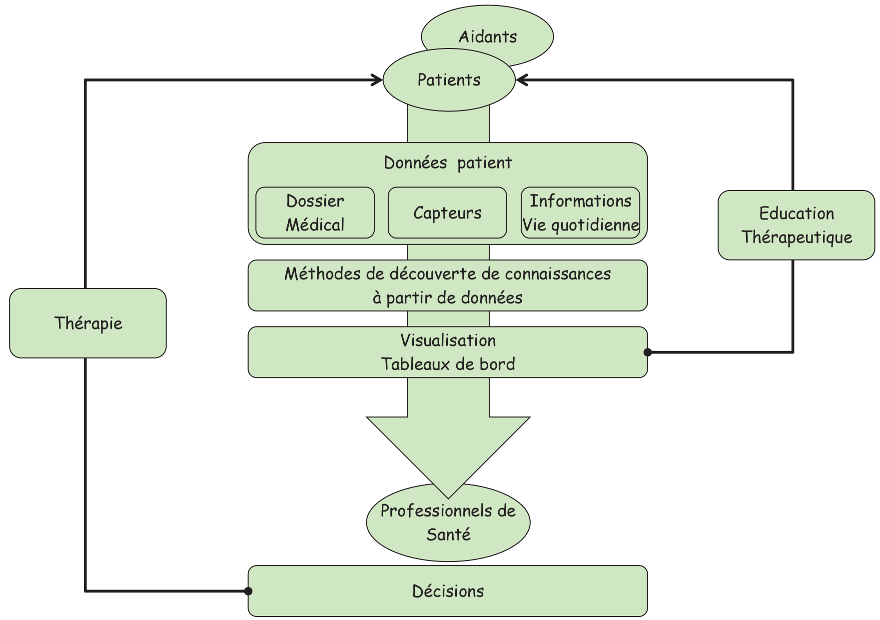
Ses yeux se mettent alors à briller. Les bons data scientists (et les data scientists eux-mêmes) sont une denrée rare, et pour quelques années encore avant que les cursus de formation ne crachent leurs diplômés, formés mais inexpérimentés.

Quant aux développeurs et autres consultants informatiques qui prétendent à l'expertise du statisticien pour exploiter les Big data, la concurrence ne sera pas rude avant longtemps. Ce n'est pas pour rien que Josh Wills nous quitte en conservant l'éclat brillant de ses yeux et son sourire entendu.

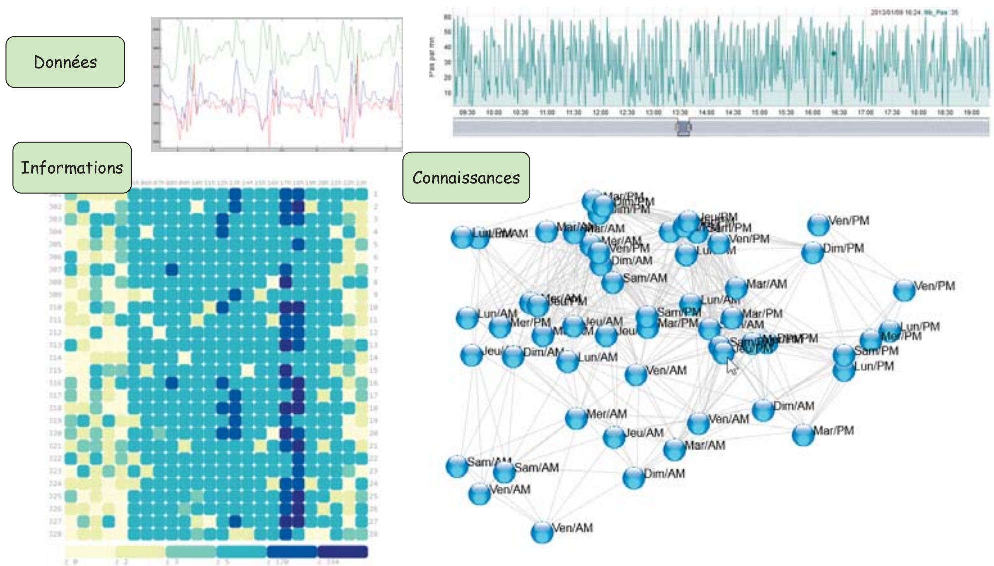
Il se murmure même chez Cloudera que le million de dollars en rémunération d'une mission de data scientist n'a rien d'extravagant au vu du service rendu. Les IT ont encore de quoi nous faire rêver…

Source : http://www.silicon.fr/josh-wills-data-scientist-chez-cloudera-82387.html
4/1/2013

Découverte de connaissances à partir de données patient

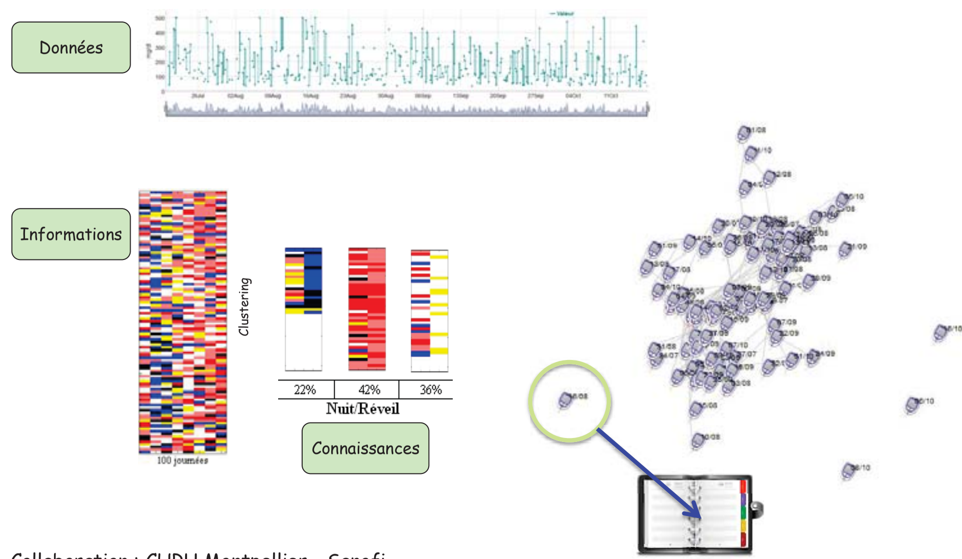


Activité physique Suivi d'un patient



Collaboration : CHRU Montpellier – Association RAP - Industriels

Diabète Suivi d'un patient



Collaboration : CHRU Montpellier - Sanofi

KDD et DM - Pourquoi ?

- **Une avalanche de données**
  - Avancées technologiques pour acquérir et sauvegarder les données
  - Le volume des données sauvegardées double tous les 20 mois
  - L'homme va créer plus d'information dans les deux à trois ans à venir qu'au cours des 40000 dernières années
  - Seulement 10% des données collectées sont analysées

Solutions :

- Centraliser les données : Datawarehouse - Datamart
- Extraire les connaissances intéressantes (KDD) par des méthodes et des techniques performantes (DM)

Problèmes et approches du KDD

- **Problèmes**
  - Identification de données pertinentes
  - Représentativité des données
  - Recherche de caractéristiques et/ou modèles valides
- **Approches**
  - Top-down : déduction par experts
  - Interactive : visualisation des données

    (OLAP : On Line Analytical Processing)
  - Bottom-up : induction à partir des données (Data Mining)

KDD qu'est-ce que c'est ? un processus !

- **Sélection et traitement des données pour :**
  - La modélisation de phénomènes réels
  - La mise en évidence de caractéristiques essentielles
- **Le Data Mining est un composant majeur du processus de KDD**
  - Développement de modèles prévisionnels et explicatifs
  - Découverte automatique de caractéristiques

KDD - La démarche

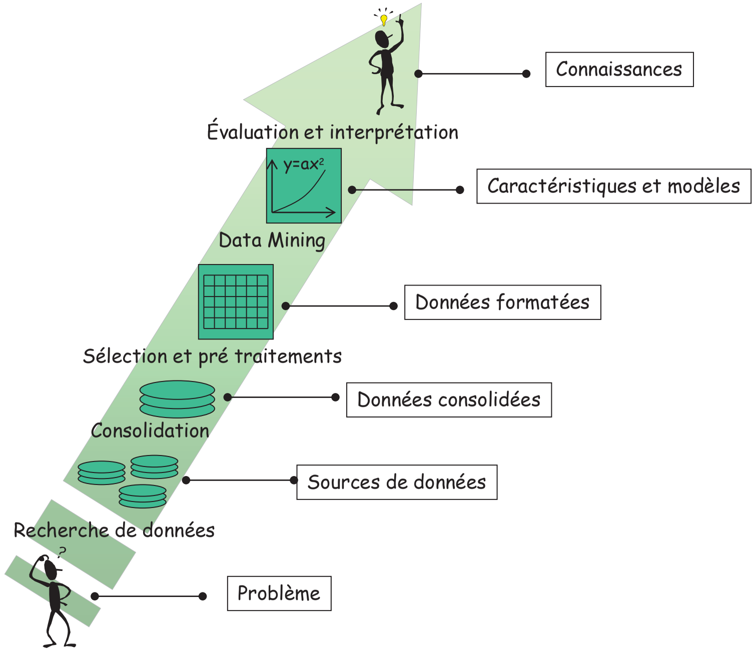
Décision
- Promouvoir le produit P dans la région R durant la période N
- Réaliser un mailing sur le produit P aux familles du profil F

Connaissance
- Une quantité Q du produit P est vendue en région R
- Les familles de profil F utilisent M% de P durant la période N

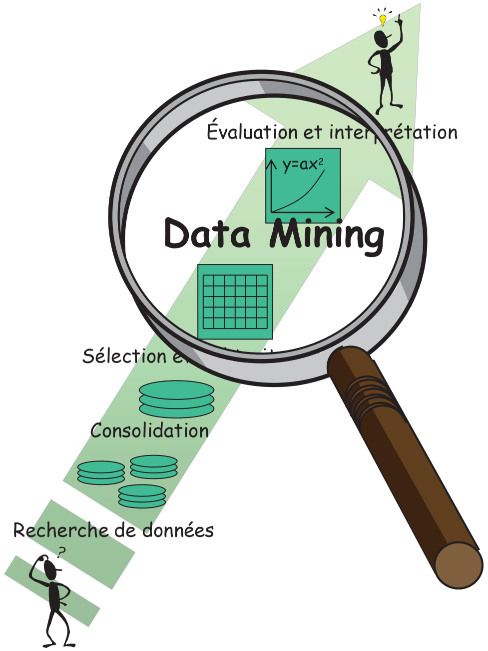
Information
- X habite la région R
- Y a A ans
- Z dépense son argent dans la ville V de la région R

Données
- Consommateurs
- Magasins
- Ventes
- Démographie
- Géographie

KDD - Le processus

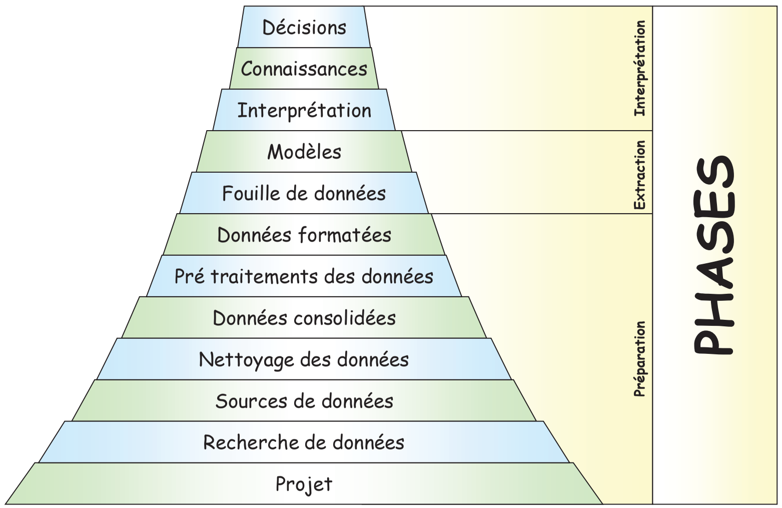


KDD - Le processus



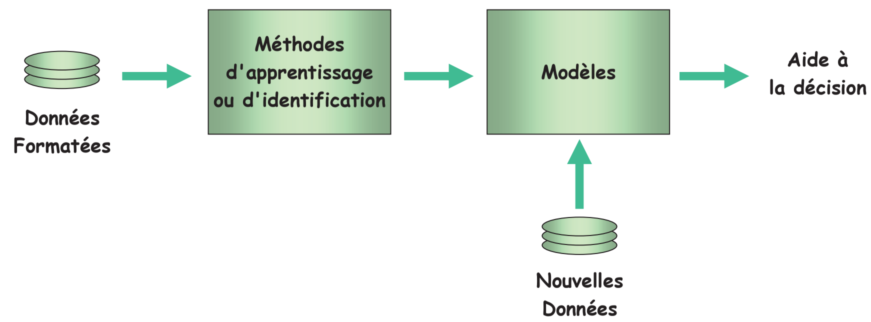
Objet de ce cours

KDD - Le processus



Data Mining

- Processus d'induction à partir des données



Nature des données

- Tableau (Individus x Caractères) de dimension (n x p)

		Caractères										
		Age	Revenu	Sexe		Situation Matrimoniale			…	Caractère j	…	Caractère p
				M	F	Marié	Célibataire	Veuf Divorcé				
		X^1	X^2	X^3	X^4	X^5	X^6	X^7	…	X^j	…	X^p
Individus	X_1	x_1^1	x_1^2	x_1^3	x_1^4	x_1^5	x_1^6	x_1^7	…	x_1^j	…	x_1^p
	X_2	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_2^6	x_2^7	…	x_2^j	…	x_2^p
	…	…	…	…	…	…	…	…	…	…	…	…
	X_i	x_i^1	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	x_i^7	…	x_i^j	…	x_i^p
	…	…	…	…	…	…	…	…	…	…	…	…
	X_n	x_n^1	x_n^2	x_n^3	x_n^4	x_n^5	x_n^6	x_n^7	…	x_n^j	…	x_n^p

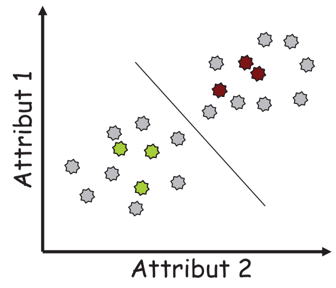
Vecteur Individus $X_i = [x_i^1, x_i^2, ..., x_i^j, ..., x_i^p]$

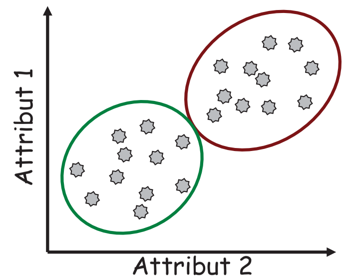
Vecteur Caractère $X^j = [x_1^j, x_2^j, ..., x_i^j, ..., x_n^j]^T$

- Les caractères Age et Revenu sont quantitatifs
- Les caractères Sexe et Situation matrimoniale sont qualitatifs
- Modalité du caractère Situation matrimoniale : (Célibataire, Marié, Veuf ou divorcé)
- Les variables : X^3, X^4, X^5, X^6 et X^7 sont booléennes (1 = vrai, 0 = faux)

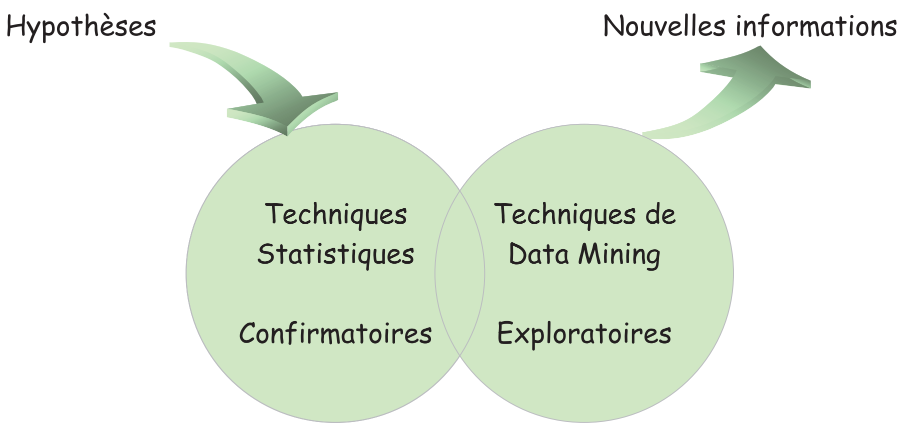
Classification supervisée
Classification non supervisée (clustering)

- Les objets possèdent un ou plusieurs caractères
- Classification
  - Quelques points sont labélisés
  - Recherche de règles pour labéliser les autres points
  - Apprentissage supervisé
- Clustering
  - Pas de points labélisés
  - Grouper les points par similarité
  - Apprentissage non supervisé





Data Mining vs Statistiques



KDD et DM - Domaines d'application

- **Science et médecine** : prévision, classification, diagnostic, …
- **Marketing** : segmentation, ciblage, géomarketing, …
- **Finance** : gestion des portefeuilles, prévision, …
- **Banques** : acceptation de crédit, segmentation, …
- **Sécurité** : détection de fraudes, d'impayés, …
- **Industrie** : contrôle de qualité, allocation de ressources, …
- **Ingénierie** : analyse de signaux, reconnaissance de motifs, …
- **Internet** : moteurs de recherche intelligents, e-marketing, text mining, …
- **Environnement** : prévision des pics de pollution, analyse des risques, …

Organisation du cours

Introduction

Bases du KDD et du DM

Bases de l'analyse des données

Analyse factorielle

Classification

Clustering

Logique floue

Réseaux de neurones artificiels

Conclusion

Les étapes du processus de KDD

- Étape 1 : Poser le problème
- Étape 2 : Rechercher les données
- Étape 3 : Sélectionner les données pertinentes
- Étape 4 : Nettoyer les données
- Étape 5 : Transformer les données
- Étape 6 : Rechercher les caractéristiques et les modèles
- Étape 7 : Évaluer et valider les résultats

Les étapes du processus de KDD
Étape 1 : Poser le problème

- Quel est la problématique ?
- Quels sont les objectifs ?
- Quels sont les résultats attendus ?
- Comment seront-ils évalués ?

Les étapes du processus de KDD
Étape 1: Poser le problème

Un exemple : Classification de fleurs d'iris

- Problème :

  Connaissant :
  - la longueur des sépales d'une fleur d'iris
  - la largeur des sépales d'une fleur d'iris
  - la longueur des pétales d'une fleur d'iris
  - la largeur des pétales d'une fleur d'iris

  Est-il possible de classer les fleurs dans les catégories suivantes ?
  - Iris Setosa - classe 1
  - Iris Versicolour - classe 2
  - Iris Virginica - classe 3

- L'objectif est de fournir un modèle de classification qui accepte en entrée les 4 caractéristiques des sépales et des pétales des fleurs d'iris et qui fournit en sortie le numéro de la classe de la fleur (1, 2 ou 3).

- Les résultats de la classification seront évalués sur une base de données de test représentant 25% des données collectées sous la forme d'une matrice de confusion.

Les étapes du processus de KDD
Étape 2 : Rechercher les données

- Identifier les informations
- Identifier les sources
- Vérifier leur qualité
- Vérifier leur facilité d'accès
  - Documents papier
  - Supports électroniques
  - Fichiers internes ou externes
  - Sources multiples, Data Warehouse ou Data Mart

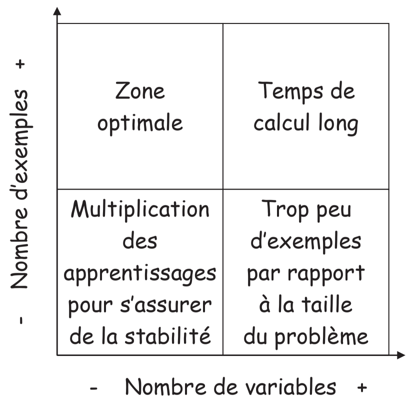
Les étapes du processus de KDD
Étape 2 : Rechercher les données

Un exemple : Classification de fleurs d'iris

- Title: Iris Plants Database Updated Sept 21 2000 by C.Blake
- File : iris.dat
- Sources:
  - (a) Creator: R.A. Fisher
  - (b) Donor: Michael Marshall
  - (c) Date: July, 1988
- Number of Instances: 150 (50 in each of three classes)
- Number of Attributes: 4 numeric attributes and the class
- Attribute Information:
  - 1. sepal length in cm
  - 2. sepal width in cm
  - 3. petal length in cm
  - 4. petal width in cm
  - 5. class:
  - -- Iris Setosa
  - -- Iris Versicolour
  - -- Iris Virginica
- Missing Attribute Values: None

Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

- Réduire les dimensions
  - Expertise humaine
  - Analyses graphiques
  - Analyses de corrélation
  - Analyse en composantes principales
  - …



Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

Statistiques sur toutes les classes

Coefficient de corrélation	L sépale	l sépale	L pétale	l pétale	Classe
L sépale	1,00	-0,11	0,87	0,82	0,78
l sépale	-0,11	1,00	-0,42	-0,36	-0,42
L pétale	0,87	-0,42	1,00	0,96	0,95
l pétale	0,82	-0,36	0,96	1,00	0,96
Classe	0,78	-0,42	0,95	0,96	1,00
Minimum	4,30	2,00	1,00	0,10	1,00
Maximum	7,90	4,40	6,90	2,50	3,00
Moyenne	5,84	3,05	3,76	1,20	2,00
Ecart-type	0,83	0,43	1,76	0,76	0,82
Médiane	5,80	3,00	4,35	1,30	2,00
Interquartile	1,30	0,50	3,50	1,50	2,00

Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

Statistiques sur la classe 1

Coefficient de corrélation	L sépale	l sépale	L pétale	l pétale
L sépale	1,00	0,75	0,26	0,28
l sépale	0,75	1,00	0,18	0,28
L pétale	0,26	0,18	1,00	0,31
l pétale	0,28	0,28	0,31	1,00
Minimum	4,30	2,30	1,00	0,10
Maximum	5,80	4,40	1,90	0,60
Moyenne	5,01	3,42	1,46	0,24
Ecart-type	0,35	0,38	0,17	0,11
Médiane	5,00	3,40	1,50	0,20
Interquartile	0,40	0,60	0,20	0,10

Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

Statistiques sur la classe 2

Coefficient de corrélation	L sépale	l sépale	L pétale	l pétale
L sépale	1,00	0,53	0,75	0,55
l sépale	0,53	1,00	0,56	0,66
L pétale	0,75	0,56	1,00	0,79
l pétale	0,55	0,66	0,79	1,00
Minimum	4,90	2,00	3,00	1,00
Maximum	7,00	3,40	5,10	1,80
Moyenne	5,94	2,77	4,26	1,33
Ecart-type	0,52	0,31	0,47	0,20
Médiane	5,90	2,80	4,35	1,30
Interquartile	0,70	0,50	0,60	0,30

Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

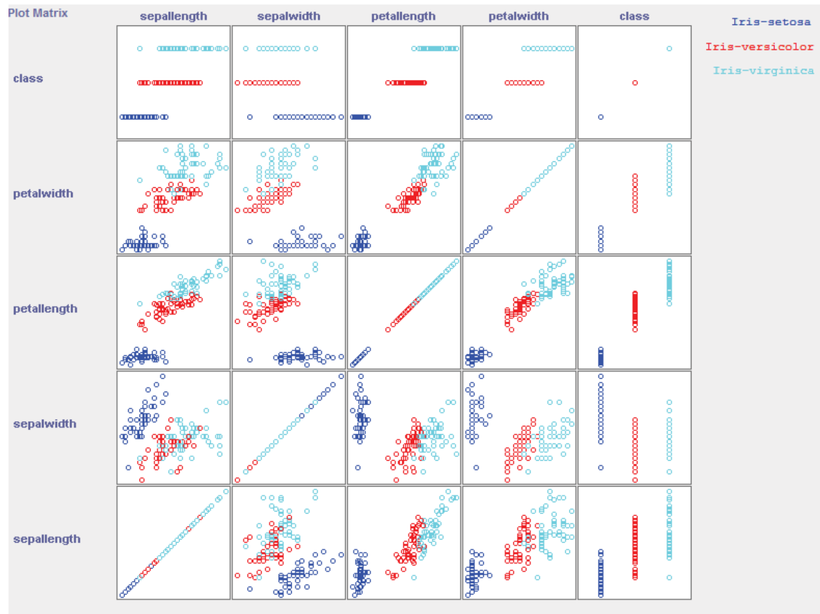
Un exemple : Classification de fleurs d'iris

Statistiques sur la classe 3

Coefficient de corrélation	L sépale	l sépale	L pétale	l pétale
L sépale	1,00	0,46	0,86	0,28
l sépale	0,46	1,00	0,40	0,54
L pétale	0,86	0,40	1,00	0,32
l pétale	0,28	0,54	0,32	1,00
Minimum	4,90	2,20	4,50	1,40
Maximum	7,90	3,80	6,90	2,50
Moyenne	6,59	2,97	5,55	2,03
Ecart-type	0,64	0,32	0,55	0,27
Médiane	6,50	3,00	5,55	2,00
Interquartile	0,70	0,40	0,80	0,50

Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

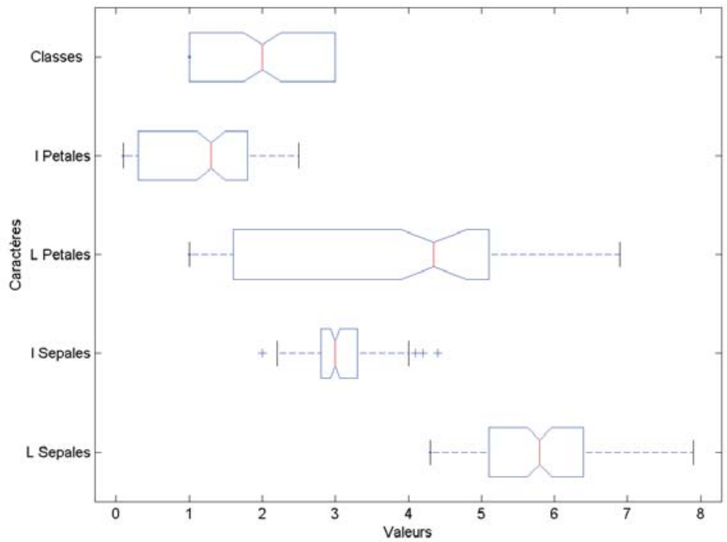
Un exemple : Classification de fleurs d'iris



Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

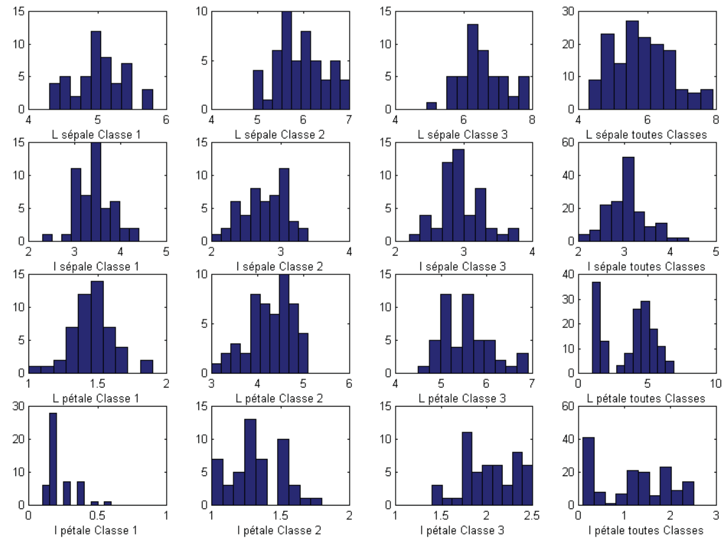
Médiane et quartiles



Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

Histogrammes des caractéristiques



Les étapes du processus de KDD
Étape 3 : Sélectionner les données pertinentes

Un exemple : Classification de fleurs d'iris

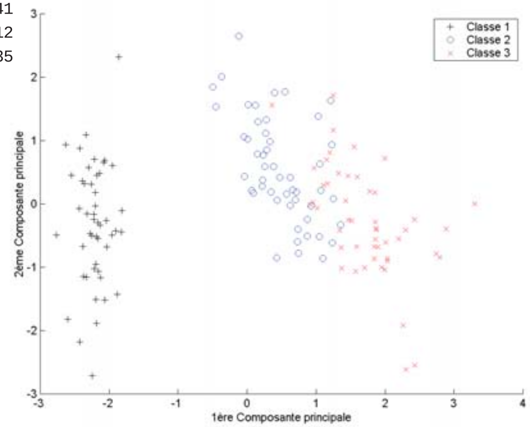
Analyse en composantes principales

Coefficients

	CP1	CP2	CP3	CP4
L sépale	0.5224	-0.3723	0.7210	-0.2620
l sépale	-0.2634	-0.9256	-0.2420	0.1241
L pétale	0.5813	-0.0211	-0.1409	0.8012
l pétale	0.5656	-0.0654	-0.6338	-0.5235

Valeurs propres et % d'inertie

CP1	2.9108	72.77%
CP2	0.9212	23.03%
CP3	0.1474	03.69%
CP4	0.0206	00.51%



Les étapes du processus de KDD
Étape 4 : Nettoyer les données

- Vérifier l'origine des données
- Traiter les valeurs aberrantes
- Traiter les valeurs manquantes
- Traiter les valeurs nulles
- Estimer la qualité des données

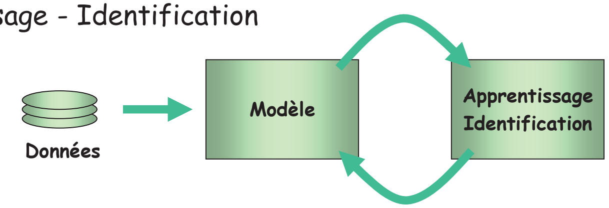
Les étapes du processus de KDD
Étape 5 : Transformer les données

- Cette étape peut consister à :
  - Coder les informations qualitatives
  - Coder en ratios (pourcentages)
  - Normaliser les données
  - Transformer les dates en durées
  - Transcoder les données,

    exemple : code postal en coordonnées géographiques
  - Exprimer des fréquences
  - Exprimer des tendances
  - Réaliser des combinaisons de variables
  - ...

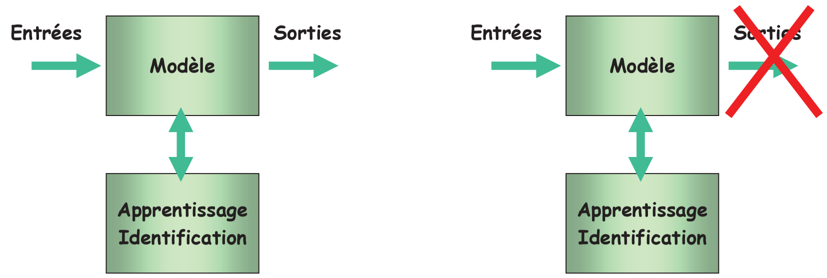
Les étapes du processus de KDD
Étape 6 : Rechercher les caractéristiques et les modèles

- Apprentissage - Identification



- Apprentissage Supervisé

Apprentissage Non supervisé



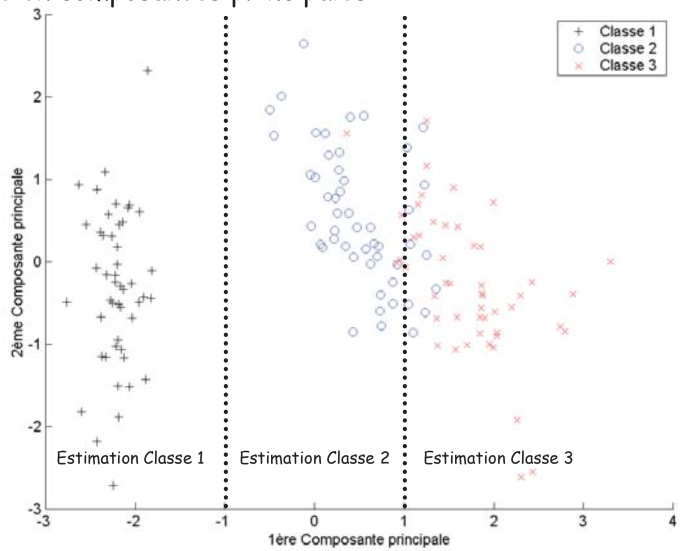
Les étapes du processus de KDD
Étape 6 : Rechercher les caractéristiques et les modèles

- Choix d'une méthode ou d'une technique :
  - Réseaux de neurones
  - Régression
  - Règles
  - Arbres de décision
  - Analyses factorielles
  - Logique Floue
  - Algorithmes génétiques
  - Le raisonnement à base de cas
  - Les « Rough Sets »
  - Les réseaux Bayésiens
  - ...

Les étapes du processus de KDD
Étape 6 : Rechercher les caractéristiques et les modèles

Un exemple : *Classification de fleurs d'iris*

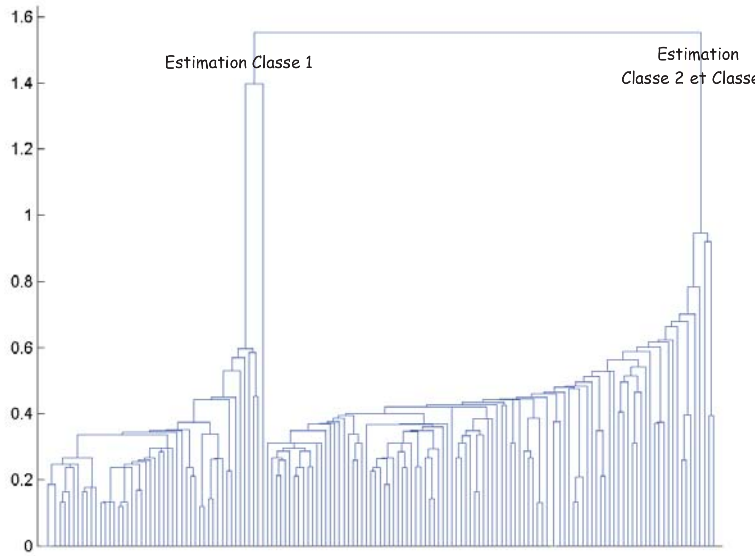
Analyse en composantes principales



Les étapes du processus de KDD
Étape 6 : Rechercher les caractéristiques et les modèles

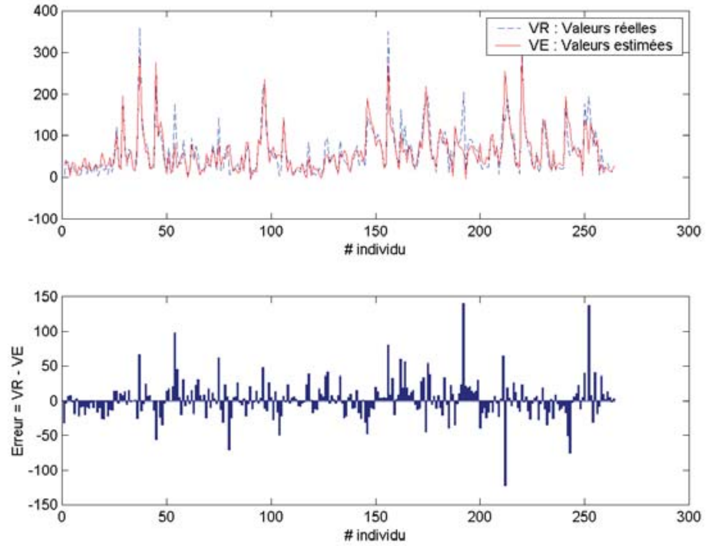
Un exemple : Classification de fleurs d'iris

Groupement hiérarchique



Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation qualitative - exemple

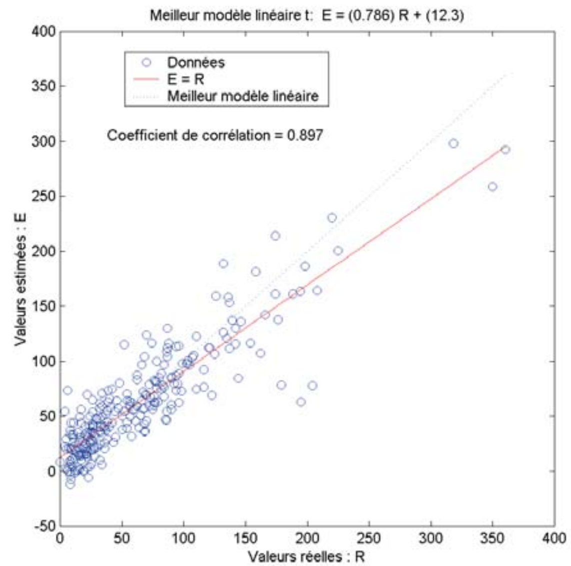


Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Erreur quadratique moyenne
    (MSE : Mean Square Error)
  - Erreur absolue moyenne
    (MAE : Mean Absolute Error)
  - Racine carrée de l'erreur quadratique moyenne
    (RMSE : Root Mean Square error)
  - Erreur relative absolue moyenne
    (MAPE : Mean Absolute Percentage Error)
  - Coefficient de corrélation

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative - exemple

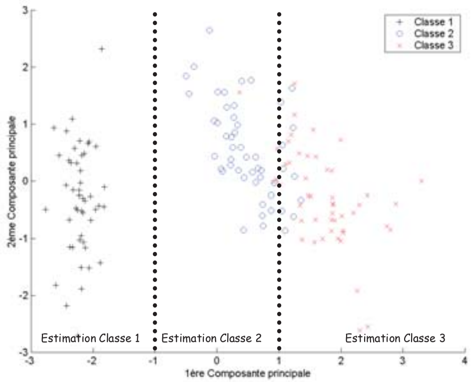


Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - matrice de confusion

    exemple : Classification de fleurs d'iris

		Classes estimées		
		C1	C2	C3
Classes réelles	C1	50	0	0
	C2	0	40	10
	C3	0	4	46



Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - matrice de confusion

    exemple : Classification de fleurs d'iris

		Classes estimées		
		C1	C2	C3
Classes réelles	C1	49	1	0
	C2	0	47	3
	C3	0	2	48

```
Arbre de décision
--------------------

petalwidth <= 0.6: Iris-setosa
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica
|   |   |   petalwidth > 1.5: Iris-versicolor
|   petalwidth > 1.7: Iris-virginica
```


Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Correctly Classified Instances et

    Incorrectly Classified Instances

    exemple : Classification de fleurs d'iris

		Classes estimées		
		C1	C2	C3
Classes réelles	C1	49	1	0
	C2	0	47	3
	C3	0	2	48

Arbre de décision

Correctly Classified Instances	**144**	**96%**
Incorrectly Classified Instances	**6**	**4%**

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Kappa Statistic

    exemple : Classification de fleurs d'iris

		Classes estimées		
		C1	C2	C3
Classes réelles	C1	49	1	0
	C2	0	47	3
	C3	0	2	48

$$P_0 = \frac{49+47+48}{150} = 0.96$$

$$P_e = \frac{50\times 49 + 50\times 50 + 50\times 51}{150\times 150} = \frac{1}{3}$$

$$K = \frac{0.96 - 1/3}{1 - 1/3} = 0.94$$

Arbre de décision

Kappa statistic **0.94**

Mesure du degré de concordance de 2 juges (Classifieur et Expert)

$$K = \frac{P_0 - P_e}{1 - P_e}$$

P_0 : proportion de l'échantillon pour laquelle les juges sont d'accord

$$P_e = \frac{\sum_i p_i \times p_{.i}}{n^2}$$

p_i : somme des éléments de la ligne i

$p_{.i}$: somme des éléments de la colonne i

n : taille de l'échantillon

Le coefficient Kappa est un nombre réel, sans dimension, compris entre **-1** et **1**. L'accord sera d'autant plus élevé que la valeur de Kappa est proche de 1 et l'accord maximal est atteint (K = 1) lorsque P_o = 1 et P_e = 0,5.

Lorsqu'il y a **indépendance des jugements**, le coefficient Kappa est égal à **zéro** (P_o = P_e), et dans le cas d'un désaccord total entre les juges, le coefficient Kappa prend la valeur **-1** avec P_o = 0 et P_e = 0,5.

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Calculs d'erreurs

    exemple : Classification de fleurs d'iris

Classes réelles		Classes estimées C1	C2	C3
	C1	49	1	0
	C2	0	47	3
	C3	0	2	48

Arbre de décision

Mean absolute error	0.035
Root mean squared error	0.1586
Relative absolute error	7.8705 %
Root relative squared error	33.6353 %

Soient $p_1, p_2, \cdots, p_n$ les probabilités calculées par le classifieur pour chaque exemple d'appartenir à sa *vrai classe*.
Soient $a_1, a_2, \cdots, a_n$ les probabilités à priori pour chaque exemple d'appartenir à la classe qui leur a été fixée par définition (en général, les a_i valent toujours 1, mais on peut imaginer qu'on soit un peu moins catégorique, et que la classe attribuée ne le soit qu'avec une certaine confiance).
soit $\bar{a} = \frac{1}{n}\sum_i a_i$

Mean absolute error
$$\frac{|p_1 - a_1| + |p_2 - a_2| + \cdots + |p_n - a_n|}{n}$$

Root mean-squared error
$$\sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{n}}$$

Relative absolute error
$$\frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{|a_1 - \bar{a}| + \cdots |a_n - \bar{a}|}$$

Root relative squared error
$$\sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \cdots (a_n - \bar{a})^2}}$$

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Notions de Vrais/Faux Positifs et Vrais/Faux Négatifs

Classe réelles	Classes estimées C	$\overline{C}$
C	VP	FN
$\overline{C}$	FP	VN

VP :Vrais Positifs, VN : Vrais Négatifs,
FN : Faux Négatifs, FP : Faux Positifs

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Taux de VP et Taux de FP
  - exemple : Classification de fleurs d'iris

Classe réelles	Classes estimées C	$\overline{C}$
C	VP	FN
$\overline{C}$	FP	VN

	C1	$\overline{C1}$
C1	49	1
$\overline{C1}$	0	100

	C2	$\overline{C2}$
C2	47	3
$\overline{C2}$	3	97

	C3	$\overline{C3}$
C3	48	2
$\overline{C3}$	3	97

Arbre de décision

TP Rate	FP Rate	Class
0.98	0	Iris-setosa
0.94	0.03	Iris-versicolor
0.96	0.03	Iris-virginica

Taux de VP = VP/(VP+FN)   Taux de FP = FP/(FP+VN)

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative :
  - Précision, Rappel, F-Mesure
  - exemple : Classification de fleurs d'iris

Classe réelles	Classes estimées C	$\overline{C}$
C	VP	FN
$\overline{C}$	FP	VN

	C1	$\overline{C1}$
C1	49	1
$\overline{C1}$	0	100

	C2	$\overline{C2}$
C2	47	3
$\overline{C2}$	3	97

	C3	$\overline{C3}$
C3	48	2
$\overline{C3}$	3	97

Arbre de décision

Precision	Recall	F-Measure	Class
1	0.98	0.99	Iris-setosa
0.94	0.94	0.94	Iris-versicolor
0.94	0.96	0.95	Iris-virginica

Précision = VP/(VP+FP)   Rappel = VP/(VP+FN) = Taux VP

F-Mesure = 2xPrécisionxRappel/(Précision+Rappel)

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

		Condition (as determined by "Gold standard")		
		Condition positive	Condition negative	
Test outcome	Test outcome positive	True positive	False positive (Type I error)	Precision = Σ True positive / Σ Test outcome positive
	Test outcome negative	False negative (Type II error)	True negative	Negative predictive value = Σ True negative / Σ Test outcome negative
		Sensitivity = Σ True positive / Σ Condition positive	Specificity = Σ True negative / Σ Condition negative	Accuracy = Σ True positive + Σ True negative / Σ Total population

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation quantitative

A

63	37
28	72

TPR = 0.63
FPR = 0.28

B

77	23
77	23

TPR = 0.77
FPR = 0.77

C

24	76
88	12

TPR = 0.24
FPR = 0.88

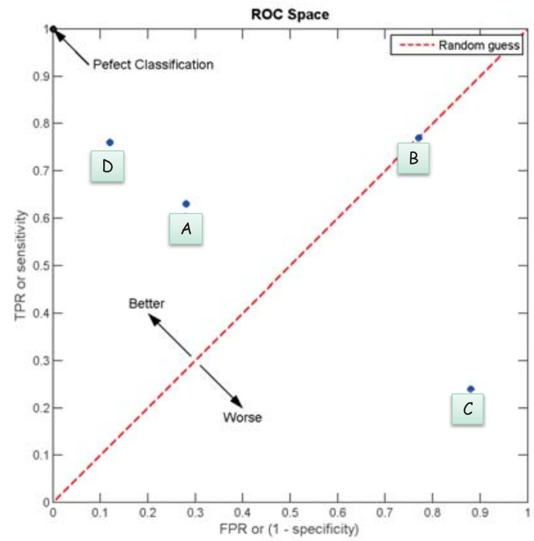
D

76	24
12	88

TPR = 0.76
FPR = 0.12

TP	FN
FP	TN

TPR= TP/(TP+FN)
FPR= FP/(FP+TN)



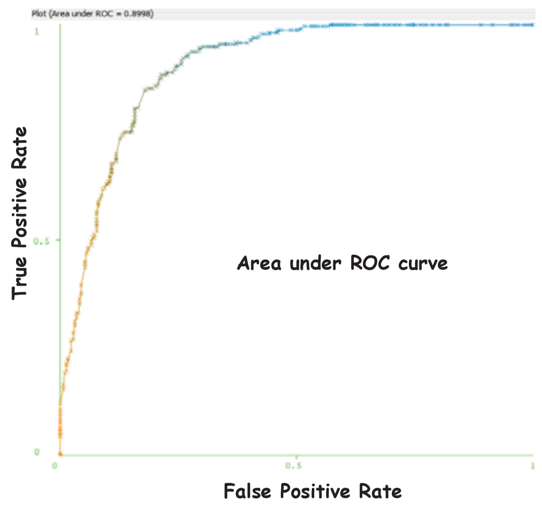
ROC : Receiver Operating Characteristic

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

Model			
Positive	Negative		
TP	FN	Positive	Real
FP	TN	Negative	

True Positive Rate = TP / (TP + FN)

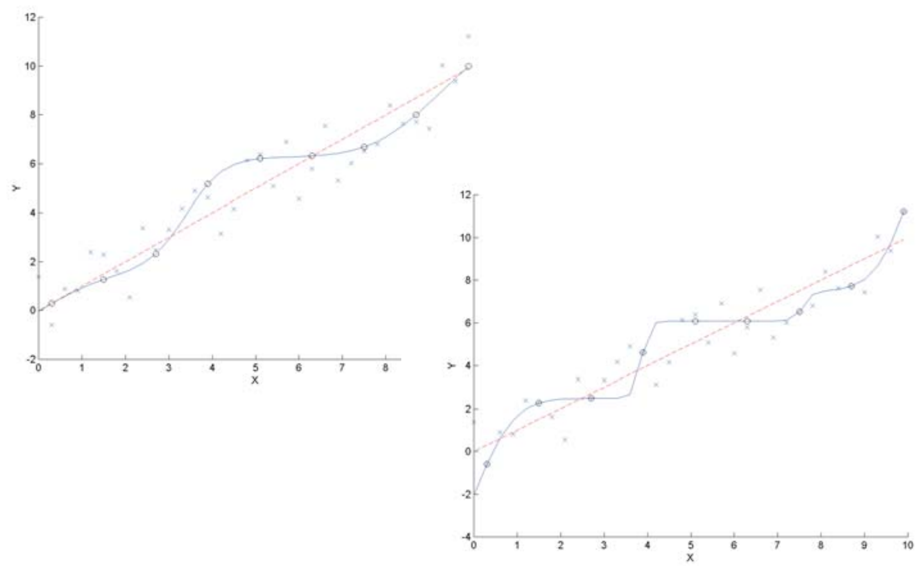
False Positive Rate = FP / (FP + TN)



ROC : Receiver Operating Characteristic

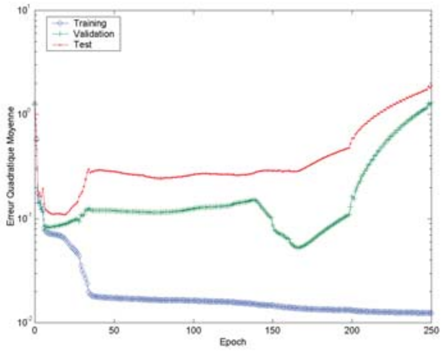
Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

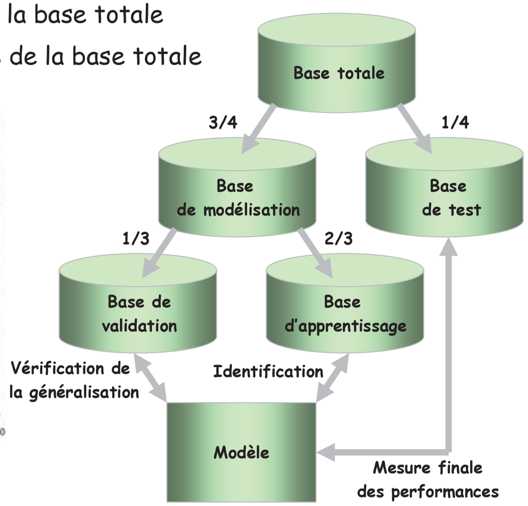
- Sur apprentissage - Overfitting



Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Évaluation par le test
  - Base de test 25% de la base totale
  - Base de validation 25% de la base totale
  - Base d'apprentissage 50% de la base totale

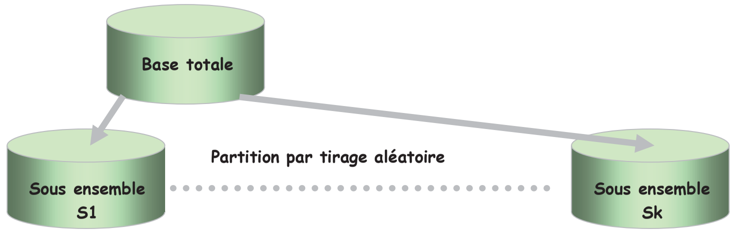




Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Validation croisée

Attention cette technique ne fournit pas de modèles mais permet d'évaluer les performances de la modélisation dans les cas où la base totale est petite



Pour tout i de 1 à k
- Appliquer une méthode d'identification sur (Base totale - Si)
- Calculer la performance du modèle sur Si

Fin

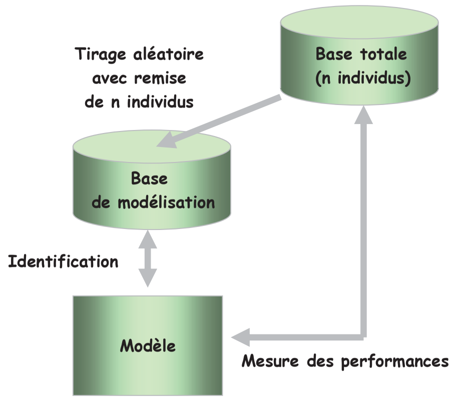
Agréger les performances individuelles pour établir une mesure globale

Si la taille des Si est de 1 individu on parle alors de « Leave one out »

Les étapes du processus de KDD
Étape 7 : Évaluer et valider les résultats

- Bootstrap

Attention cette technique ne fournit pas de modèles mais permet d'évaluer les performances de la modélisation dans les cas où la base totale est petite



L'expérience est réalisée un grand nombre de fois et les performances sont agrégées pour fournir une évaluation globale

Organisation du cours

Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

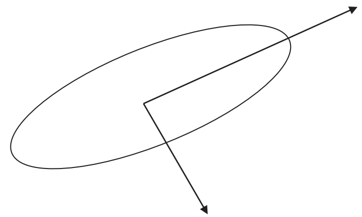
Au programme

- Introduction
- Pré traitement des données
- Quelques mesures et graphiques de statistiques descriptives
- Éléments du traitement du signal : transformée et série de Fourier
- Modèle linéaire : régression linéaire

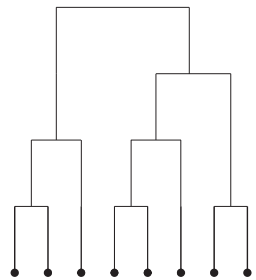
Introduction

L'analyse des données contient deux grands groupes de méthodes

Analyse Factorielle



Classification/Régression



L'espace des données

- Considérons l'espace de n points de dimension p suivant :

Caractères

$$X \;:\; \text{Individus}\;\begin{array}{c|ccccc|} & X^1 & \dots & X^j & \dots & X^p \\ \hline X_1 & x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_i & x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_n & x_n^1 & \dots & x_n^j & \dots & x_n^p \\ \hline \end{array}$$

$$X_i = [x_i^1, x_i^2, \dots, x_i^j, \dots, x_i^p]$$

$$X^j = [x_1^j, x_2^j, \dots, x_i^j, \dots, x_n^j]^T$$

Pré Traitement des données

- Données Centrées - Réduites

$$X^j = [x_1^j, x_2^j, \dots, x_i^j, \dots, x_n^j]^T$$

$$x_i^{j*} = \frac{x_i^j - \overline{X^j}}{s_{X^j}}$$

```
Fonctions Matlab : prestd, poststd
```

- Données Bornées [0,1]

$$x_i^{j*} = \frac{x_i^j - X_{min}^j}{X_{max}^j - X_{min}^j}$$

Données Bornées [a,b]

$$x_i^{j*} = \frac{x_i^j - \text{offset}}{\text{quotient}}$$

$$\text{quotient} = \left(X_{max}^j - X_{min}^j\right)/(b-a)$$

$$\text{offset} = X_{max}^j - \text{quotient} \times b$$

```
Fonctions Matlab : premnmx, postmnmx
```

Quelques mesures de statistiques descriptives

$X^j = [x_1^j, x_2^j, ..., x_i^j, ..., x_n^j]^T$

Mesures de bornage	Fonctions MATLAB
Maximum de $X^j = X^j_{max} = \max\{x_1^j, x_2^j, ..., x_i^j, ..., x_n^j\}$	`max`
Minimum de $X^j = X^j_{min} = \min\{x_1^j, x_2^j, ..., x_i^j, ..., x_n^j\}$	`min`

Quelques mesures de statistiques descriptives

$X^j = [x_1^j, x_2^j, ..., x_i^j, ..., x_n^j]^T$

Mesures de tendance centrale	Fonctions MATLAB
Moyenne arithmétique $= \overline{X^j} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$	`mean`
Moyenne géométrique $= \left[\prod_{i=1}^{n} x_i^j\right]^{\frac{1}{n}}$	`geomean`
Moyenne harmonique $= \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i^j}}$	`harmmean`
Médiane $M : F(M) = 0.5$	`median`

Quelques mesures de statistiques descriptives

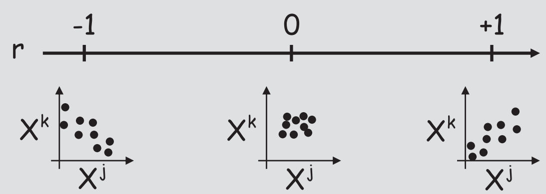
$X^j = [x_1^j, x_2^j, ..., x_i^j, ..., x_n^j]^T$

Mesures de dispersion	Fonctions MATLAB
Intervalle interquartile $= \lvert Q_3 - Q_1 \rvert$ $F(Q_1) = 0.25, F(Q_2) = 0.50, F(Q_3) = 0.75$	`iqr`
Étendue ou intervalle de variation $w = \lvert X^j_{max} - X^j_{min} \rvert$	`range`
Variance : $s_j^{\,2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^j - \overline{X^j}\right)^2$	`var`
Écart-type : $s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^j - \overline{X^j}\right)^2}$	`std`

Quelques mesures de statistiques descriptives

$X^j = [x_1^j, x_2^j, ..., x_i^j, ..., x_n^j]^T \qquad X^k = [x_1^k, x_2^k, ..., x_i^k, ..., x_n^k]^T$

Mesures de liaison entre deux caractères quantitatifs	Fonctions MATLAB
Covariance : $s_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^j - \overline{X^j}\right)\left(x_i^k - \overline{X^k}\right)$	`cov`
Coefficient de corrélation : $r_{jk} = \dfrac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^j - \overline{X^j}\right)\left(x_i^k - \overline{X^k}\right)}{s_j s_k}$	`corrcoef`



Quelques mesures de statistiques descriptives

Mesures de liaison entre caractères quantitatifs

Matrice de variance

$$V = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ & s_2^2 & & \\ & & \ddots & \vdots \\ & & & s_p^2 \end{pmatrix}$$

Matrice de corrélation

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ & 1 & & \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \qquad D_{1/s} = \begin{pmatrix} 1/s_1 & 0 & \cdots & 0 \\ 0 & 1/s_2 & 0 & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1/s_p \end{pmatrix}$$

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

Quelques mesures de statistiques descriptives

Tableau de contingence entre caractères qualitatifs

Exemple :

Individus x Caractères

Individus	Niveau hiérarchique x1: Cadre Supérieur	Niveau hiérarchique x1: Cadre	Niveau hiérarchique x1: Employé	Sexe x2: M	Sexe x2: F
1	1	0	0	1	0
2	0	1	0	1	0
3	1	0	0	0	1
4	1	0	0	1	0
5	0	0	1	0	1
6	0	0	1	0	1
7	0	1	0	1	0
8	0	0	1	1	0

Tableau de contingence $X1^T X2$

Niveau hiérarchique X1	Sexe X2: M	Sexe X2: F
Cadre Supérieur	2	1
Cadre	2	0
Employé	1	2

Commandes MATLAB

```
>> X1=[1,0,0;0,1,0;1,0,0;1,0,0;0,0,1;0,0,1;0,1,0;0,0,1];
>> X2=[1,0;1,0;0,1;1,0;0,1;0,1;1,0;1,0];
>> X1'*X2

ans =

     2     1
     2     0
     1     2
```


Quelques mesures de statistiques descriptives

Mesures de liaison entre deux caractères qualitatifs

Matrice de contingence : $\begin{bmatrix} n_{11} & n_{12} & . & n_{1j} & . & n_{1s} \\ n_{21} & n_{22} & . & . & . & . \\ . & . & . & . & . & . \\ n_{i1} & . & . & n_{ij} & . & n_{is} \\ . & . & . & . & . & . \\ n_{r1} & . & . & n_{rj} & . & n_{rs} \end{bmatrix}$

$$n_{i.} = \sum_{j=1}^{s} n_{ij} \qquad n_{.j} = \sum_{i=1}^{r} n_{ij} \qquad n = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{s} n_{.j}$$

Écart à l'indépendance : $D^2 = \displaystyle\sum_{i=1,j=1}^{r,s} \frac{\left(n_{ij} - \frac{n_{i.} \times n_{.j}}{n}\right)^2}{\frac{n_{i.} \times n_{.j}}{n}} = n\left[\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{n_{ij}^2}{n_{i.} \times n_{.j}}\right]$

$$0 \leq D^2 \leq n \times \inf(s-1, r-1)$$

Indépendance — Dépendance

Quelques mesures de statistiques descriptives

Mesure de liaison entre un caractère quantitatif Y et un caractère qualitatif X

Si X possède k catégories, on note :

$n_1, n_2, \ldots, n_k$ les effectifs observés

$\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_k$ les moyennes de Y pour chaque catégorie

Individus	Sexe X: H	Sexe X: F	Revenu Y
1	1	0	150
2	1	0	200
3	0	1	120
4	0	1	110
5	1	0	250
6	0	1	130

$n_1 = 3, n_2 = 3$
$\overline{y}_1 = 200, \overline{y}_2 = 120$
$\overline{y} = 160, s_y^2 = 2960$
$e^2 = 0.54$

Rapport de corrélation : $e^2 = \dfrac{\frac{1}{n}\sum_{i=1}^{k} n_i (\overline{y}_i - \overline{y})^2}{s_y^2}$

$$0 \leq e^2 \leq 1$$

Indépendance — Dépendance

Quelques mesures de statistiques descriptives

$X_i = [x_i^1, x_i^2, ..., x_i^j, ..., x_i^p]$ $\quad X_k = [x_k^1, x_k^2, ..., x_k^j, ..., x_k^p]$

Mesures de distance
Formule Générale : $d_{ik}^2 = d^2(X_i; X_k) = (X_i - X_k)^T M (X_i - X_k)$
Distance Euclidienne : $M = I = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}$
Distance standardisée : $M = D_{1/s^2} = \begin{bmatrix} 1/s_1^2 & & & 0 \\ & 1/s_2^2 & & \\ & & \ddots & \\ 0 & & & 1/s_p^2 \end{bmatrix}$
Distance de Mahalanobis : $M = V^{-1}$ $\quad V$ = matrice de covariance
Distance "City Block" : $d_{jk}^2 = \sum_{i=1}^{n} \lvert x_i^j - x_i^k \rvert$

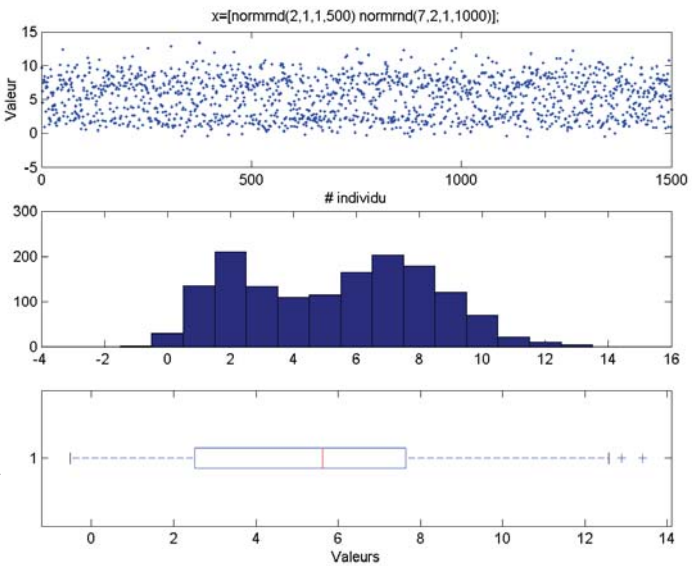
Quelques mesures de statistiques descriptives

$X^{cible} = [x_1^{cible}, x_2^{cible}, ..., x_i^{cible}, ..., x_n^{cible}]^T \quad X^{estimé} = [x_1^{estimé}, x_2^{estimé}, ..., x_i^{estimé}, ..., x_n^{estimé}]^T$

Mesures d'erreur
Erreur quadratique moyenne : $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^{cible} - x_i^{estimé}\right)^2$
Racine de l'erreur quadratique moyenne : $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^{cible} - x_i^{estimé}\right)^2}$
Erreur absolue moyenne : $MAE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert x_i^{cible} - x_i^{estimé}\right\rvert$
Erreur relative absolue moyenne : $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert \frac{x_i^{cible} - x_i^{estimé}}{x_i^{cible}}\right\rvert$

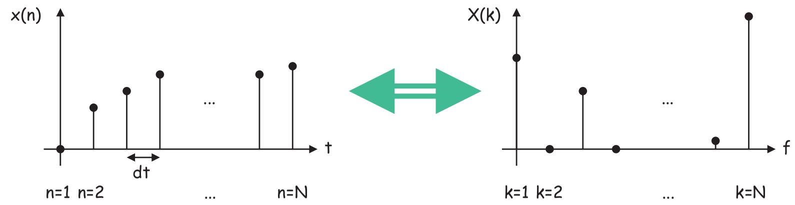
Quelques graphiques de statistiques descriptives

- Tracé de courbes
  `Fonction MATLAB : plot`
- Histogramme
  `Fonction MATLAB : hist`
- Boxplot
  `Fonction MATLAB : boxplot`



Éléments du traitement du signal

- Transformée de Fourier discrète : permet de convertir un signal du domaine temporel dans le domaine fréquentiel



Transformée

$$X_k = \sum_{n=1}^{N} x_n e^{-i2\pi(n-1)(k-1)/N}$$

pour $k = 1,2,..., N$

`Fonction Matlab : fft`

Transformée inverse

$$x_n = \frac{1}{N}\sum_{k=1}^{N} X_k e^{-i2\pi(n-1)(k-1)/N}$$

pour $n = 1,2,..., N$

`Fonction Matlab : ifft`

Éléments du traitement du signal

- Série de Fourier

$$x_n = a_0 + \sum_{k=1}^{N/2} a_k \cos(2\pi k t_n/(N.dt)) + b_k \sin(2\pi k t_n/(N.dt))$$

$$a_0 = X_1/N$$
$$a_k = 2 \times \Re(X_{k+1})/N$$
$$b_k = -2 \times \Im(X_{k+1})/N$$

- exemple : reconstruction d'un signal carré (N=100)

Régression linéaire

$$X^j = [x_1^j, x_2^j, \dots, x_i^j, \dots, x_n^j]^T$$

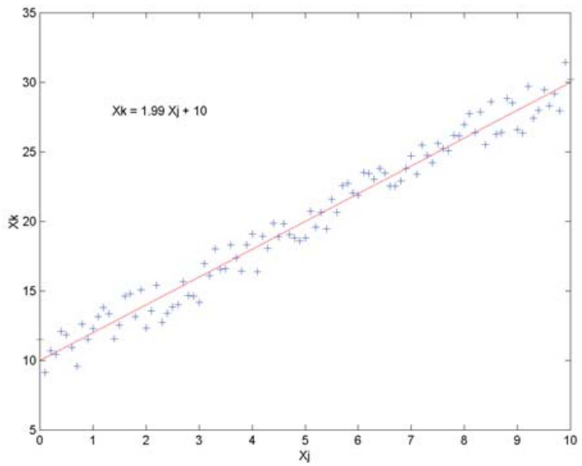
$$X^k = [x_1^k, x_2^k, \dots, x_i^k, \dots, x_n^k]^T$$

$$X^k = \hat{a} X^j + \hat{b}$$

$$\hat{b} = \frac{\sum_{i=1}^{N} (x_i^j - \overline{X^j})(x_i^k - \overline{X^k})}{\sum_{i=1}^{N} (x_i^j - \overline{X^j})^2}$$

$$\hat{a} = \overline{X^k} - \hat{b} \times \overline{X^j}$$

`Fonction Matlab : regress`

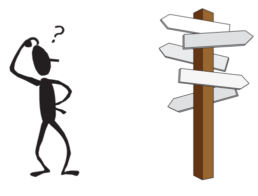


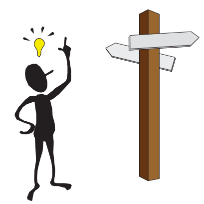
Organisation du cours

Analyse Factorielle

- Terme générique pour parler :
  - de l'analyse en composantes principales
  - et de l'analyse factorielle proprement dite

- Objectif :
  - Réduire les dimensions de l'espace des données
  - Création de variables « latentes » : combinaisons linéaires des variables réelles





Analyse en Composantes Principales

- X étant un tableau de p variables numériques décrivant n individus $e_1, e_2, \cdots, e_n$
- L'ACP permet de rechercher une représentation des n individus dans un sous espace de l'espace initial.
- Autrement dit, l'objectif est de définir k nouvelles variables, combinaisons des p variables de l'espace initial, qui feraient perdre le « moins d'information possible ».
- *Ces k variables sont appelées « composantes principales », et les axes qu'elles déterminent « axes principaux ».*

Analyse en Composantes Principales

- Notations

Vecteurs individus dans l'espace initial : $e_1, e_2, \cdots, e_n$

Vecteurs individus dans l'espace de projection : $f_1, f_2, \cdots, f_n$

Vecteurs variables : $x_1, x_2, \cdots, x_n$

Poids associés à chaque individus : $p_1, p_2, \cdots, p_n$

Centre de gravité du nuage de points initial : g

Inertie totale du nuage de points : I_g

Remarque : on supposera les variables centrées

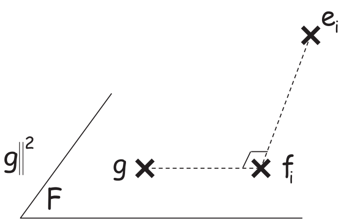
Analyse en Composantes Principales

- On cherche un sous-espace de l'espace initial tel que :

$\sum_{i=1}^{n} p_i \|e_i - f_i\|^2$ soit minimale

Or $\|e_i - g\|^2 = \|e_i - f_i\|^2 + \|f_i - g\|^2$

L'objectif est donc de maximiser : $\sum_{i=1}^{n} p_i \|f_i - g\|^2$



En d'autres termes,
l'objectif est de maximiser la dispersion (ou inertie totale) des f_i

Soit l'inertie du nuage initial : $I_g^{init} = \sum_{i=1}^{n} p_i \, e_i^t \, M \, e_i = \mathrm{Trace}(MV)$

avec V matrice de variance et M une métrique

Il faut donc maximiser l'inertie du nuage projeté : $I_g^{proj} = \mathrm{Trace}(MVP)$

avec P : matrice de projection

Analyse en Composantes Principales

- Axes principaux

On cherche une droite maximisant l'inertie du nuage projeté sur cette droite. Soit a un vecteur de cette droite

On a : $P = a(a^t M a)^{-1} a^t M$

Le but est de trouver a maximisant Trace(VMP)

On montre que a est le vecteur propre de la matrice VM associé à la plus grande valeur propre.

Le sous-espace des projetés de dimension k est engendré par les k vecteurs propres de VM associés aux k plus grandes valeurs propres

Analyse en Composantes Principales

- Facteurs principaux

  A l'axe principal a_i est associé le facteur principal u_i

  $$u_i = Ma_i$$

- Composantes principales

  Ce sont les variables c_i définies par les facteurs principaux, combinaisons linéaires des $x_1, \ldots, x_p$

  $$c_i = Xu_i$$

La variance de c_i est égale à la valeur propre λ_i correspondante.

Ces composantes sont orthogonales et donc non corrélées entre elles.

Analyse en Composantes Principales

- Interprétation : exemple sur la base des iris

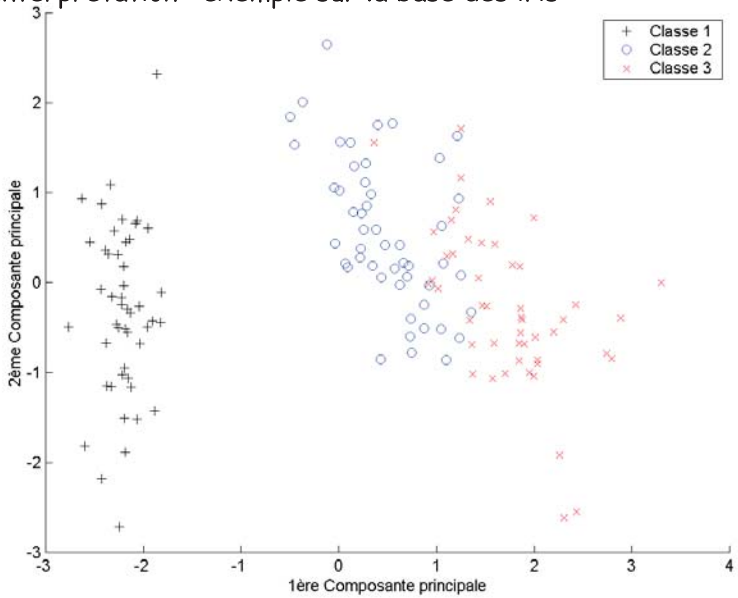
Vecteurs propres

	CP1	CP2	CP3	CP4
L sépale	0.5224	-0.3723	0.7210	-0.2620
l sépale	-0.2634	-0.9256	-0.2420	0.1241
L pétale	0.5813	-0.0211	-0.1409	0.8012
l pétale	0.5656	-0.0654	-0.6338	-0.5235

	valeurs propres	% d'inertie	% cumulé
CP1	2.9108	72.77%	72.77%
CP2	0.9212	23.03%	95.80%
CP3	0.1474	03.69%	99.49%
CP4	0.0206	00.51%	100.00%

Analyse en Composantes Principales

- Interprétation : exemple sur la base des iris



Organisation du cours

Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

Classification

- Objectif : identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs

Π est la population,

D est l'ensemble des descriptions,

C est l'ensemble des classes

$X : \Pi \to D$

fonction qui associe une description à tout élément de la population

$Y : \Pi \to C$

fonction qui associe une classe à tout élément de la population

$F : D \to C$

fonction de classement recherchée

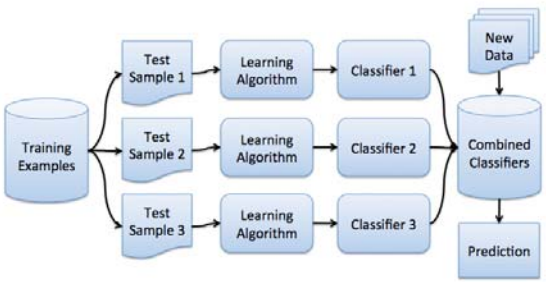
Classification

- Le but est de rechercher une fonction de classement F telle que $F \circ X$ soit une bonne approximation de Y.

$$\begin{tikzcd} \Pi \arrow[rr, "X"] \arrow[dr, "Y"'] & & D \arrow[dl, "F"] \\ & C & \end{tikzcd}$$

Méthodes de classification

- 4 grandes catégories de méthodes de classification
  - Distance : exemple k plus proches voisins
  - Probabiliste : exemple Naives Bayes
  - Arbres de décision: exemple ID3
  - Distance et optimisation : exemple Machines à vecteurs de support
- Il existe un nombre important de méthodes de classification basées sur ces 4 catégories
- Ensemble Learning



Algorithme des k plus proches voisins

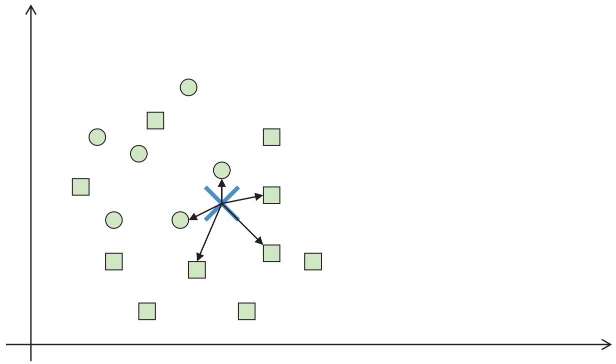
- On considère l'espace de n points de dimension p suivant :

$$X \ : \quad \begin{array}{c|ccccc|} & X^1 & \dots & X^j & \dots & X^p \\ \hline X_1 & x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_i & x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_n & x_n^1 & \dots & x_n^j & \dots & x_n^p \\ \hline \end{array}$$

- A chaque point est associée une classe connue à l'avance
- Soit $X_T = [x_T^1, x_T^2, \dots, x_T^j, \dots, x_T^p]$ un point que l'on souhaite classifier
- On calcule toutes les distances entre le point X_T et les n points de l'espace
- On conserve les k points les plus proches de X_T
- La classe majoritaire dans l'ensemble de ces k points est attribuée à X_T

Algorithme des k plus proches voisins

- Exemple :



Si k=3 le nouveau point sera associé à la classe des cercles
Si k=5 le nouveau point sera associé à la classe des carrés

Classifieur naïf de Bayes

- On note :

$x = [x^1, x^2, ..., x^i, ..., x^p]$ un vecteur de descripteurs
$C = \{c_1, \cdots, c_q\}$ l'ensemble des classes possibles
S : ensemble fini de couples de la forme (x, c_k)
n : nombre d'observations de S
n_k : nombre d'éléments de S appartenant à la classe c_k

- La règle de classification de Bayes recommande de classer le vecteur x dans la classe c_k pour laquelle $P(c_k/x)$ est maximal.

Ce qui revient à maximiser : $\dfrac{p(x/c_k)p(c_k)}{p(x)}$

soit encore $p(x/c_k)p(c_k)$, $p(x)$ ne dépendant pas de c_k

Pour pouvoir appliquer la règle de Bayes il faut donc pouvoir estimer : $p(x/c_k)$ et $p(c_k)$

Classifieur naïf de Bayes

- On estime $p(c_k)$ par $\hat{p}(c_k) = \dfrac{n_k}{n}$
- et $p(x/c_k)$ par : $\hat{p}(x/c_k) = \prod_{i=1}^{p} p(x^i/c_k)$

Ce qui revient à considérer les attributs comme indépendants les uns des autres

- La règle de classification de Bayes devient alors : classer le vecteur x dans la classe c_k qui maximise :

$$\prod_{i=1}^{p} p(x^i/c_k)p(c_k)$$

- Lorsque les attributs prennent leurs valeurs dans un même ensemble, on peut faire l'hypothèse que dans chaque classe, la probabilité d'observer une valeur quelconque d'un attribut est en fait indépendante de l'attribut.

Pour toute classe c, pour toute valeur v et tous les indices i et j :

$$p(x^i = v/c) = p(x^j = v/c)$$

Classifieur naïf de Bayes

- Exemple

$S = S_1 \cup S_2$
$S_1 = \{01100, 11001, 10110, 10101, 10010\}$
$S_2 = \{01010, 11111, 11010, 11101, 10101\}$
$C = \{c_1, c_2\}$

- On demande de classer x=00111

$$\hat{p}(c_1) = \frac{n_1}{n} = \frac{5}{10} = \frac{1}{2}$$
$$\hat{p}(c_2) = \frac{n_2}{n} = \frac{5}{10} = \frac{1}{2}$$

Classifieur naïf de Bayes

- On suppose les attributs indépendants

	c_1	c_2		c_1	c_2
$\hat{p}(x_1=0/c_j)$	1/5	1/5	$\hat{p}(x_1=1/c_j)$	4/5	4/5
$\hat{p}(x_2=0/c_j)$	3/5	1/5	$\hat{p}(x_2=1/c_j)$	2/5	4/5
$\hat{p}(x_3=0/c_j)$	2/5	2/5	$\hat{p}(x_3=1/c_j)$	3/5	3/5
$\hat{p}(x_4=0/c_j)$	3/5	2/5	$\hat{p}(x_4=1/c_j)$	2/5	3/5
$\hat{p}(x_5=0/c_j)$	3/5	2/5	$\hat{p}(x_5=1/c_j)$	2/5	3/5

$$\hat{p}(00111/c_1)=\frac{1}{5}\times\frac{3}{5}\times\frac{3}{5}\times\frac{2}{5}\times\frac{2}{5}=\frac{36}{5^5}=11.5\,10^{-3}$$

$$\hat{p}(00111/c_2)=\frac{1}{5}\times\frac{1}{5}\times\frac{3}{5}\times\frac{3}{5}\times\frac{3}{5}=\frac{27}{5^5}=8.6\,10^{-3}$$

- On classera donc x=00111 en classe c_1

Classifieur naïf de Bayes

- Si on fait en plus l'hypothèse que les attributs prennent dans chaque classe les mêmes valeurs avec les mêmes probabilités

$$\hat{p}(x_i=0/c_1)=\frac{12}{25};\hat{p}(x_i=1/c_1)=\frac{13}{25}$$

$$\hat{p}(x_i=0/c_2)=\frac{8}{25};\hat{p}(x_i=1/c_2)=\frac{17}{25}$$

$$\hat{p}(00111/c_1)=\frac{12^2\times 13^3}{25^5}=32.4\,10^{-3}$$

$$\hat{p}(00111/c_2)=\frac{8^2\times 17^3}{25^5}=32.2\,10^{-3}$$

- On classera donc x=00111 en classe c_1

Exercice 1 classifieur naïf de Bayes

- Objectif : Classifier des phrases selon leur thème : la radio ou la télévision
- Échantillon :
  - Classe télévision :
    - Le programme TV n'est pas intéressant.
    - La TV m'ennuie.
    - Les enfants aiment la TV.
    - On reçoit la TV par onde radio.
  - Classe radio :
    - Il est intéressant d'écouter la radio.
    - Sur les ondes, les programmes pour enfants sont rares.
    - Les enfants vont écouter la radio; c'est rare.
- Vocabulaire : V={TV, programme, intéressant, enfants, radio, onde, écouter, rare}

- En utilisant un classifieur de Bayes, à quel thème serait associée la phrase :
  « J'ai vu la radio de mes poumons à la TV »

Exercice 1 classifieur naïf de Bayes

- Codage des informations

- Échantillon :

	TV	programme	intéressant	enfants	radio	onde	écouter	rare	
	x1	x2	x3	x4	x5	x6	x7	x8	Classe
1									
2									
3									
4									
5									
6									
7									

- Phrase à classer :
  x=x1 x2 x3 x4 x5 x6 x7 x8

Exercice 1 classifieur naïf de Bayes

$\hat{p}(c_1) = -$ $\hat{p}(c_2) = -$

$\hat{p}(x_i = 1/c_1) = -$ $\hat{p}(x_i = 1/c_2) = -$

$\hat{p}(x_i = 0/c_1) = -$ $\hat{p}(x_i = 0/c_2) = -$

$\hat{p}(c_1) \times \hat{p}(10001000/c_1) =$

$\hat{p}(c_2) \times \hat{p}(10001000/c_2) =$

Exercice 1 classifieur naïf de Bayes

$\hat{p}(c_1) = \frac{4}{7}$ $\hat{p}(c_2) = \frac{3}{7}$

$\hat{p}(x_1 = 0/c_1) = \frac{0}{4}$ $\hat{p}(x_1 = 0/c_2) = \frac{3}{3}$ $\hat{p}(x_1 = 1/c_1) = \frac{4}{4}$ $\hat{p}(x_1 = 1/c_2) = \frac{0}{3}$

$\hat{p}(x_2 = 0/c_1) = \frac{3}{4}$ $\hat{p}(x_2 = 0/c_2) = \frac{2}{3}$ $\hat{p}(x_2 = 1/c_1) = \frac{1}{4}$ $\hat{p}(x_2 = 1/c_2) = \frac{1}{3}$

$\hat{p}(x_3 = 0/c_1) = \frac{3}{4}$ $\hat{p}(x_3 = 0/c_2) = \frac{2}{3}$ $\hat{p}(x_3 = 1/c_1) = \frac{1}{4}$ $\hat{p}(x_3 = 1/c_2) = \frac{1}{3}$

$\hat{p}(x_4 = 0/c_1) = \frac{3}{4}$ $\hat{p}(x_4 = 0/c_2) = \frac{1}{3}$ $\hat{p}(x_4 = 1/c_1) = \frac{1}{4}$ $\hat{p}(x_4 = 1/c_2) = \frac{2}{3}$

$\hat{p}(x_5 = 0/c_1) = \frac{3}{4}$ $\hat{p}(x_5 = 0/c_2) = \frac{1}{3}$ $\hat{p}(x_5 = 1/c_1) = \frac{1}{4}$ $\hat{p}(x_5 = 1/c_2) = \frac{2}{3}$

$\hat{p}(x_6 = 0/c_1) = \frac{3}{4}$ $\hat{p}(x_6 = 0/c_2) = \frac{2}{3}$ $\hat{p}(x_6 = 1/c_1) = \frac{1}{4}$ $\hat{p}(x_6 = 1/c_2) = \frac{1}{3}$

$\hat{p}(x_7 = 0/c_1) = \frac{4}{4}$ $\hat{p}(x_7 = 0/c_2) = \frac{1}{3}$ $\hat{p}(x_7 = 1/c_1) = \frac{0}{4}$ $\hat{p}(x_7 = 1/c_2) = \frac{2}{3}$

$\hat{p}(x_8 = 0/c_1) = \frac{4}{4}$ $\hat{p}(x_8 = 0/c_2) = \frac{1}{3}$ $\hat{p}(x_8 = 1/c_1) = \frac{0}{4}$ $\hat{p}(x_8 = 1/c_2) = \frac{2}{3}$

$$\hat{p}(c_1) \times \hat{p}(10001000/c_1) = \frac{4}{7} \times \frac{4}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{4}{4} \times \frac{4}{4} = \frac{20736}{458752} = 0{,}0452$$

$$\hat{p}(c_2) \times \hat{p}(10001000/c_2) = \frac{3}{7} \times \frac{0}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{0}{45927} = 0$$

Exercice 1 classifieur naïf de Bayes

$$\hat{p}(x_i/c_k) = \frac{1 + \textit{Nombre d'occurences de } x_i \textit{ de } V \textit{ dans l'ensemble des textes de classe } k}{Card(V) + \textit{Nombre total d'occurences de mots dans l'ensemble des textes de classe } k}$$

$\hat{p}(c_1) = \frac{4}{7}$ $\hat{p}(c_2) = \frac{3}{7}$

$\hat{p}(x_1 = 1/c_1) = \frac{1+4}{8+9}$ $\hat{p}(x_1 = 1/c_2) = \frac{1+0}{8+11}$

$\hat{p}(x_2 = 1/c_1) = \frac{1+1}{8+9}$ $\hat{p}(x_2 = 1/c_2) = \frac{1+1}{8+11}$

$\hat{p}(x_3 = 1/c_1) = \frac{1+1}{8+9}$ $\hat{p}(x_3 = 1/c_2) = \frac{1+1}{8+11}$

$\hat{p}(x_4 = 1/c_1) = \frac{1+1}{8+9}$ $\hat{p}(x_4 = 1/c_2) = \frac{1+2}{8+11}$

$\hat{p}(x_5 = 1/c_1) = \frac{1+1}{8+9}$ $\hat{p}(x_5 = 1/c_2) = \frac{1+2}{8+11}$

$\hat{p}(x_6 = 1/c_1) = \frac{1+1}{8+9}$ $\hat{p}(x_6 = 1/c_2) = \frac{1+1}{8+11}$

$\hat{p}(x_7 = 1/c_1) = \frac{1+0}{8+9}$ $\hat{p}(x_7 = 1/c_2) = \frac{1+2}{8+11}$

$\hat{p}(x_8 = 1/c_1) = \frac{1+0}{8+9}$ $\hat{p}(x_8 = 1/c_2) = \frac{1+2}{8+11}$

$$\hat{p}(c_1) \times \hat{p}(10001000/c_1) = \frac{4}{7} \times \frac{5}{17} \times \frac{2}{17} = \frac{40}{2023} = 0{,}020$$

$$\hat{p}(c_2) \times \hat{p}(10001000/c_2) = \frac{3}{7} \times \frac{1}{19} \times \frac{3}{19} = \frac{9}{2527} = 0{,}0035$$

Exercice 2 classifieur naïf de Bayes

Diagnostic de la grippe

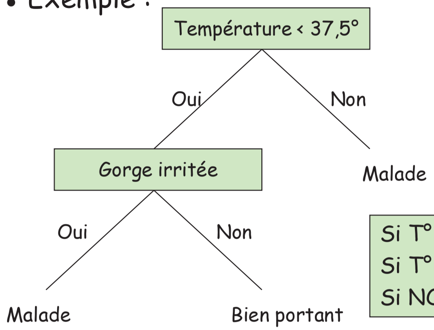
#	Frissons	Nez qui coule	Maux de tête	Fièvre	Grippe
1	Oui	Non	Moyen	Oui	Non
2	Oui	Oui	Non	Non	Oui
3	Oui	Non	Fort	Oui	Oui
4	Non	Oui	Moyen	Oui	Oui
5	Non	Non	Non	Non	Non
6	Non	Oui	Fort	Oui	Oui
7	Non	Oui	Fort	Non	Non
8	Oui	Oui	Moyen	Oui	Oui
9	Oui	Non	Non	Oui	?

Arbres de Décision

- Un arbre de décision est une représentation graphique d'une procédure de classification

- Un arbre de décision peut être traduit sous forme de règles de décision

- Exemple :



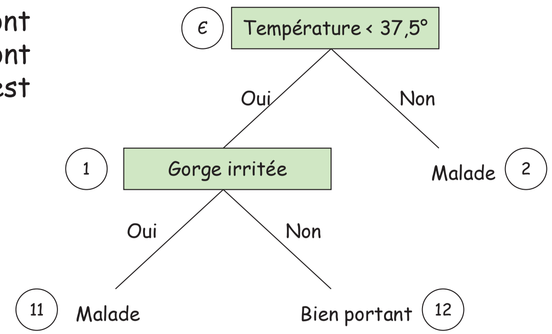
Si T° < 37,5 ET Gorge irritée ALORS Malade
Si T° < 37,5 ET NON(Gorge irritée) ALORS Bien portant
Si NON(T° < 37,5) ALORS Malade

Arbres de Décision

- Un arbre de décision est un arbre au sens informatique du terme.
  - Chaque nœud interne test un attribut
  - Chaque branche correspond à une valeur d'un attribut
  - Chaque nœud feuille est une classe

Les nœuds de l'arbre sont repérés par des positions qui sont des mots de {1, … , p}*, où p est l'arité maximale des nœuds.

On note ϵ le mot vide.

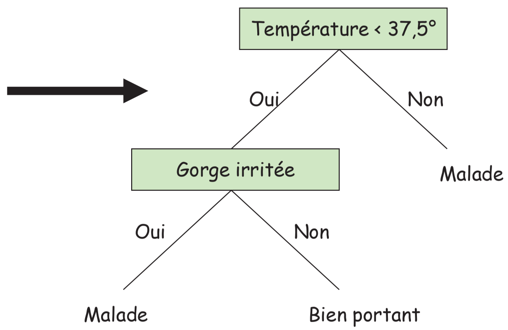


Arbres de Décision

- Comment généré automatiquement un arbre à partir de données ?

- Exemple :

Patients	T°	Gorge	Malade
Dupond	37,2	Normale	Non
Durand	38,5	Normale	Oui
⋮	⋮	⋮	⋮
Martin	37,2	Irritée	Oui



Arbres de Décision

- Notation

S : échantillon

$\{1, \cdots, c\}$: ensemble de classes

t : arbre de décision

p : position dans l'arbre

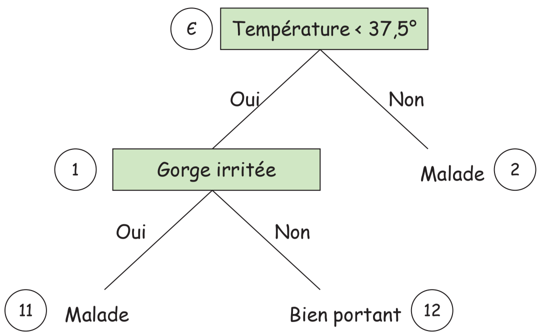
$N(p)$: cardinal de l'ensemble des exemples associé à p
$N(k/p)$: cardinal de l'ensemble des exemples associé à p et de classe k
$P(k/p) = N(k/p)/N(p)$: proportion d'éléments de classe k à la position p

Arbres de Décision

- Exemple

	Gorge irritée	Gorge non irritée
T° < 37,5	(6 S, 37 M)	(91 S, 1 M)
T° ≥ 37,5	(2 S, 21 M)	(1 S, 41 M)

M : Malade
S : Bien portant



$N(11) = 43;$
$N(S/11) = 6;$
$N(M/11) = 37;$
$P(S/11) = N(S/11)/N(11) = 6/43;$
$P(M/11) = N(M/11)/N(11) = 37/43$

Arbres de Décision

- Exemple introductif à l'apprentissage automatique d'arbre de décision
- Données banque

	M	A	R	E	I
1	moyen	moyen	village	oui	oui
2	élevé	moyen	bourg	non	non
3	faible	âgé	bourg	non	non
4	faible	moyen	bourg	oui	oui
5	moyen	jeune	ville	oui	oui
6	élevé	âgé	ville	oui	non
7	moyen	âgé	ville	oui	non
8	faible	moyen	village	non	non

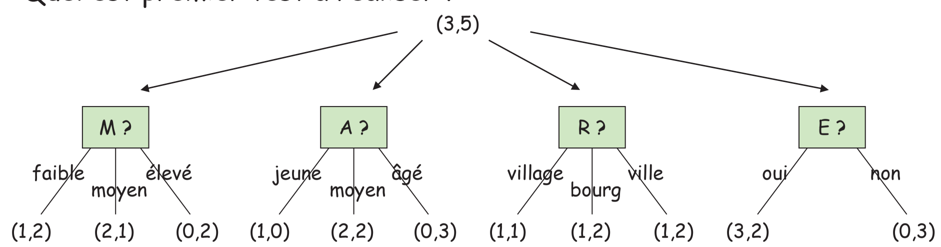
Moyenne du solde du compte :
$M = \{élevé, moyen, faible\}$
Tranche d'âge du client :
$A = \{jeune, moyen, âgé\}$
Type de localité de résidence du client :
$R = \{village, bourg, ville\}$
Niveau d'études supérieures du client :
$E = \{oui, non\}$
Consultatio n du compte par Internet :
$I = \{oui, non\}$

- On souhaite construire un arbre de décision qui soit capable de déterminer si un client consultera son compte par Internet en fonction des attributs : M, A, R et E

Arbres de Décision

- Racine de l'arbre (pas de test)
  Etiquette (3,5) correspondant à : (N(ϵ,oui), N(ϵ,non))

- Quel est premier test à réaliser ?



- Intuitivement :
  - Le test sur R n'est pas discriminatoire
  - Le test sur A est interessant sur les branches jeune et âgé

Arbres de Décision

- Quelles fonctions permettraient de répresenter ces intuitions ?

- Fonctions qui seraient :
  - Minimum lorsque le nœud est pur (tous les exemples sont dans une même classe
  - et Maximum lorsque les exemples sont équirépartis.

- Exemples de fonctions possédant ces propriétés :

  - Entropie : $Entropie(p) = -\sum_{k=1}^{c} P(k/p) \times \log(P(k/p))$

  - Fonction de Gini : $Gini(p) = 1 - \sum_{k=1}^{c} P(k/p)^2 = 2 \times \sum_{k<k'}^{c} P(k/p) \times P(k'/p)$

Arbres de Décision

- Quelles fonction permettrait choisir un test ?

  - Fonction gain : $Gain(p, test) = i(p) - \sum_{j=1}^{n} P_j \times i(p_j)$

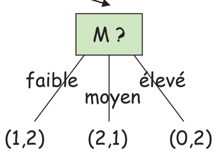
    p : position
    test : test d'arité n
    P_j : proportion d'éléments de S à la position p qui vont en position p_j
    $i(p)$: Entropie(p) ou Gini(p)

  - Le test choisi est celui qui possède le gain le plus grand

Arbres de Décision

- Exemple traité avec l'entropie :

(3,5) $\quad \text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



$\text{Entropie}(1) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$

$\text{Entropie}(2) = -\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right) \approx 0{,}918$

$\text{Entropie}(3) = -\frac{0}{2}\log\left(\frac{0}{2}\right) - \frac{2}{2}\log\left(\frac{2}{2}\right) = 0$

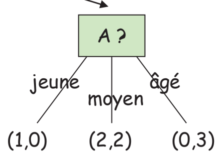
$Gain(\in, M) = \text{Entropie}(\in) - \left(\frac{3}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2) + \frac{2}{8}\text{Entropie}(3)\right)$

$= \text{Entropie}(\in) - 0{,}688$

Arbres de Décision

- Exemple traité avec l'entropie :

(3,5) $\quad \text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



$\text{Entropie}(1) = -\frac{1}{1}\log\left(\frac{1}{1}\right) - \frac{0}{1}\log\left(\frac{0}{1}\right) = 0$

$\text{Entropie}(2) = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$

$\text{Entropie}(3) = -\frac{0}{3}\log\left(\frac{0}{3}\right) - \frac{3}{3}\log\left(\frac{3}{3}\right) = 0$

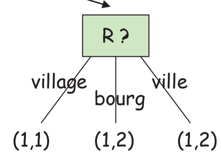
$Gain(\in, A) = \text{Entropie}(\in) - \left(\frac{1}{8}\text{Entropie}(1) + \frac{4}{8}\text{Entropie}(2) + \frac{3}{8}\text{Entropie}(3)\right)$

$= \text{Entropie}(\in) - 0{,}5$

Arbres de Décision

- Exemple traité avec l'entropie :

(3,5) $\quad \text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



$\text{Entropie}(1) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$

$\text{Entropie}(2) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$

$\text{Entropie}(3) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$

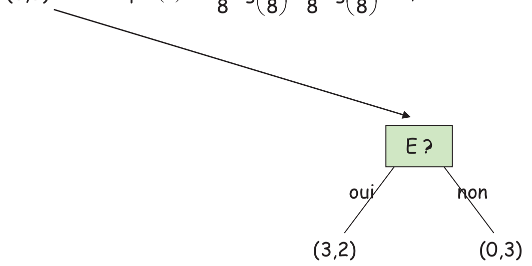
$Gain(\in, R) = \text{Entropie}(\in) - \left(\frac{2}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2) + \frac{3}{8}\text{Entropie}(3)\right)$

$= \text{Entropie}(\in) - 0{,}938$

Arbres de Décision

- Exemple traité avec l'entropie :

(3,5) $\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



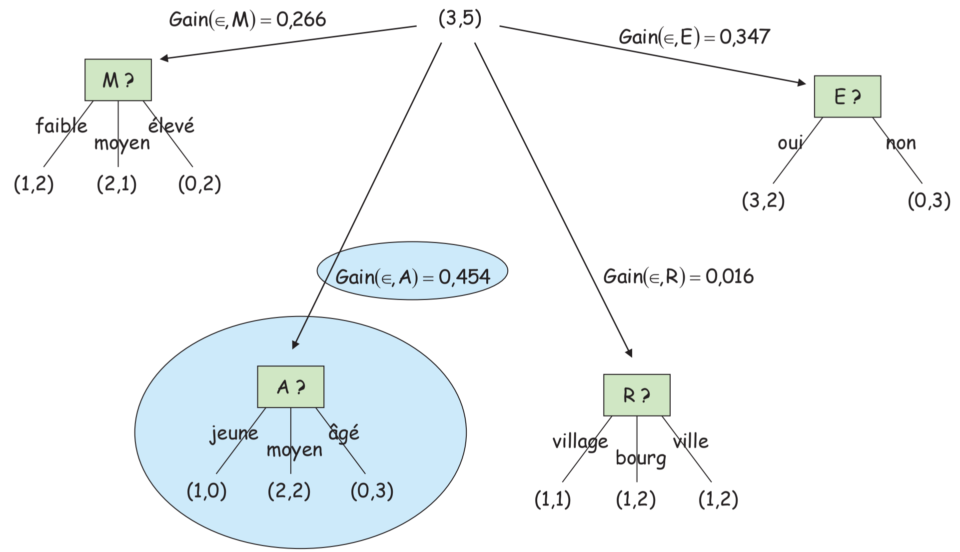
$\text{Entropie}(1) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) \approx 0.97$

$\text{Entropie}(2) = -\frac{0}{3}\log\left(\frac{0}{3}\right) - \frac{3}{3}\log\left(\frac{3}{3}\right) = 0$

$\text{Gain}(\in, E) = \text{Entropie}(\in) - \left(\frac{5}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2)\right) = \text{Entropie}(\in) - 0{,}607$

Arbres de Décision

- Exemple traité avec l'entropie :



Arbres de Décision

- Algorithme générique de construction d'un arbre

entrée : langage de description; échantillon S
début
Initialiser à l'arbre vide; la racine est le nœud courant
répéter
Décider si le nœud courant est terminal
si le nœud est terminal **alors**
Affecter une classe
sinon
Sélectionner un test et créer le sous-arbre
finsi
Passer au nœud suivant non exploré si il en existe
jusqu'à obtenir un arbre de décision
fin

Arbres de Décision

- Algorithme CART (Breiman 1984) - Arbres de décision binaires

Un nœud p est terminal si :

$Gini(p) \le i_0$ ou $N(p) \le n_0$, i_0 et n_0 sont des paramètres à fixer

On choisit le test qui maximise :

$$Gain(p, test) = Gini(p) - (P_{gauche} \times Gini(p_1) + P_{droite} \times Gini(p_2))$$

On attribut la classe majoritaire à une feuille

Cet algorithme de construction est suivi d'une phase d'élagage de l'arbre qui ne sera pas traitée dans le cadre de ce cours

Arbres de Décision

- Algorithme ID3

Un nœud p est terminal si : tous les éléments associés à ce nœud sont dans une même classe ou si aucun test n'a pu être séléctionner

On choisit le test qui maximise :

$$Gain(p, test) = Entropie(p) - \sum_{j=1}^{n} P_j \times Entropie(p_j)$$

On attribut la classe majoritaire à une feuille

Cet algorithme de construction est suivi d'une phase d'élagage de l'arbre qui ne sera pas traitée dans le cadre de ce cours

Arbres de Décision

- Algorithme C4.5 (Quilan 1993) (initialement ID3)

Un nœud p est terminal si : tous les éléments associés à ce nœud sont dans une même classe ou si aucun test n'a pu être séléctionner

On choisit le test qui maximise :

$$GainRatio(p, test) = Gain(p, test) / SplitInfo(p, test)$$

$$Gain(p, test) = Entropie(p) - \sum_{j=1}^{n} P_j \times Entropie(p_j)$$

$$SplitInfo(p, test) = -\sum_{j=1}^{n} P'(j/p) \times \log(P'(j/p))$$

$P'(j/p)$: proportion des éléments de p prenant la jème valeur de test

On attribut la classe majoritaire à une feuille

Cet algorithme de construction est suivi d'une phase d'élagage de l'arbre qui ne sera pas traitée dans le cadre de ce cours

Arbres de Décision : Exercice

- Exemple introductif à l'apprentissage automatique d'arbre de décision
- Données banque

	M	A	R	E	I
1	moyen	moyen	village	oui	oui
2	élevé	moyen	bourg	non	non
3	faible	âgé	bourg	non	non
4	faible	moyen	bourg	oui	oui
5	moyen	jeune	ville	oui	oui
6	élevé	âgé	ville	oui	non
7	moyen	âgé	ville	oui	non
8	faible	moyen	village	non	non

Moyenne du solde du compte :

$M = \{élevé, moyen, faible\}$

Tranche d'âge du client :

$A = \{jeune, moyen, âgé\}$

Type de localité de résidence du client :

$R = \{village, bourg, ville\}$

Niveau d'études supérieures du client :

$E = \{oui, non\}$

Consultation du compte par Internet :

$I = \{oui, non\}$

- On souhaite construire un arbre de décision qui soit capable de déterminer si un client consultera son compte par Internet en fonction des attributs : M, A, R et E

Arbres de Décision : Exercice

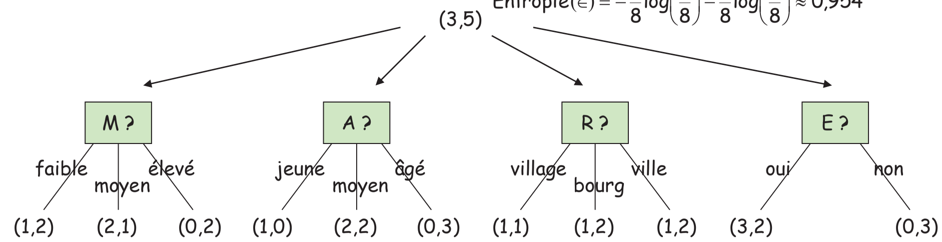

```
Fonction ID3 (I,O,T)
I ensemble des attributs d 'entrée
O attribut de sortie
T ensemble des individus d 'apprentissage

Si (T est vide)
        Renvoyer Erreur

Si (tous les individus de T appartiennent à la même classe)
        Renvoyer un nœud avec le label de la classe

Si (I est vide)
        Renvoyer un nœud avec le label le plus fréquent sur l 'attribut de sortie de T

Calculer le gain d 'information pour tous les attributs de I relativement à T

X est l 'attribut avec le plus grand gain

{x_j / j=1,2,…,m} sont les valeurs de X

{T_j / j=1,2,…,m} sont les sous ensembles de T décomposé par rapport aux x_j

Renvoyer un arbre avec X comme label du nœud racine
        et x_1, x_2, …, x_m comme labels des arcs allant aux arbres
        ID3(I-{X},O,T_1), ID3(I-{X},O,T_2), …, ID3(I-{X},O,T_m)
```

Arbres de Décision : Exercice

```
ID3 ({M,A,R,E},{I},{1,2,3,4,5,6,7,8})
```


Arbres de Décision : Exercice

$$\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$$



$$\text{Entropie}(1) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$$

$$\text{Entropie}(2) = -\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right) \approx 0{,}918$$

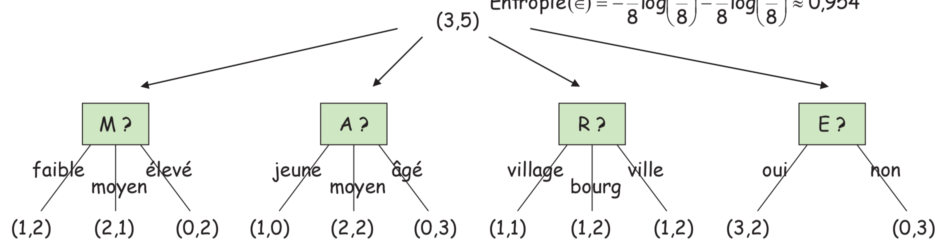
$$\text{Entropie}(3) = -\frac{0}{2}\log\left(\frac{0}{2}\right) - \frac{2}{2}\log\left(\frac{2}{2}\right) = 0$$

$$Gain(\in, M) = \text{Entropie}(\in) - \left(\frac{3}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2) + \frac{2}{8}\text{Entropie}(3)\right)$$

$$= \text{Entropie}(\in) - 0{,}688$$

Arbres de Décision : Exercice

$$\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$$



$$\text{Entropie}(1) = -\frac{1}{1}\log\left(\frac{1}{1}\right) - \frac{0}{1}\log\left(\frac{0}{1}\right) = 0$$

$$\text{Entropie}(2) = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$$

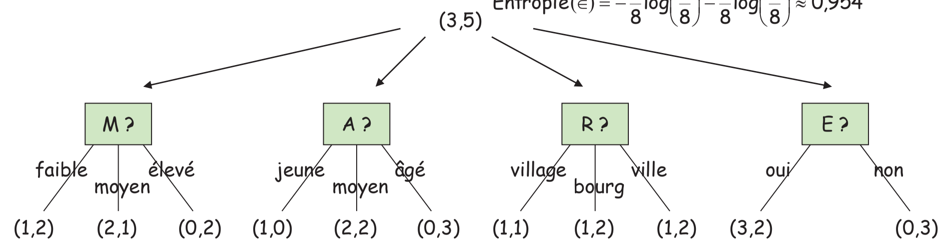
$$\text{Entropie}(3) = -\frac{0}{3}\log\left(\frac{0}{3}\right) - \frac{3}{3}\log\left(\frac{3}{3}\right) = 0$$

$$Gain(\in, A) = \text{Entropie}(\in) - \left(\frac{1}{8}\text{Entropie}(1) + \frac{4}{8}\text{Entropie}(2) + \frac{3}{8}\text{Entropie}(3)\right)$$

$$= \text{Entropie}(\in) - 0{,}5$$

Arbres de Décision : Exercice

$$\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$$



$$\text{Entropie}(1) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

$$\text{Entropie}(2) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$$

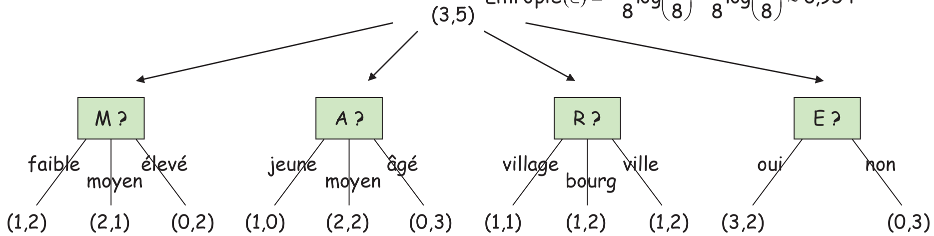
$$\text{Entropie}(3) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) \approx 0{,}918$$

$$Gain(\in, R) = \text{Entropie}(\in) - \left(\frac{2}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2) + \frac{3}{8}\text{Entropie}(3)\right)$$

$$= \text{Entropie}(\in) - 0{,}938$$

Arbres de Décision : Exercice

$\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



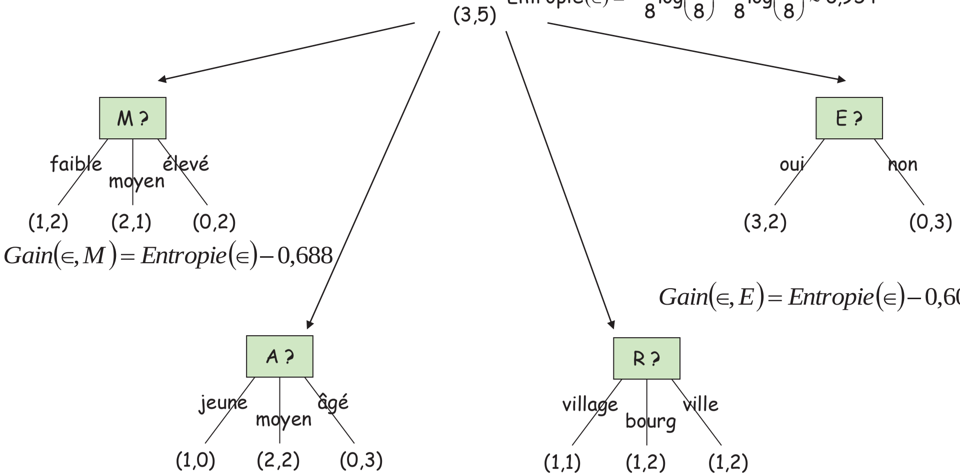
$\text{Entropie}(1) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) \approx 0.97$

$\text{Entropie}(2) = -\frac{0}{3}\log\left(\frac{0}{3}\right) - \frac{3}{3}\log\left(\frac{3}{3}\right) = 0$

$\text{Gain}(\in, E) = \text{Entropie}(\in) - \left(\frac{5}{8}\text{Entropie}(1) + \frac{3}{8}\text{Entropie}(2)\right) = \text{Entropie}(\in) - 0{,}607$

Arbres de Décision : Exercice

$\text{Entropie}(\in) = -\frac{3}{8}\log\left(\frac{3}{8}\right) - \frac{5}{8}\log\left(\frac{5}{8}\right) \approx 0{,}954$



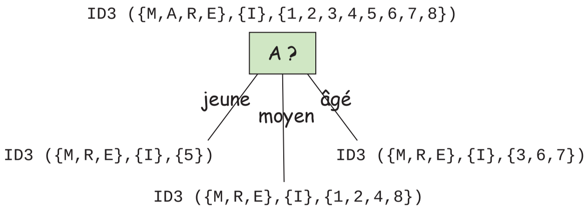
$Gain(\in, M) = Entropie(\in) - 0{,}688$

$Gain(\in, E) = Entropie(\in) - 0{,}607$

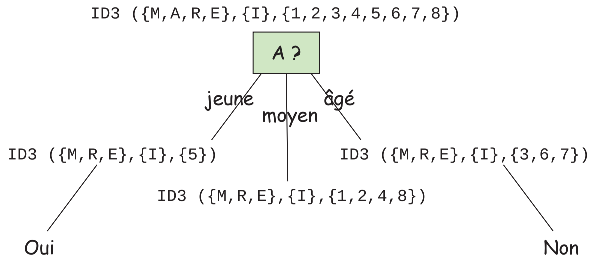
$Gain(\in, A) = Entropie(\in) - 0{,}5$

$Gain(\in, R) = Entropie(\in) - 0{,}938$

Arbres de Décision : Exercice



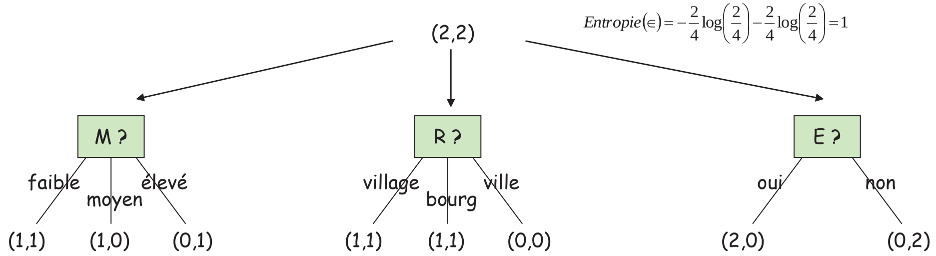
Arbres de Décision : Exercice



	M	R	E	I
1	moyen	village	oui	oui
2	élevé	bourg	non	non
4	faible	bourg	oui	oui
8	faible	village	non	non

Arbres de Décision : Exercice

$$Entropie(\in) = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$$



$$Entropie(1) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

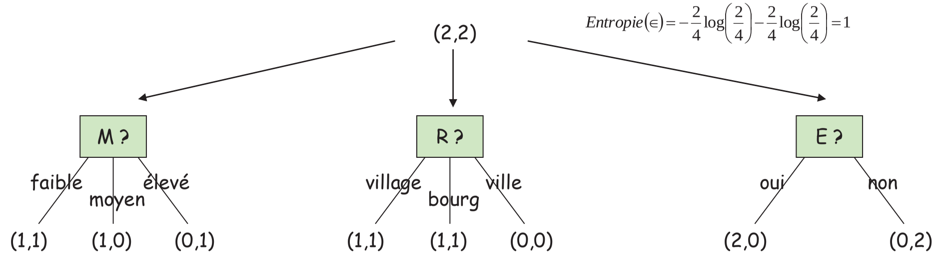
$$Entropie(2) = -\frac{1}{1}\log\left(\frac{1}{1}\right) - \frac{0}{1}\log\left(\frac{0}{1}\right) = 0$$

$$Entropie(3) = -\frac{0}{1}\log\left(\frac{0}{1}\right) - \frac{1}{1}\log\left(\frac{1}{1}\right) = 0$$

$$Gain(\in, M) = Entropie(\in) - \left(\frac{2}{4}Entropie(1) + \frac{1}{4}Entropie(2) + \frac{1}{4}Entropie(3)\right) = Entropie(\in) - 0{,}5$$

Arbres de Décision : Exercice

$$Entropie(\in) = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$$



$$Entropie(1) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

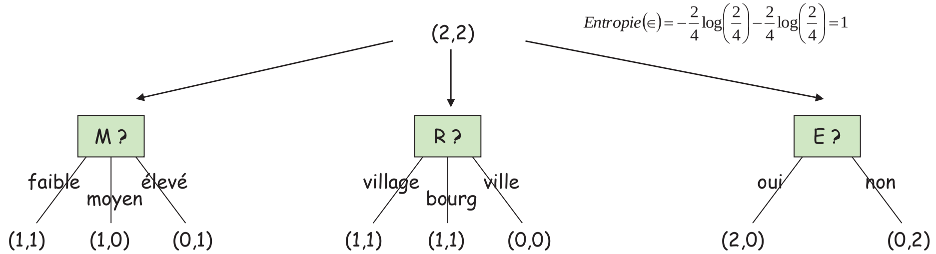
$$Entropie(2) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

$$Entropie(3) = -\frac{0}{0}\log\left(\frac{0}{0}\right) - \frac{0}{0}\log\left(\frac{0}{0}\right) = 0$$

$$Gain(\in, R) = Entropie(\in) - \left(\frac{2}{4}Entropie(1) + \frac{2}{4}Entropie(2) + \frac{0}{4}Entropie(3)\right) = Entropie(\in) - 1$$

Arbres de Décision : Exercice

$$Entropie(\in) = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$$

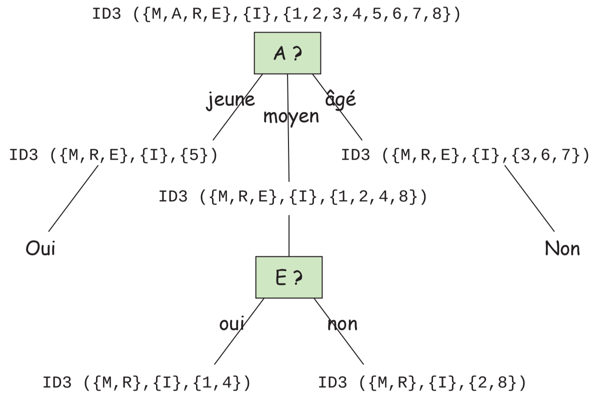


$$Entropie(1) = -\frac{2}{2}\log\left(\frac{2}{2}\right) - \frac{0}{2}\log\left(\frac{0}{2}\right) = 0$$

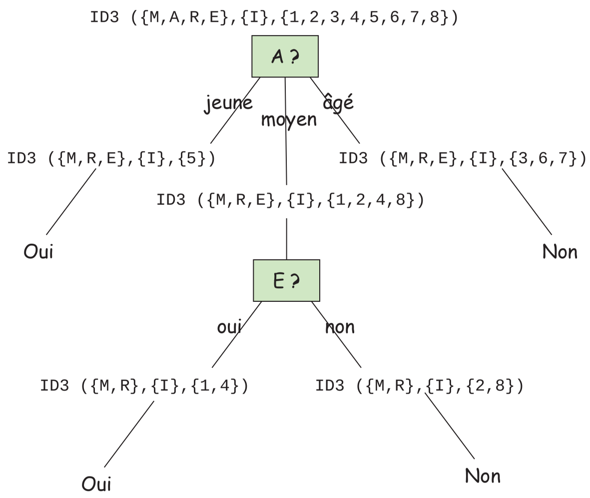
$$Entropie(2) = -\frac{0}{2}\log\left(\frac{0}{2}\right) - \frac{2}{2}\log\left(\frac{2}{2}\right) = 0$$

$$Gain(\in, E) = Entropie(\in) - \left(\frac{2}{4}Entropie(1) + \frac{2}{4}Entropie(2)\right) = Entropie(\in) - 0$$

Arbres de Décision : Exercice

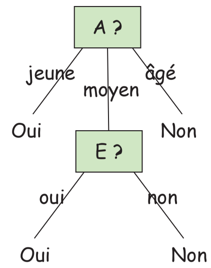


Arbres de Décision : Exercice



Arbres de Décision : Exercice

- Arbre final



Arbres de Décision : Exercice

- Evaluer les performances

	M	A	R	E	Ir	Ie
1	moyen	moyen	village	oui	oui	oui
2	élevé	moyen	bourg	non	non	non
3	faible	âgé	bourg	non	non	non
4	faible	moyen	bourg	oui	oui	oui
5	moyen	jeune	ville	oui	oui	oui
6	élevé	âgé	ville	oui	non	non
7	moyen	âgé	ville	oui	non	non
8	faible	moyen	village	non	non	non
9	moyen	âgé	village	oui	oui	non
10	élevé	jeune	ville	non	oui	oui
11	faible	âgé	village	non	non	non
12	moyen	moyen	bourg	oui	non	oui

Arbres de Décision : Exercice

- Peut-on jouer au tennis ?

	Outlook	Temperature	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

Organisation du cours

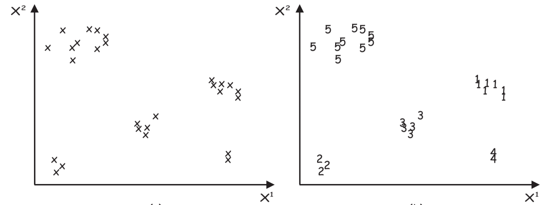
Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

Clustering

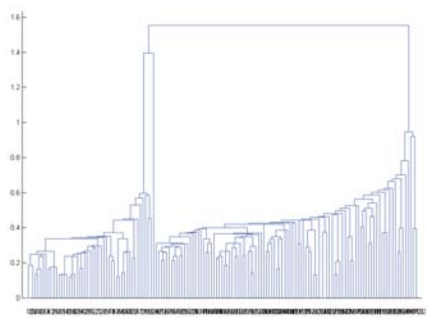
- Principe: action de découper un ensemble d'objets en groupes (clusters) de telle sorte que les caractéristiques des objets dans un même cluster soient similaires et que les caractéristiques des objets dans des clusters différents soient distinctes



- Méthodes :
  - Clustering hiérarchique
  - K-means
  - Fuzzy C-means
  - Substractive clustering
  - ...

Le clustering hiérarchique

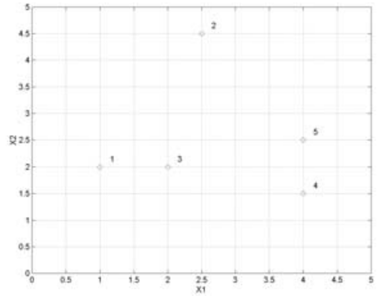
- Étape 1 : trouver les similarités entre paires d'objets
- Étape 2 : Grouper les objets sous la forme d'un arbre
- Étape 3 : Déterminer la coupe de l'arbre



Le clustering hiérarchique

- Étape 1 : trouver les similarités entre toutes les paires d'objets

#Point	X1	X2
1	1.0000	2.0000
2	2.5000	4.5000
3	2.0000	2.0000
4	4.0000	1.5000
5	4.0000	2.5000



Distance entre les points

	1	2	3	4	5
1	0	2.9155	1.0000	3.0414	3.0414
2	2.9155	0	2.5495	3.3541	2.5000
3	1.0000	2.5495	0	2.0616	2.0616
4	3.0414	3.3541	2.0616	0	1.0000
5	3.0414	2.5000	2.0616	1.0000	0

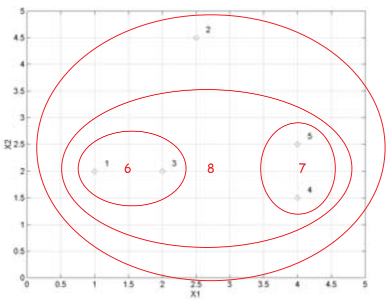
Fonctions Matlab : pdist, squareform

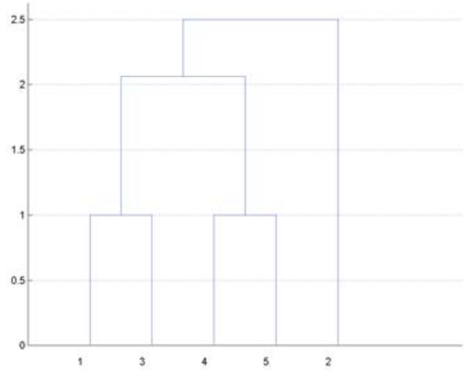
Le clustering hiérarchique

- Étape 2 : Grouper les objets sous la forme d'un arbre

```
      point 1   point 2   distance
G1    1.0000    3.0000    1.0000
G2    4.0000    5.0000    1.0000
G3    6.0000    7.0000    2.0616
G4    8.0000    2.0000    2.5000
```

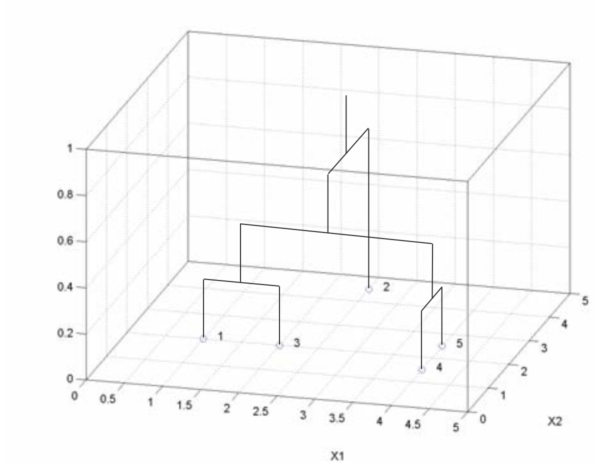
```
Fonctions Matlab : linkage, dendrogram
```





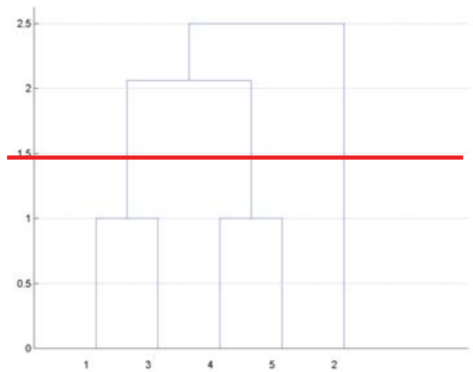
Le clustering hiérarchique

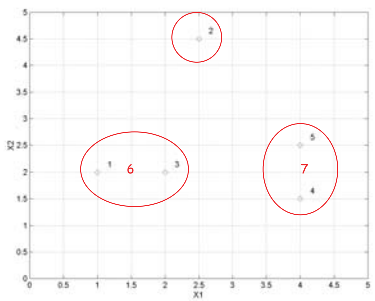
- Étape 2 : Grouper les objets sous la forme d'un arbre



Le clustering hiérarchique

- Étape 3 : Déterminer la coupe de l'arbre





L'algorithme du K-means

- On considère l'espace de n points de dimension p suivant :

$$X : \begin{array}{c|ccccc|} & X^1 & \dots & X^j & \dots & X^p \\ \hline X_1 & x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_i & x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_n & x_n^1 & \dots & x_n^j & \dots & x_n^p \\ \hline \end{array}$$

- On suppose que les n points peuvent être groupés en c clusters c < n
- Les clusters sont décrits par leurs centres

$$V_k = [v_k^1, v_k^2, \dots, v_k^j, \dots, v_k^p], 1 \le k \le c$$

- On note d(i,k) la distance entre le point X_i et le centre V_k
- Le point X_i appartient au cluster dont le centre est le plus proche
- On note m_k la moyenne des vecteurs dans le cluster k

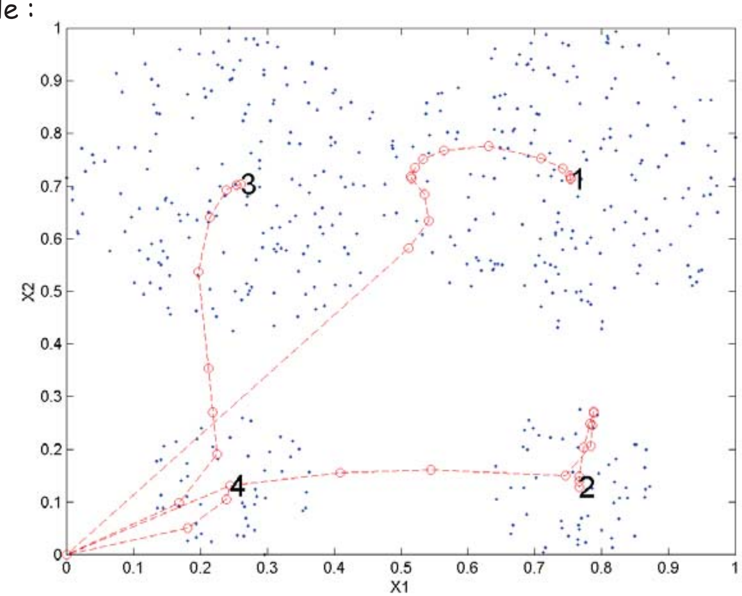
L'algorithme du K-means

- Algorithme :
  - Initialiser la position des centres : $V_k = [v_k^1, v_k^2, ..., v_k^j, ..., v_k^p], 1 \le k \le c$
  - Calculer les m_k
  - Jusqu'à ce qu'il n'y ait plus de changement sur les m_k

    Faire
    - Chaque point X_i est affecté au cluster le plus proche
    - Calculer les nouveaux m_k
  - Fin Jusqu'à

L'algorithme du K-means

- Exemple :

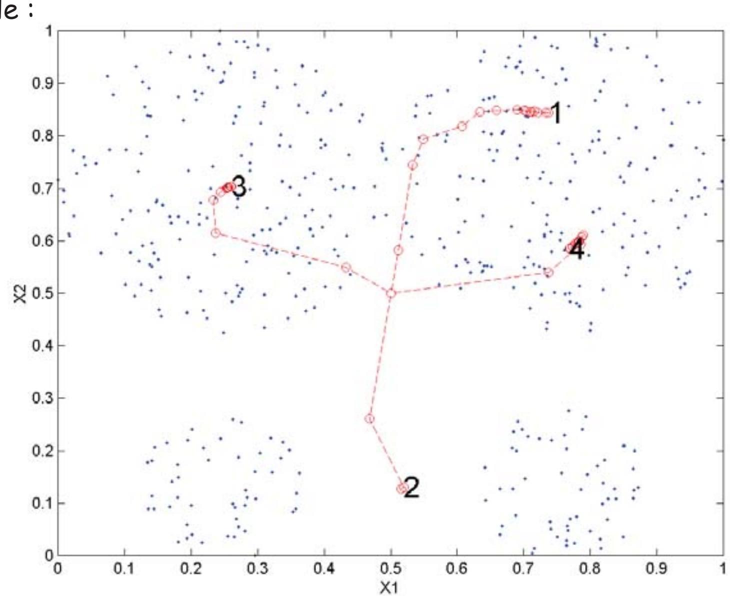


L'algorithme du K-means

- Algorithme très simple mais :
  - L'initialisation des positions des centres conditionne le résultat final (cf. transparent suivant)
  - Il peut arriver qu'un cluster soit vide
  - Le résultat dépend de la métrique utilisée pour calculer les distances
  - Le nombre de clusters k doit être fixé a priori

L'algorithme du K-means

- Exemple :



Organisation du cours

Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

Introduction

Traiter plus efficacement, et/ou plus simplement des problèmes que l'informatique classique peinait à résoudre : Classification, Contrôle de processus ...

Années 60 - Lofti Zadeh : Sous-ensembles Flous

Base formelle pour le traitement des
- Incertitudes
- Imprécisions
- Incomplétudes

Interface entre :
- l'information numérique (quantitatif)
- et l'information symbolique (qualitatif)

Années 70/80 - Raisonnement Flou - Logique Floue

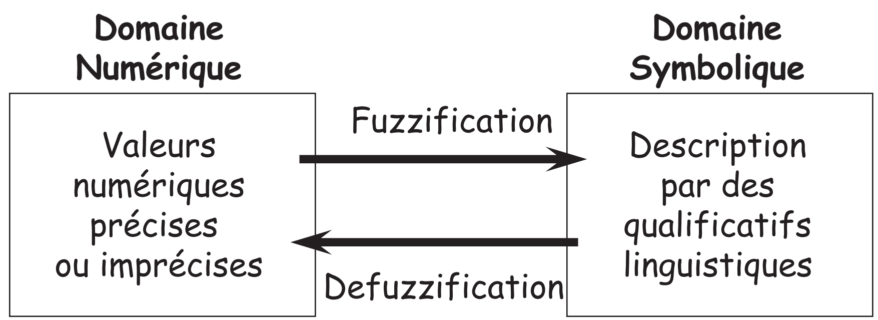
Premières applications industrielles

Introduction

- **Imprécisions** (vague, flou, général, ambigu)
  Difficultés dans l'énoncé de la connaissance
  "2000 à 3000 manifestants"
  "à peu près 1kg"
  "environ 1m80"
- **Incertitudes**
  Doutes sur la validité d'une connaissance
  "je crois que la voiture était blanche"
- **Incomplétudes**
  Absences de connaissances ou connaissances partielles

Objectif

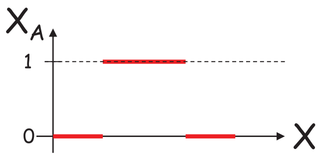
Interface entre :
l'information numérique (quantitatif)
et l'information symbolique (qualitatif)



Concepts de base : Définitions

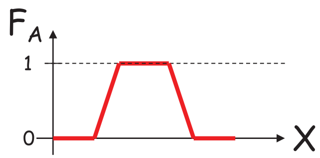
Un **sous ensemble classique** A de X est défini par une fonction caractéristique X_A

$$X_A : X \to \{0,1\}$$



Un **sous ensemble flou** A de X est défini par une fonction d'appartenance F_A

$$F_A : X \to [0,1]$$



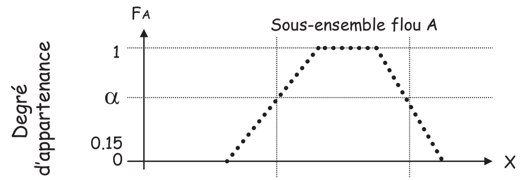
Notation :

$$A = \sum_{x \in X} F_A(x)/x \quad \text{si X est dénombrable}$$

$$A = \int F_A(x)/x \quad \text{si X est indénombrable}$$

Concepts de base : Fonction d'appartenance

$$F_A : \begin{array}{l} X \to [0,1] \\ x \to F_A(x) \end{array}$$



Support de A : $\text{supp}(A) = \{x \in X / F_A(x) \neq 0\}$

Noyau de A : $\text{noy}(A) = \{x \in X / F_A(x) = 1\}$

Hauteur de A : $h(A) = \sup_{x \in X}(F_A(x))$

Cardinalité de A : $|A| = \sum_{x \in X} F_A(x)$

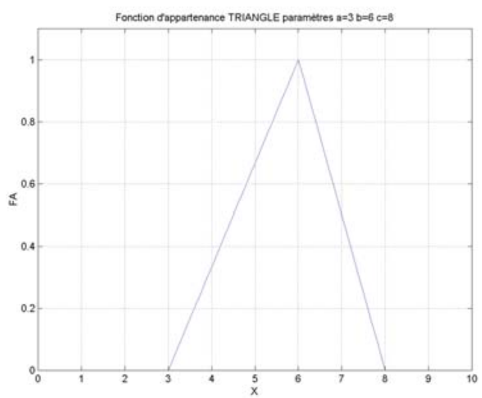
α - coupe de A : $A_\alpha = \{x \in X / F_A(x) \geq \alpha\}$

Concepts de base : Fonction d'appartenance

- Triangle

$$F_A(x;a,b,c,d) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases}$$

$$F_A(x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

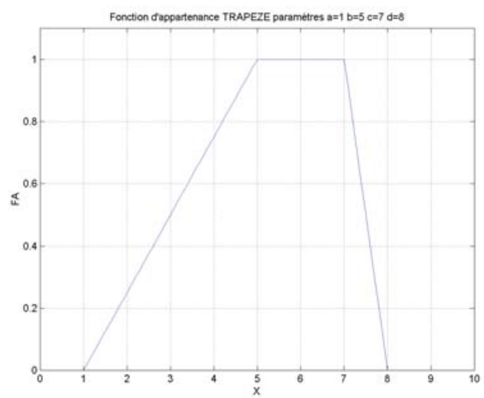


Concepts de base : Fonction d'appartenance

- Trapèze

$$F_A(x;a,b,c,d) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \dfrac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$$

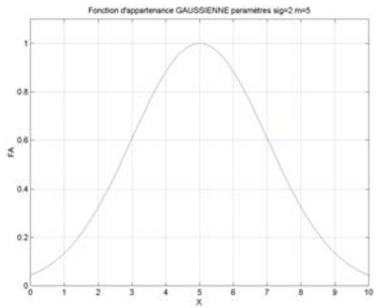
$$F_A(x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$$



Concepts de base : Fonction d'appartenance

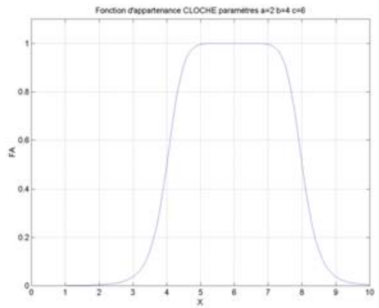
- Gaussienne

$$F_A(x;\sigma,m) = e^{\frac{-(x-m)^2}{2\sigma^2}}$$



- Cloche

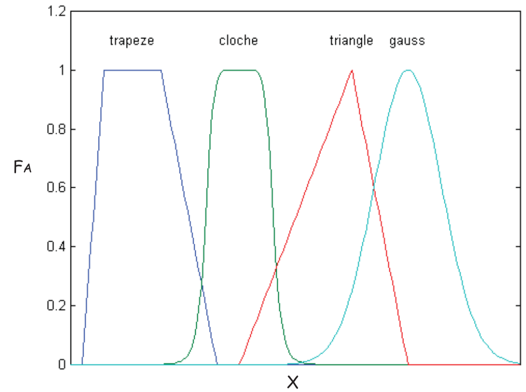
$$F_A(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}$$



Concepts de base : Nombre flou

Définition : un nombre flou est un sous-ensemble flou dont la fonction d'appartenance est normale et convexe.

- F_A est une fonction d'appartenance convexe ssi :
  $\forall(x,y)\in R\times R,\quad \forall z\in[x,y],\quad F_A(z)\geq \min(F_A(x),F_A(y))$
- F_A est est une fonction d'appartenance normale ssi : h(A)=1



Concepts de base : Norme Triangulaire

Une **norme triangulaire** (t-norme) est une fonction
$T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ qui vérifie pour tous a, b, c et d de [0, 1]

• T(0,0)=0 et T(1,1)=1	conditions aux limites
• T(a,1)=T(1,a)=a	élément neutre
• T(a,b)=T(b,a)	commutativité
• T(a,b) ≤T(c,d) si a≤c et b≤d	monotonie
• T(a,T(b,c))=T(T(a,b),c)	associativité

Exemple : T(a,b)=min(a,b)

Concepts de base : Conorme Triangulaire

Une **conorme triangulaire** (t-conorme) est une fonction
$S : [0, 1] \times [0, 1] \rightarrow [0, 1]$ qui vérifie pour tous a, b, c et d de [0, 1]

• S(1,1)=1 et S(0,0)=0	conditions aux limites
• S(a,0)=S(0,a)=a	élément neutre
• S(a,b)=S(b,a)	commutativité
• S(a,b) ≤S(c,d) si a≤c et b≤d	monotonie
• S(a,S(b,c))=S(S(a,b),c)	associativité

Exemple : S(a,b)=max(a,b)

Concepts de base : Norme et Conorme Triangulaire

- **Dualité**

  S(a,b) = 1 - T(1-a, 1-b)

  T(a,b) = 1 - S(1-a , 1-b)

- **Exemples**

  S(a,b) = max(a,b) T(a,b) = min(a,b)

  S(a,b) = a+b-ab T(a,b) = ab

- **Utilisation**

  Opérateur de conjonction $\cap$: T-norme

  Opérateur de disjonction $\cup$: T-conorme

Concepts de base : Norme et Conorme Triangulaire

- **Principales t-normes et t-conormes duales**

Nom	T-norme	T-conorme	Négation
Zadeh	$\min(x,y)$	$\max(x,y)$	$1-x$
Probabiliste	xy	$x+y-xy$	$1-x$
Lukasiewicz	$\max(x+y-1,0)$	$\min(x+y,1)$	$1-x$
Hamacher $\gamma > 0$	$\dfrac{xy}{\gamma+(1-\gamma)(x+y-xy)}$	$\dfrac{x+y-xy-(1-\gamma)xy}{1-(1-\gamma)xy}$	$1-x$
Weber	$\begin{cases} x \text{ si } y=1 \\ y \text{ si } x=1 \\ 0 \text{ sinon} \end{cases}$	$\begin{cases} x \text{ si } y=0 \\ y \text{ si } x=0 \\ 1 \text{ sinon} \end{cases}$	$1-x$

Concepts de base : Opérations

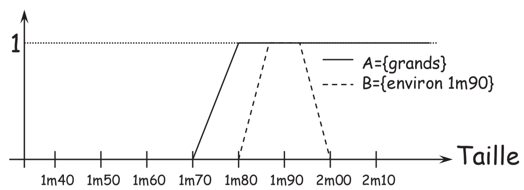
- Soit *A* et B deux sous ensembles flous de X et x un élément de X

  - Égalité

    $A = B \text{ ssi } \forall x \in X : F_A(x) = F_B(x)$

  - Inclusion

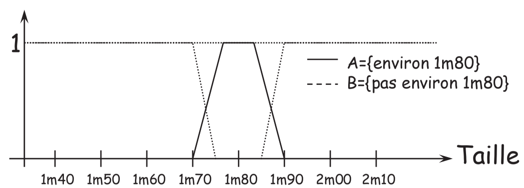
    $B \subset A \text{ ssi } \forall x \in X : F_A(x) \geq F_B(x)$



  - Complément

    $\bar{A}$ est défini $\forall x \in X$ par :

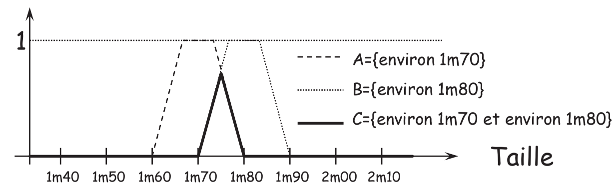
    $F_{\bar{A}}(x) = 1 - F_A(x)$



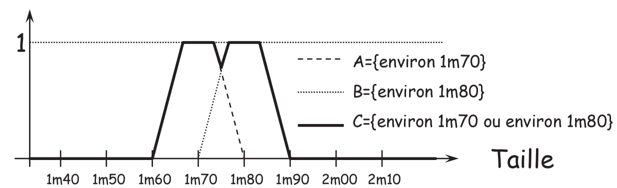
Concepts de base : Opérations

- Soit *A* et B deux sous ensembles flous de X et x un élément de X

  - Intersection $C = A \cap B \text{ ssi } \forall x \in X : F_C(x) = \min(F_A(x), F_B(x))$



  - Union $C = A \cup B \text{ ssi } \forall x \in X : F_C(x) = \max(F_A(x), F_B(x))$



Concepts de base : Relation Floue

- **Produit Cartésien :**

  Soient r sous-ensembles flous $A_1, A_2, \ldots, A_r$, respectivement définis sur $X_1, X_2, \ldots, X_r$, on définit leur produit cartésien $A = A_1 \times A_2 \times \ldots \times A_r$, comme un sous-ensemble flou de X de fonction d'appartenance :

  $$\forall\ x = (x_1, x_2; \ldots, x_r) \in X,\ F_A(x) = \min(F_{A1}(x_1), F_{A2}(x_2), \ldots, F_{Ar}(x_r))$$

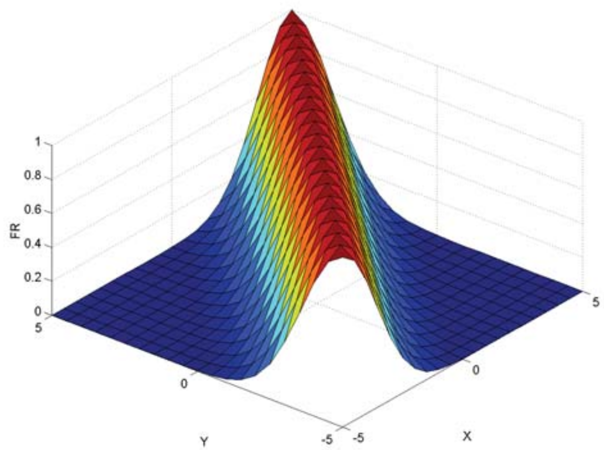
- **Relation floue :**

  Une relation floue entre r ensembles de référence $X_1, X_2, \ldots, X_r$, est un sous-ensemble flou de $X_1 \times X_2 \times \ldots \times X_r$, de fonction d'appartenance F_R.

Concepts de base : Relation Floue

Exemple : Relation floue traduisant "x est proche de y".

$$F_R(x,y) = e^{-k(x-y)^2},\ \forall (x,y) \in R \times R,\ k > 0$$

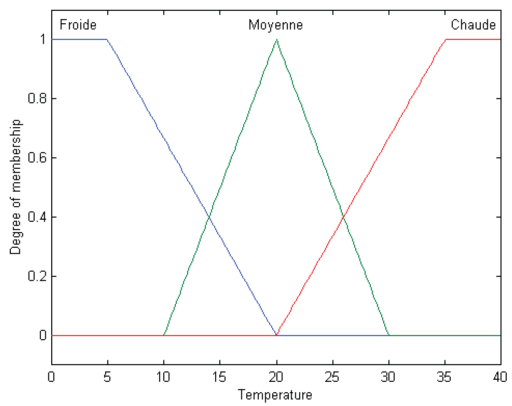


Concepts de base : Fuzzification - Variable Linguistique

Définition

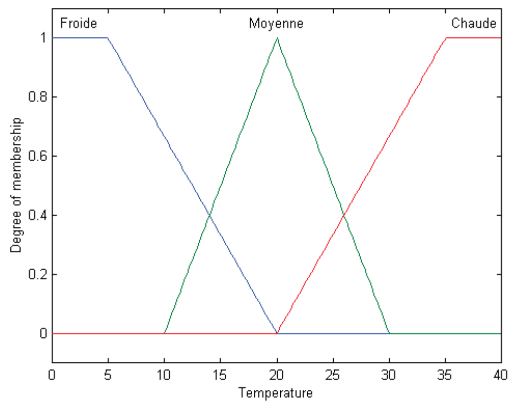
- Soit une variable E définie sur un univers de discours X, et T_E un ensemble de qualificatifs linguistiques ou ensemble des termes, caractérisant la variable E
- Une variable linguistique est formée du triplet (E, X, T_E)

Exemple : (Température, [0,40],{Froide, Moyenne, Chaude})



Concepts de base : Fuzzification en tableau

Exemple : (Température, [0,40],{Froide, Moyenne, Chaude})



Température	Froide	Moyenne	Chaude
0,00	1,00	0,00	0,00
2,00	1,00	0,00	0,00
4,00	1,00	0,00	0,00
6,00	0,93	0,00	0,00
8,00	0,80	0,00	0,00
10,00	0,67	0,00	0,00
12,00	0,53	0,20	0,00
14,00	0,40	0,40	0,00
16,00	0,27	0,60	0,00
18,00	0,13	0,80	0,00
20,00	0,00	1,00	0,00
22,00	0,00	0,80	0,13
24,00	0,00	0,60	0,27
26,00	0,00	0,40	0,40
28,00	0,00	0,20	0,53
30,00	0,00	0,00	0,67
32,00	0,00	0,00	0,80
34,00	0,00	0,00	0,93
36,00	0,00	0,00	1,00
38,00	0,00	0,00	1,00
40,00	0,00	0,00	1,00

Système d'inférence floue (SIF)

- Petits rappels de logique classique

NON

p	$\bar{p}$
0	1
1	0

ET

p \ q	0	1
0	0	0
1	0	1

OU

p \ q	0	1
0	0	1
1	1	1

Implication $p \rightarrow q$

p \ q	0	1
0	1	1
1	0	1

Modus ponens

SI	p	→ q	vrai
et	p		vrai
alors		q	vrai

Modus tollens

SI	p	→ q	vrai
et		q	faux
alors	p		faux

Système d'inférence floue

- Voici un exemple de relation d'implication :

« il fait beau » → « je suis heureux ».
Cette proposition est vraie si je suis toujours heureux quand il fait beau.

- *A ne pas confondre avec la relation d'équivalence (↔) qui elle implique que je ne soit heureux QUE lorsqu'il fait beau.*

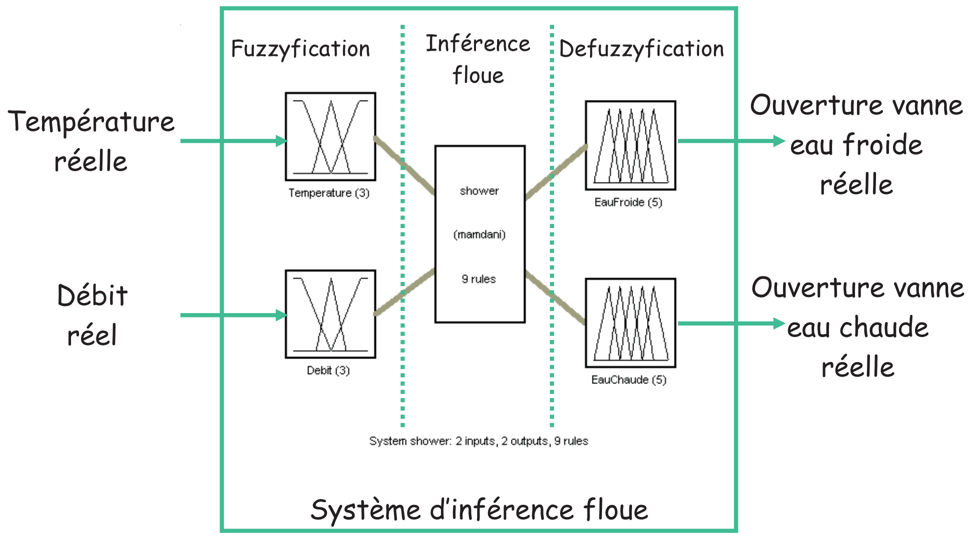
- La relation d'implication représente le SI une condition suffisante dans un sens, une condition nécessaire dans l'autre : dans $A \rightarrow B$, A est une condition suffisante de B, et B est une condition nécessaire de A

- la relation d'équivalence représente le SI ET SEULEMENT SI (↔), une condition nécessaire et suffisante ;

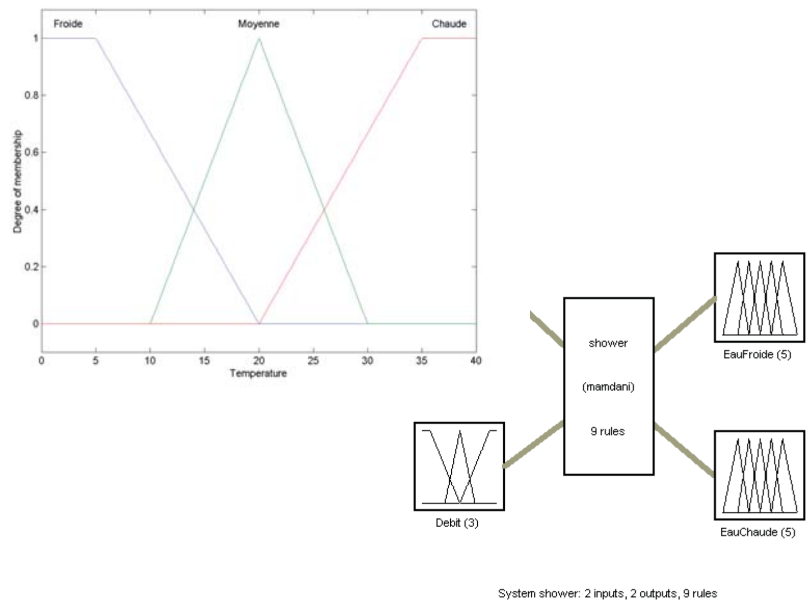
$A \leftrightarrow B$ équivaut à $(A \rightarrow B)$ ET $(B \rightarrow A)$

Système d'inférence floue (SIF)

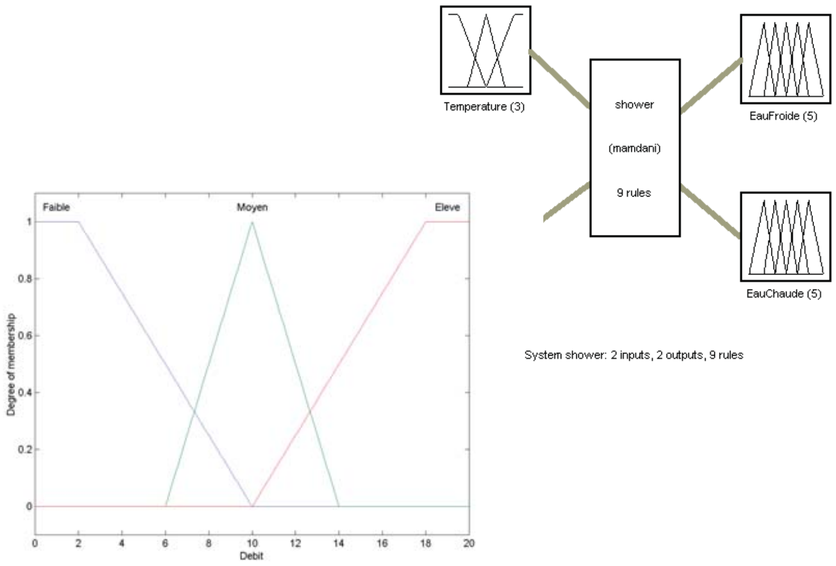
- Un exemple



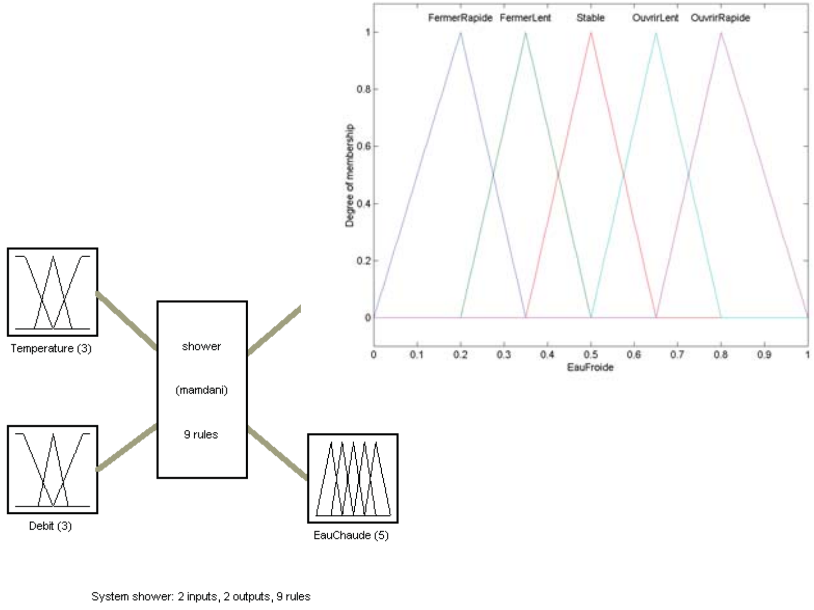
Système d'inférence floue (SIF)



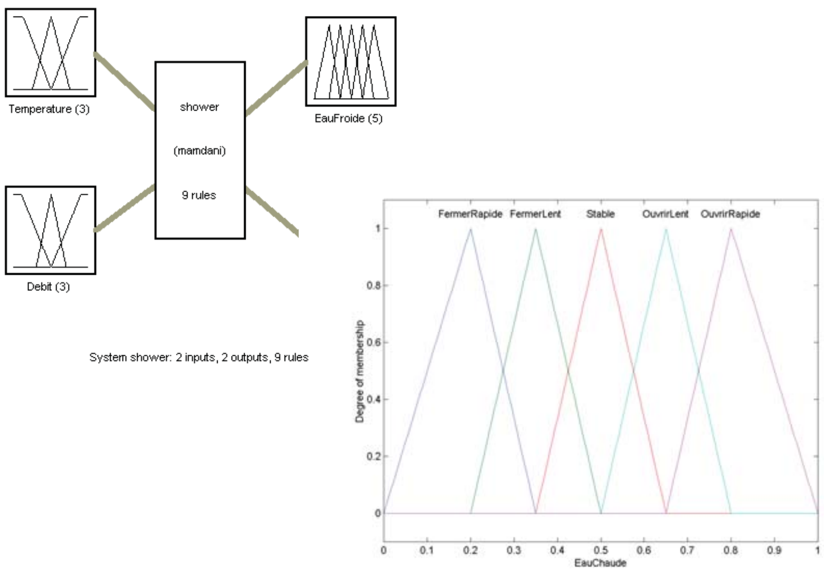
Système d'inférence floue (SIF)



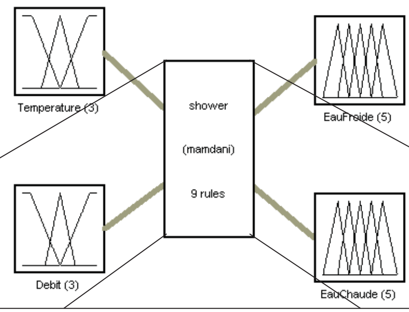
Système d'inférence floue (SIF)



Système d'inférence floue (SIF)

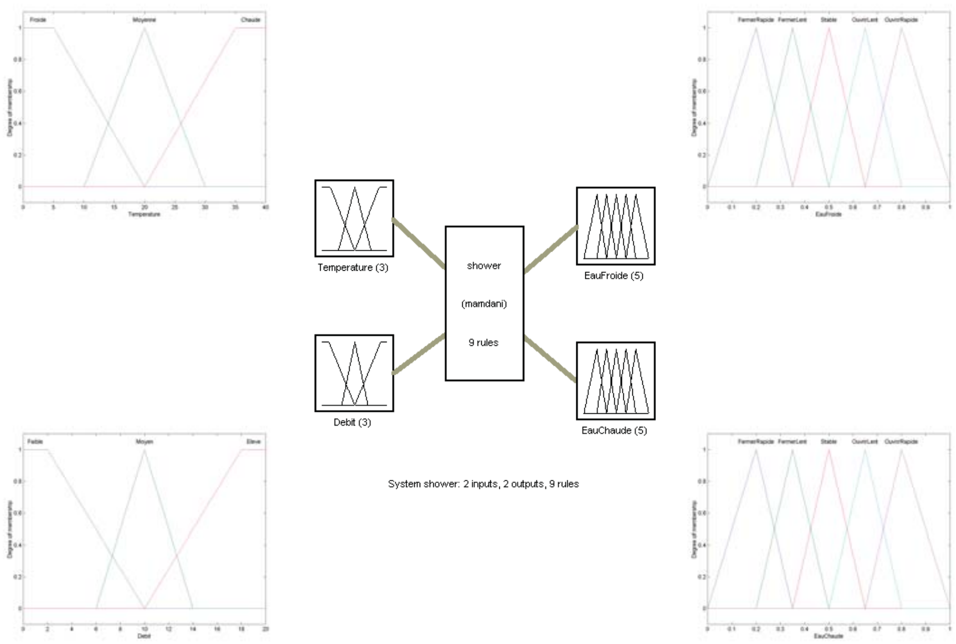


Système d'inférence floue (SIF)



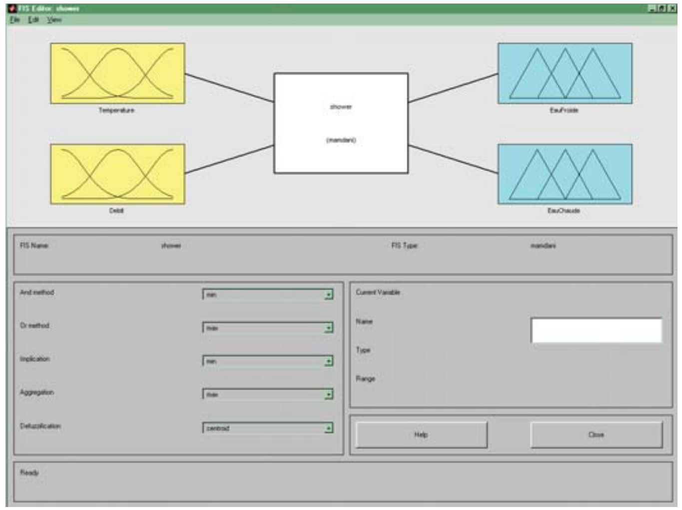
- 1 Si (Température est Froide) et (Débit est Faible) alors (EauFroide est OuvrirLent)(EauChaude est OuvrirRapide)
- 2 Si (Température est Froide) et (Débit est Moyen) alors (EauFroide est FermerLent)(EauChaude est OuvrirLent)
- 3 Si (Température est Froide) et (Débit est Eleve) alors (EauFroide est FermerRapide)(EauChaude est FermerLent)
- 4 Si (Température est Moyenne) et (Débit est Faible) alors (EauFroide est OuvrirLent)(EauChaude est OuvrirLent)
- 5 Si (Température est Moyenne) et (Débit est Moyen) alors (EauFroide est Stable)(EauChaude est Stable)
- 6 Si (Température est Moyenne) et (Débit est Eleve) alors (EauFroide est FermerLent)(EauChaude est FermerLent)
- 7 Si (Température est Chaude) et (Débit est Faible) alors (EauFroide est OuvrirRapide)(EauChaude est OuvrirLent)
- 8 Si (Température est Chaude) et (Débit est Moyen) alors (EauFroide est OuvrirLent)(EauChaude est FermerLent)
- 9 Si (Température est Chaude) et (Débit est Eleve) alors (EauFroide est FermerLent)(EauChaude est FermerRapide)

Système d'inférence floue (SIF)



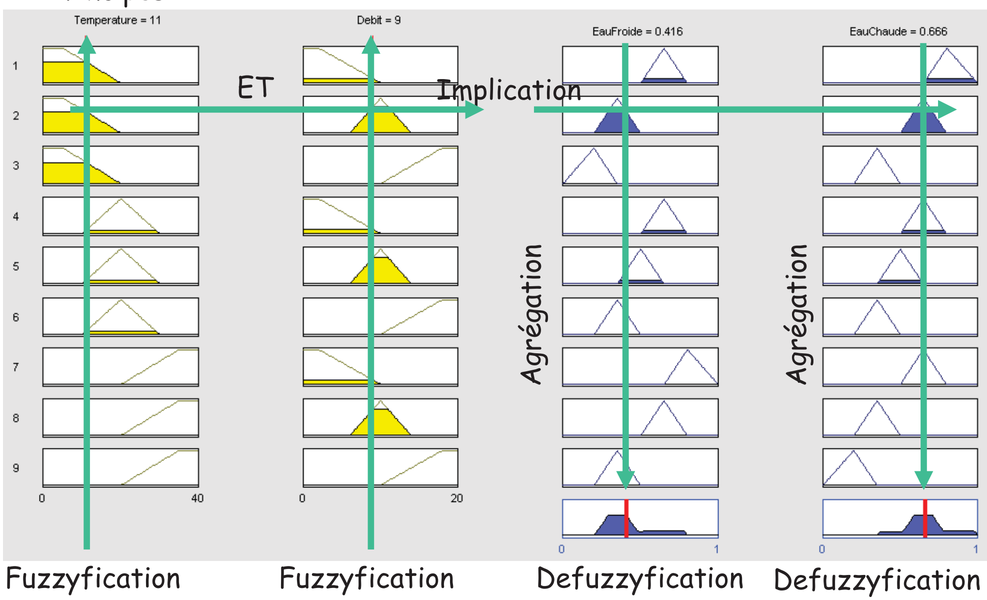
Système d'inférence floue (SIF)

- Interface MATLAB
  commande fuzzy



Système d'inférence floue (SIF)

- Principes



SIF : Proposition Floue

- Une **proposition floue** s'exprime à partir d'une variable linguistique (E,X,T_E) sous la forme "E est A", pour un terme de T_E, ou en appliquant un modificateur linguistique sur un élément de l'ensemble des termes.

- Exemples :
  - "L'âge est adulte"
  - "La température est très froide"

SIF : Règles Floues

- Les **règles floues** expriment un lien entre des propositions floues élémentaires ou des conjonctions de propositions élémentaires :

Si E est A **alors** U est B
Si E' est A' **ET** E''est A'' **ET** ... **alors** U est B

Exemples :

"**Si** Température est Froide **alors** Vanne_Eau_Chaude est Ouvrir_rapide"

"**Si** Température est Chaude **ET** Débit est faible **alors** Vanne_Eau_Chaude est Fermer_lent"

SIF : Opérateurs logiques

- **Opérateurs Logiques**

Classiques	Flous
ET	T-norme
OU	T-conorme
NON	négation
∀	inf
∃	sup
Implication	Implication floue

SIF : Opérateurs d'implication

- **Opérateurs d'implication floue**
  - Opérateur de Reichenbach $I(a,b) = 1 - a + ab$
  - Opérateur de Willmott $I(a,b) = \max(1 - a, \min(a,b))$
  - Opérateur de Rescher-Gaines $I(a,b) = \begin{cases} 1 & \text{si } a \le b \\ 0 & \text{si } a > b \end{cases}$
  - Opérateur de Kleene-Dienes $I(a,b) = \max(1 - a, b)$

SIF : Opérateurs d'implication

- **Opérateurs d'implication floue**
  - Opérateur de Brouwer-Gödel $I(a,b) = \begin{cases} 1 & \text{si } a \le b \\ b & \text{si } a > b \end{cases}$
  - Opérateur de Goguen $I(a,b) = \begin{cases} 1 & \text{si } a = 0 \\ \min(b/a, 1) & \text{si } a \ne 0 \end{cases}$
  - Opérateur de Lukasiewicz $I(a,b) = \min(1 - a + b/a, 1)$
  - Opérateur de Mamdani $I(a,b) = \min(a,b)$
  - Opérateur de Larsen $I(a,b) = ab$

SIF : Modus Ponens Binaire et Généralisé

- **Modus Ponens Binaire**

prémisse1	Si $E \rightarrow U$	vrai
prémisse2	Et E	vrai
Conclusion	Alors U	vrai

- **Modus Ponens Généralisé**

Expressions linguistiques				Opérateurs
prémisse1	Si E est A	Alors	U est B	Règle floue
prémisse2	Or E est A'			Proposition Floue
Conclusion	Donc		U est B'	Proposition Floue

SIF : Modus Ponens Généralisé

Décomposition du Modus Ponens Généralisé

Règle	**Si**	**E est A**	**Alors**	**U est B**
Univers		X		Y
Fonctions d'appartenance		$F_A(x)$		$F_B(y)$
Observation		**E est A'**		
Univers		X		
Fonction d'appartenance		$F_{A'}(x)$		
Conclusion				**U est B'**
Univers				Y
Fonction d'appartenance				$F_{B'}(y)$

SIF : Modus Ponens Généralisé

Calcul de $F_{B'}(Y)$

Règle Floue
Calcul de la relation floue R décrivant le lien causal entre les valeurs de E et celles de U, à l'aide d'un opérateur d'implication floue

Proposition floue
Pour l'observation E est A', on détermine la fonction d'appartenance $F_{A'}(x)$

Proposition floue
Pour tous les y de Y $\quad F_{B'}(y) = \sup_{x \in X} T(F_{A'}(x), R(x,y))$

T est une t-norme appelée opérateur de modus ponens généralisé

SIF : Modus Ponens Généralisé

- **Choix de T opérateur de modus ponens généralisé**

Le choix de T doit rendre le modus ponens généralisé compatible avec le modus ponens ordinaire, c'est à dire que l'on doit obtenir $F_{B'}$ identique à F_B dès que $F_{A'}$ est identique à F_A .

Les choix possibles sont :

Opérateurs T	Implication R
Lukasiewicz : $T(u,v) = \max(u+v-1,0)$	Reichenbach, Willmott, Rescher-Gaines, Kleene-Dienes, Brouwer-Gödel, Goguen, Lukasiewicz, Mamdani, Larsen
Zadeh : $T(u,v) = \min(u,v)$	Rescher-Gaines, Brouwer-Gödel, Mamdani, Larsen
Probabiliste : $T(u,v) = u.v$	Rescher-Gaines, Brouwer-Gödel, Goguen, Mamdani, Larsen

SIF : Modus Ponens Généralisé

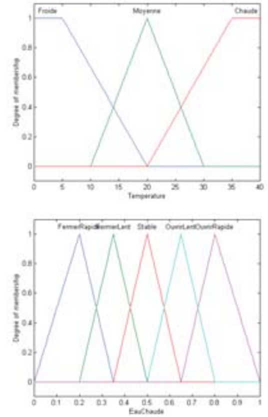
- **Exemple:**
  - Si la Température est Froide alors EauChaude est OuvrirRapide
  - Or la Température est de 11 °C
  - Donc EauChaude est ?

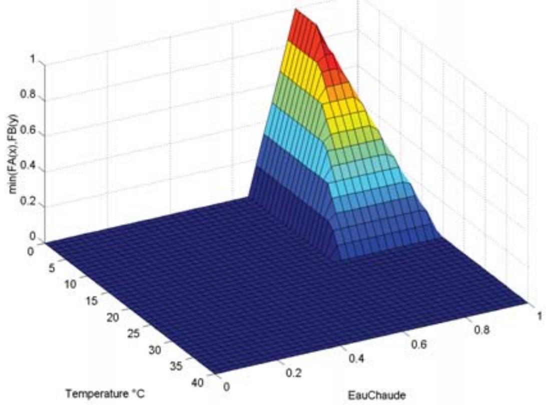
SIF : Modus Ponens Généralisé

Règle Floue

R :	Si	la Température est Froide	Alors	EauChaude est OuvrirRapide
	Si	E est A	Alors	U est B
Univers		[0°C, 40°C]		[0,1]
Appartenance		$F_A(x)$		$F_B(y)$

Opérateur d'implication : $\min(F_A(x), F_B(y))$





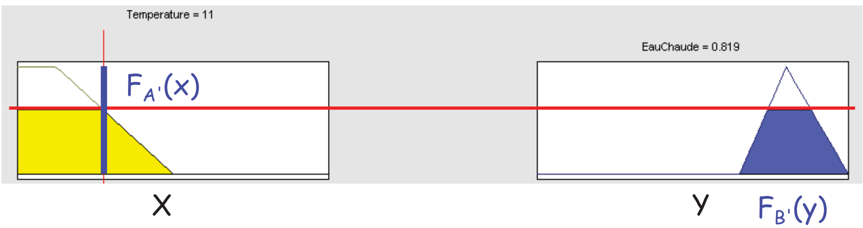
SIF : Modus Ponens Généralisé

Proposition floue : Or la Température est de 11 °C , $F_{A'}(x)$

$F_{B'}(y) = \sup_{x\in X} [T(F_{A'}(x),R(x,y))]$

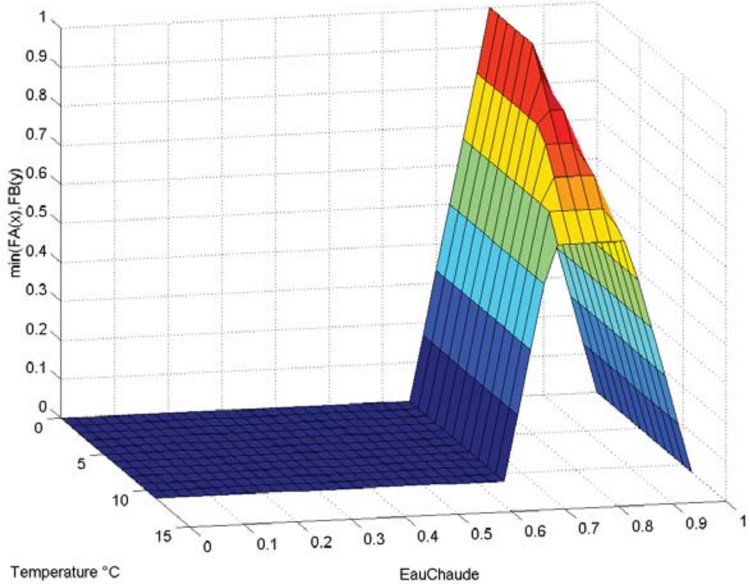
Choix de l'opérateur de modus ponens généralisé : Zadeh T=min(u,v)

$$F_{B'}(y) = \sup_{x\in X} [\min(F_{A'}(x),\min(F_A(x), F_B(y)))]$$
$$= \sup_{x\in X} [\min(F_{A'}(x), F_A(x), F_B(y))]$$
$$= \min(\sup_{x\in X} [\min(F_{A'}(x), F_A(x))], F_B(y))$$



SIF : Modus Ponens Généralisé

Proposition floue : Or la Température est de 11 °C , $F_{A'}(x)$



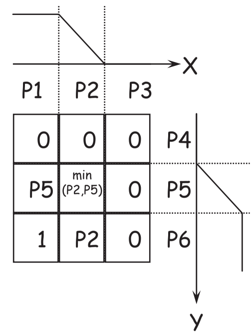
SIF : Modus Ponens Généralisé

- **Un autre exemple :**
  - Si la Distance est courte alors le Freinage est Puissant
  - Or la distance est de 70m
  - Or la Distance est d'environ 100m
  - Donc le Freinage est ?

SIF : Modus Ponens Généralisé

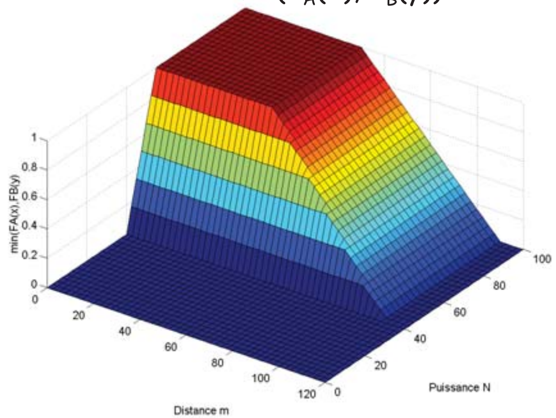
Règle Floue R :	Si	la Distance est Courte	Alors	le Freinage est Puissant
	Si	E est A	Alors	U est B
Univers		[0m, 120m]		[0N, 100N]
Fonctions d'appartenance		$F_A(x)$		$F_B(y)$

Distance Courte
$F_A(x)$



Freinage Puissant
$F_B(y)$

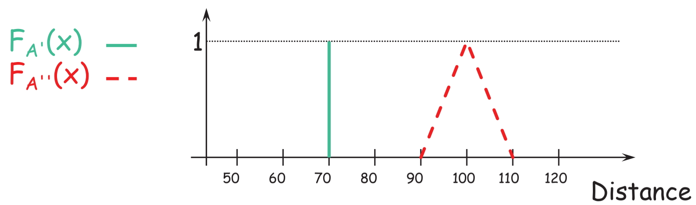
Opérateur d'implication
$\min(F_A(x), F_B(y))$



SIF : Modus Ponens Généralisé

- **Propositions floues :**

Or	la distance	est	de 70m	$F_{A'}(x)$
	E	**est**	**A'**	
Or	la Distance	est	d'environ 100m	$F_{A''}(x)$
	E	**est**	**A''**	

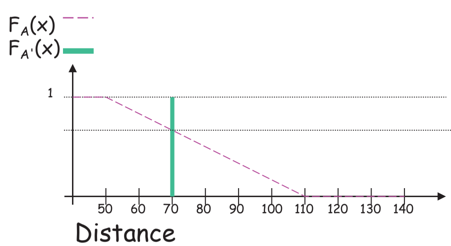


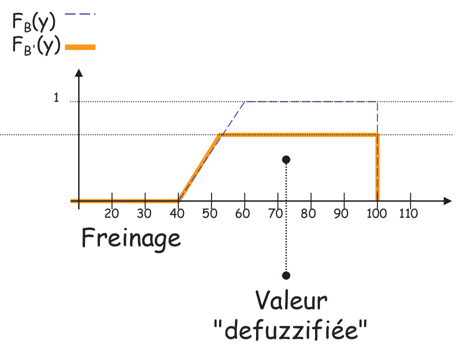
SIF : Modus Ponens Généralisé

Proposition floue : Or la distance est de 70m , $F_{A'}(x)$

$$F_{B'}(y) = \sup_{x\in X} [T(F_{A'}(x),R(x,y))] = \sup_{x\in X} [\min(F_{A'}(x),\min(F_A(x), F_B(y)))]$$

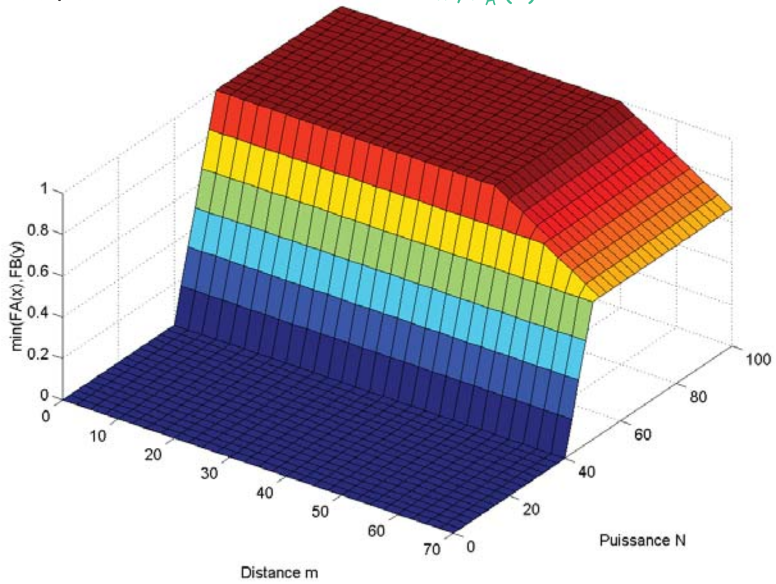
$$= \sup_{x\in X} [\min(F_{A'}(x), F_A(x), F_B(y))] = \min(\sup_{x\in X} [\min(F_{A'}(x), F_A(x))], F_B(y))$$





SIF : Modus Ponens Généralisé

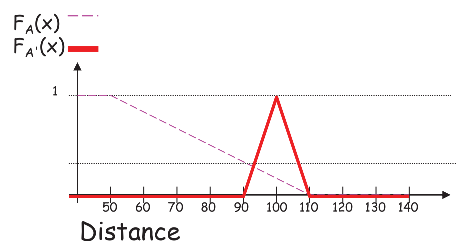
Proposition floue : Or la distance est de 70m , $F_{A'}(x)$

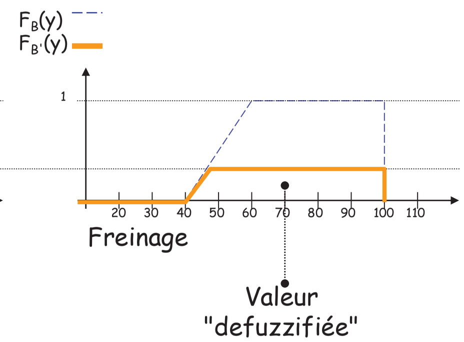


SIF : Modus Ponens Généralisé

Proposition floue : Or la Distance est d'environ 100m, $F_{A''}(x)$

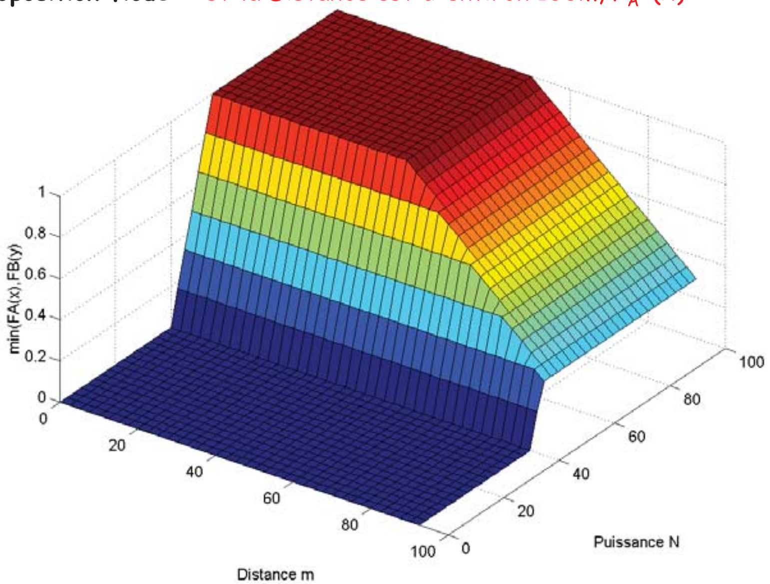
$$F_{B''}(y) = \sup_{x \in X} T(F_{A''}(x),R(x,y)) = \sup_{x \in X} [\min(\min(F_{A''}(x), F_A(x)), F_B(y))]$$
$$= \min(\sup_{x \in X} [\min(F_{A''}(x), F_A(x))], F_B(y))$$





SIF : Modus Ponens Généralisé

Proposition floue : Or la Distance est d'environ 100m, $F_{A''}(x)$

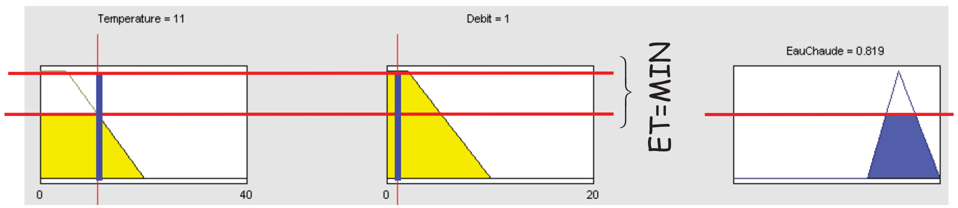


SIF : Modus Ponens Généralisé

- **Exemple une règle avec plusieurs prémices :**
- Si (Température est Froide) ET (Débit est Faible) alors (EauChaude est OuvrirRapide)
- Or la Température est de 11 °C ET le débit est 1m3/mn
- Donc EauChaude est ?

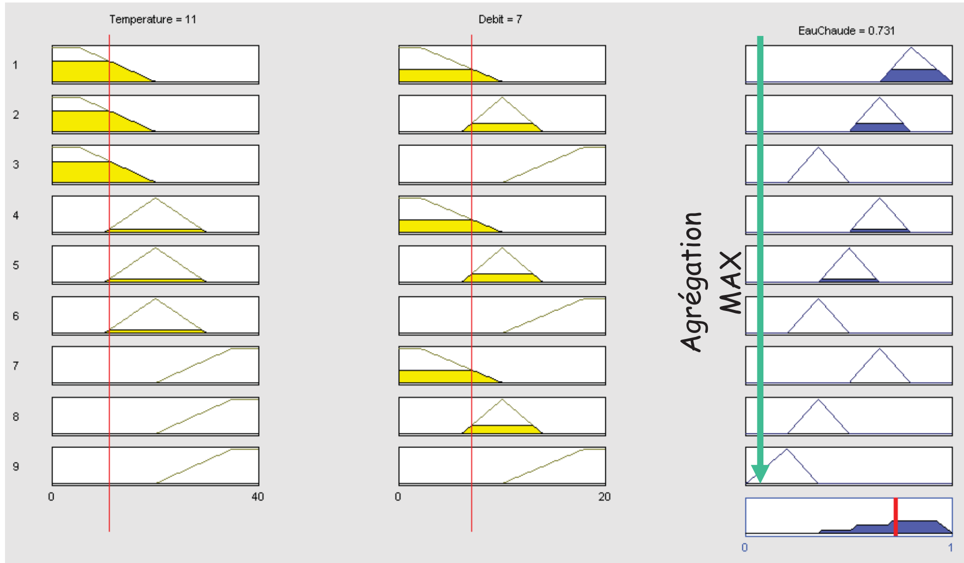
SIF : Déduction floue

- Mécanisme de déduction



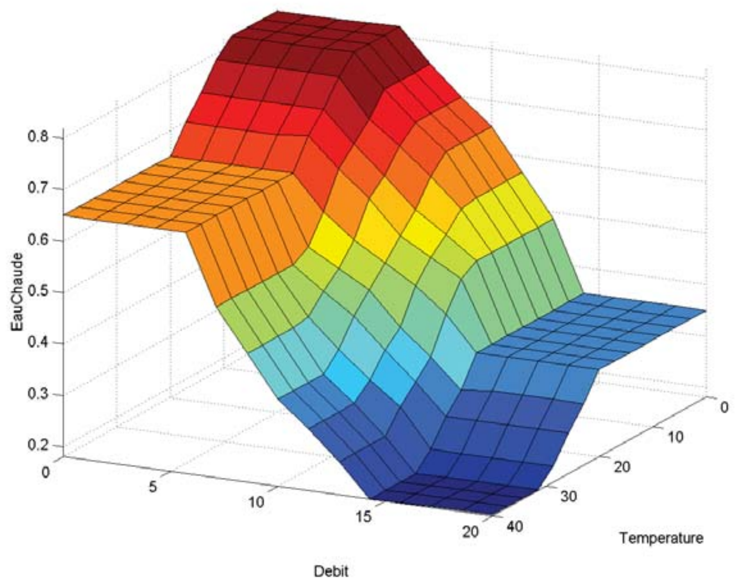
SIF : Déduction floue

- Exemple : plusieurs règles avec plusieurs prémices



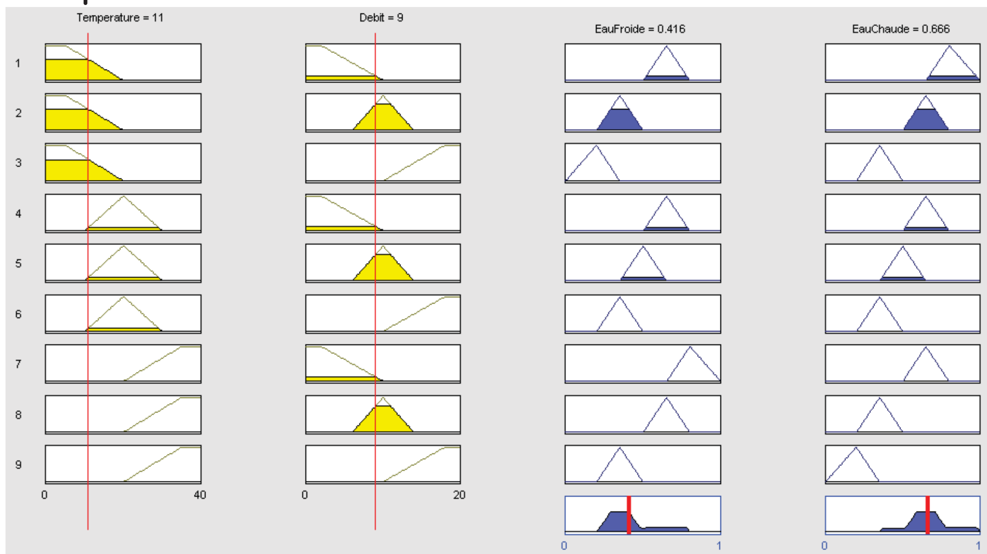
SIF : Déduction floue

- Exemple : plusieurs règles avec plusieurs prémices



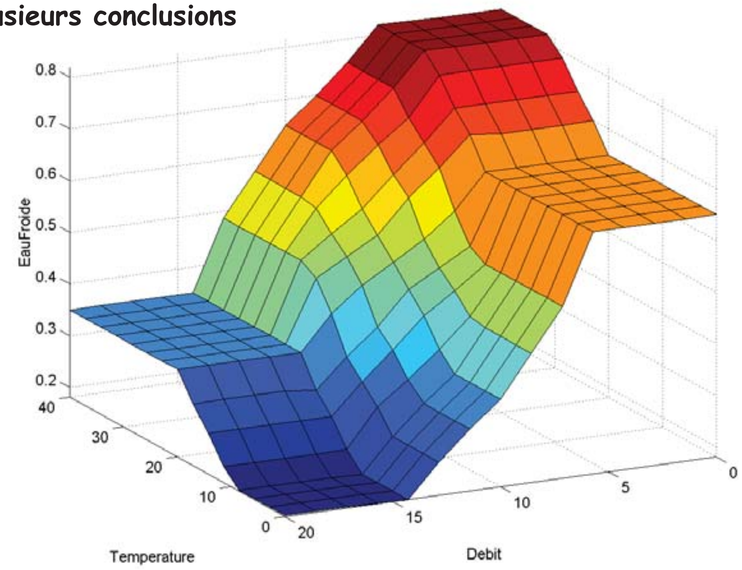
SIF : Déduction floue

- Exemple : plusieurs règles avec plusieurs prémices et plusieurs conclusions



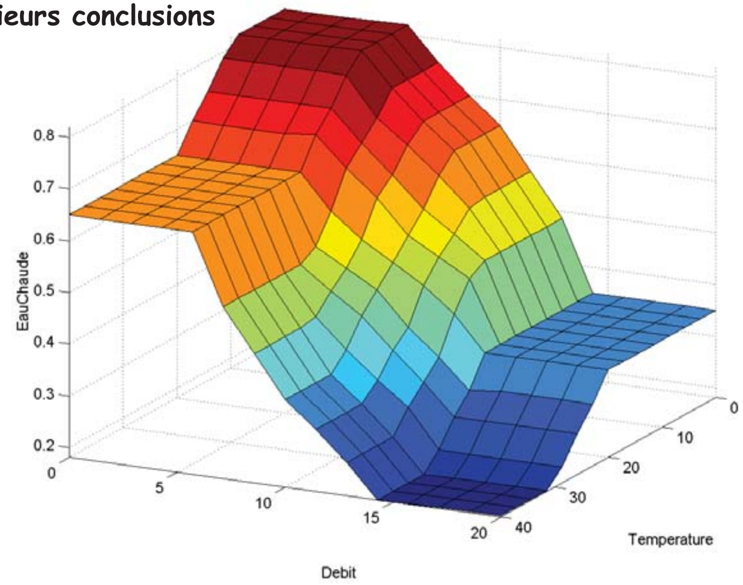
SIF : Déduction floue

- **Exemple : plusieurs règles avec plusieurs prémices et plusieurs conclusions**

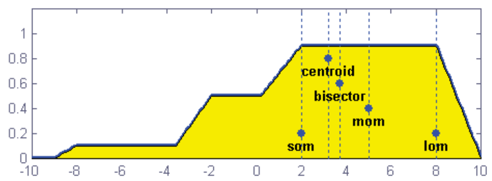


SIF : Déduction floue

- **Exemple : plusieurs règles avec plusieurs prémices et plusieurs conclusions**



SIF : Defuzzification



- Centroid : barycentre de la surface
- Bisector : bissectrice de la surface
- Som smallest of maximum : plus petite valeur des maxima
- Mom mean of maximum : moyenne des maxima
- Lom largest of maximum : plus grande valeur des maxima

SIF : types de modèles classiques

- MODELE DE TYPE MADMANI

  Les règles sont du type :
  SI x est *A* et y est B alors z est *C*

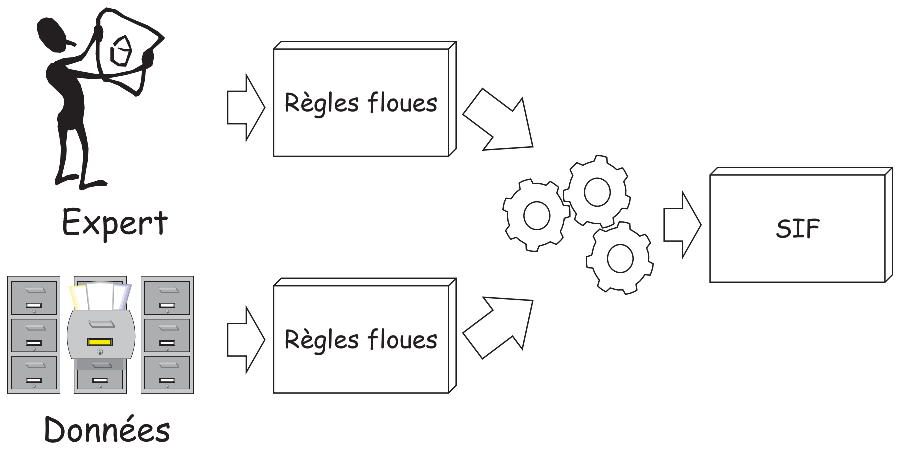
  Modélisation intuitive
  Adapté au raisonnement humain

- MODELE DE TYPE SUGENO

  Les règles sont du type:
  Si x est *A* et y est B alors z=f(x,y)
  Sugeno d'ordre O : f(x,y)=k
  Sugeno d'ordre 1 : f(x,y)=p.x+q.y+r

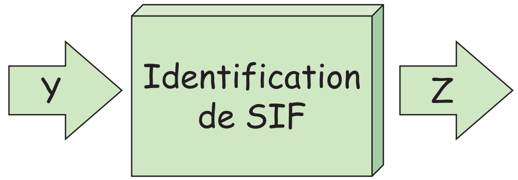
  Efficacité de calcul
  Bonne performance
  Garantie continuité des sorties

SIF : Obtention des règles



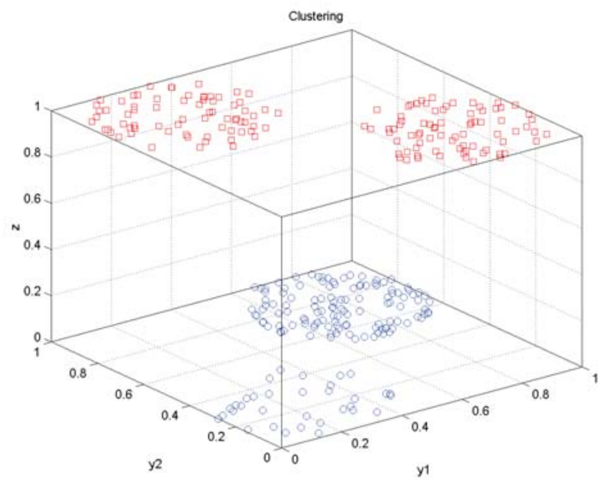
Clustering flou

- Méthode de Clustering
  - Méthode : « Substractive clustering » (SC)
  - Méthode : « Fuzzy C-means » (FCM)



Clustering flou : Substractive clustering

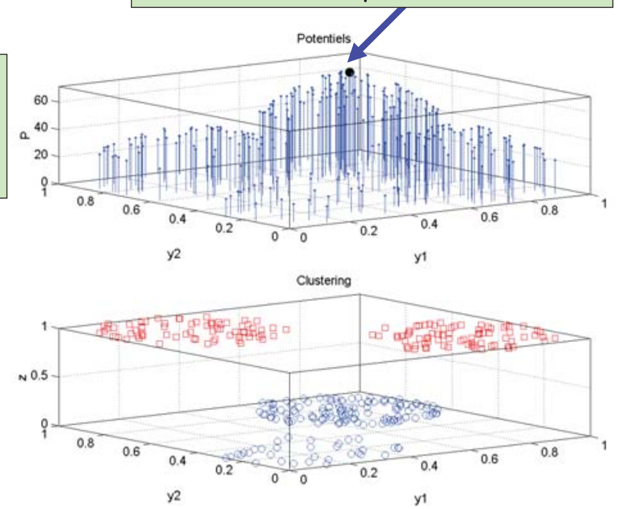
Exemple : échantillon 300 individus, 3 caractères



Clustering flou : Substractive clustering

Potentiel le plus élevé
Détermine le premier centre

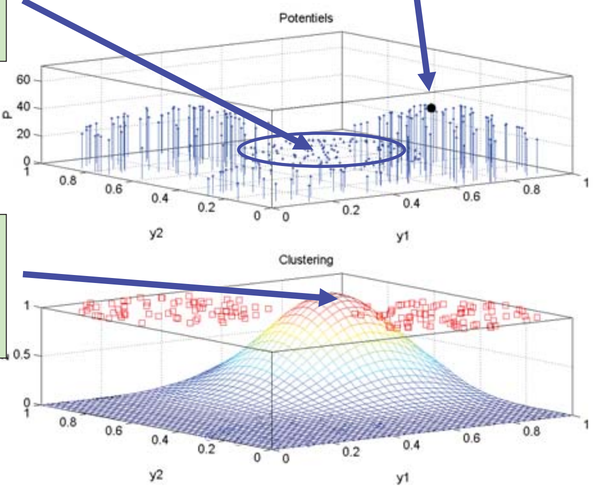
Potentiels d'un point: fonction du nombre de voisins et de leurs distances



Clustering flou : Substractive clustering

Diminution des potentiels les plus proches du centre déterminé

Potentiel le plus élevé

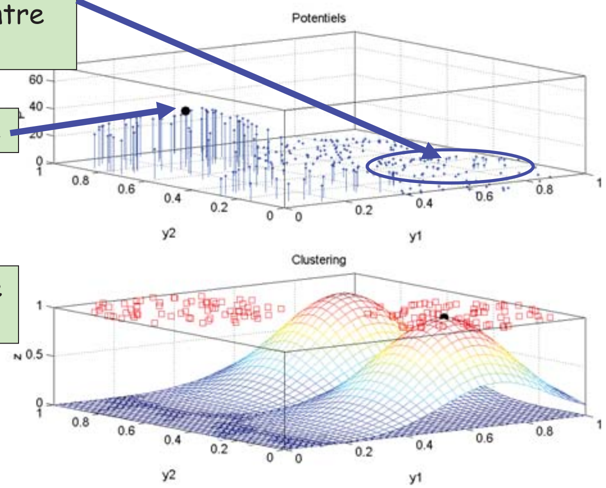


Fonction d'appartenance au premier centre déterminé

Clustering flou : Substractive clustering

Diminution des potentiels les plus proches du centre déterminé

Potentiel le plus élevé

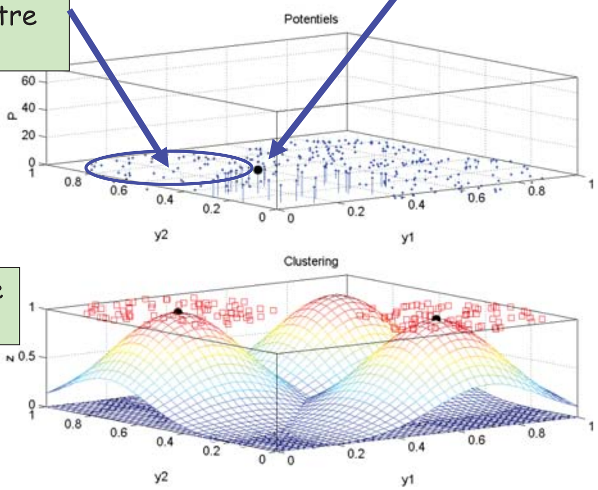


Fonctions d'appartenance aux centres

Clustering flou : Substractive clustering

Diminution des potentiels les plus proches du centre déterminé

Potentiel le plus élevé

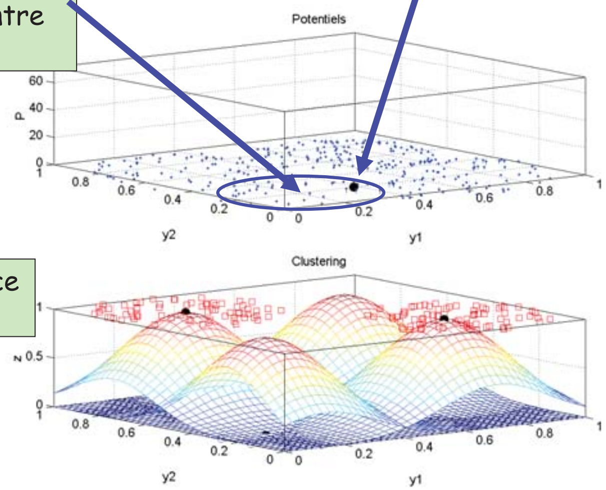


Fonctions d'appartenance aux centres

Clustering flou : Substractive clustering

Diminution des potentiels les plus proches du centre déterminé

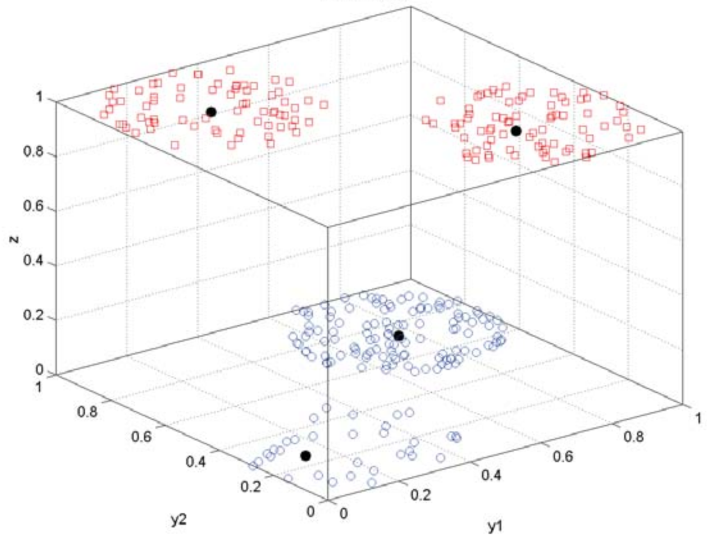
Potentiel le plus élevé



Fonctions d'appartenance aux centres

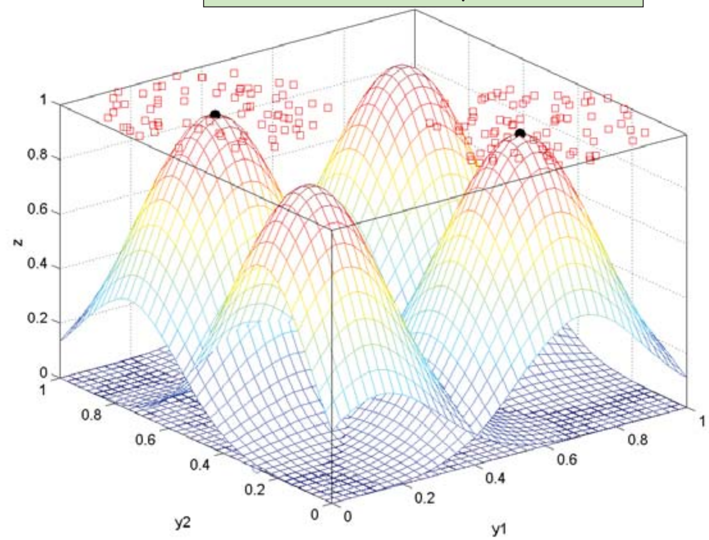
Clustering flou : Substractive clustering

Résultat du clustering



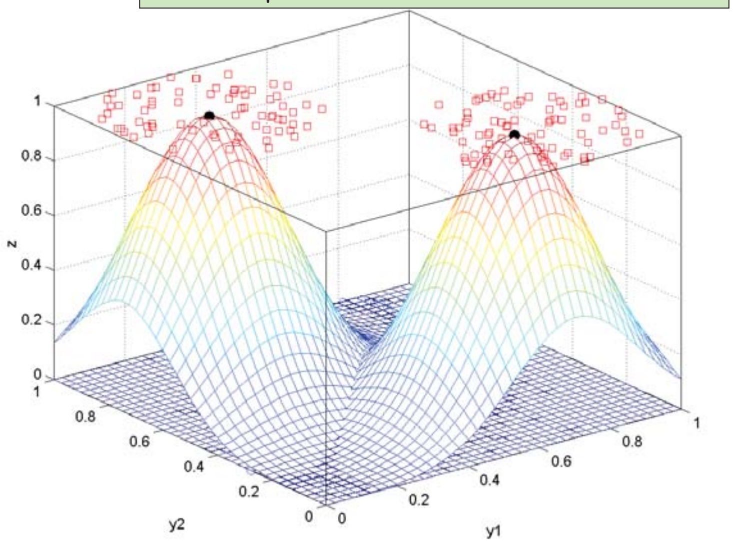
Clustering flou : Substractive clustering

Fonctions d'appartenance
associées à chaque centre



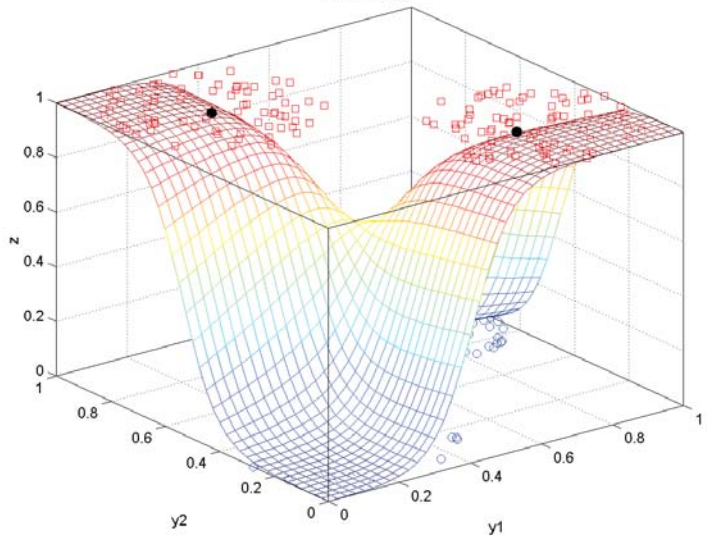
Clustering flou : Substractive clustering

Pondération des fonctions d'appartenance
par les valeurs des sorties



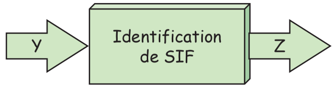
Clustering flou : Substractive clustering

Identification du SIF Final



Clustering flou : Substractive clustering

- Format des données



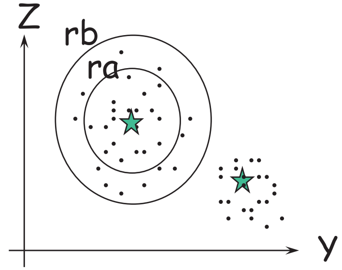
n points m entrées/sorties p entrées m-p sorties

$$\begin{array}{ccccc} [X_1] & \dots & [X_j] & \dots & [X_n] \end{array} \qquad \begin{array}{ccccc} \begin{bmatrix} Y_1 \\ Z_1 \end{bmatrix} & \dots & \begin{bmatrix} Y_j \\ Z_j \end{bmatrix} & \dots & \begin{bmatrix} Y_n \\ Z_n \end{bmatrix} \end{array}$$

$$X = \begin{bmatrix} x_1^1 & \dots & x_1^j & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & & x_i^j & & x_i^n \\ & & & & \\ x_m^1 & & x_m^j & & x_m^n \end{bmatrix} = \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} y_1^1 & \dots & y_1^j & \dots & y_1^n \\ \dots & \dots & \dots & \dots & \dots \\ y_p^1 & \dots & y_p^j & \dots & y_p^n \\ z_1^1 & \dots & z_1^j & \dots & z_1^n \\ \dots & \dots & \dots & \dots & \dots \\ z_{m-p}^1 & \dots & z_{m-p}^j & \dots & z_{m-p}^n \end{bmatrix}$$

Clustering flou : Substractive clustering

- Algorithme de clustering



1) k=0

$$P_i = \sum_{j=1}^{n} e^{-\alpha\|x_i - x_j\|^2}, \alpha = \frac{4}{r_a^2}$$

2) k=k+1

Choisir le point avec le P_i le plus grand

Ce point devient un centre :

Coordonnées : x_k^*, et de potentiel : P_k^*

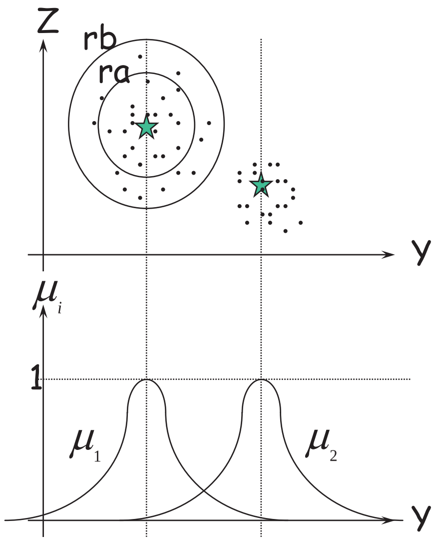
3) Modifier les P_i :

$$P_i = P_i - P_k^* e^{-\beta\|x_i - x_k^*\|^2}, \beta = \frac{4}{r_b^2}$$

4) Répéter 2 et 3 jusqu'à : $P_k^* < \varepsilon P_1^*$

Clustering flou : Substractive clustering

- Identification du modèle mathématique



$$\mu_i(Y) = e^{-\alpha\|Y - Y_i^*\|^2}$$

$$Z(Y) = \frac{\sum_{i=1}^{c} \mu_i(Y) Z_i^*}{\sum_{i=1}^{c} \mu_i(Y)}$$

c : nombre de clusters

Clustering flou : Substractive clustering

- Identification du SIF

  - A chaque centre X_k^* est associé une règle floue

    Si y_1^k est A_1^k et y_2^k est A_2^k et … alors z_1^k est B_1^k et z_2^k est B_2^k et …

  - Pour la kème règle associée à X_k^*

    Les A_i^k sont des fonctions d'appartenance exponetielle $A_i^k(q) = e^{-\alpha\|q - y_i^{k*}\|^2}$

    les B_i^k sont des singletons $B_i^k = z_i^{k*}$

Clustering flou : Substractive clustering

- Optimisation d'ordre 0

$$z_i^* = G_i y + h_i$$

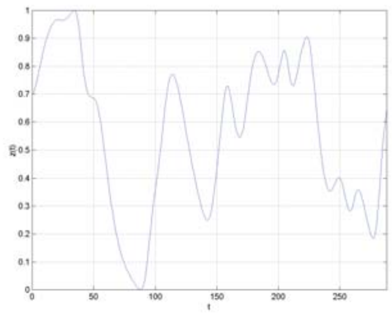
- Optimisation d'ordre 1

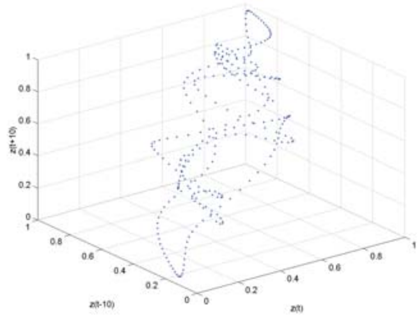
$$z_i^* = h_i$$

Clustering flou : Substractive clustering

- Un exemple d'application : prévision de série chronologique

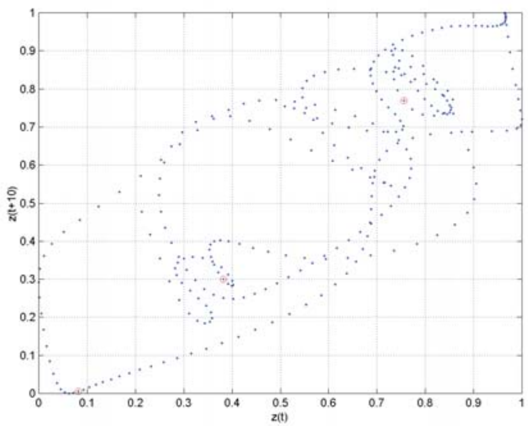
Première étape : pré traitement des données





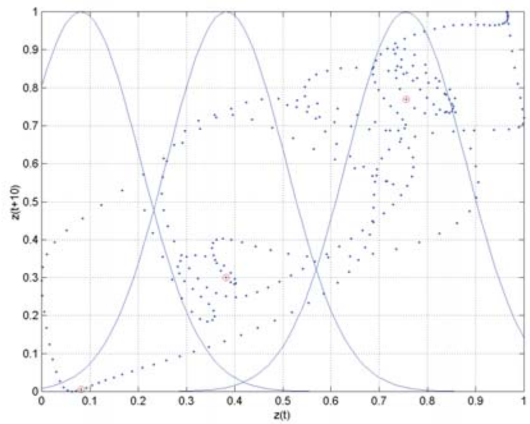
Clustering flou : Substractive clustering

Deuxième étape : Clustering



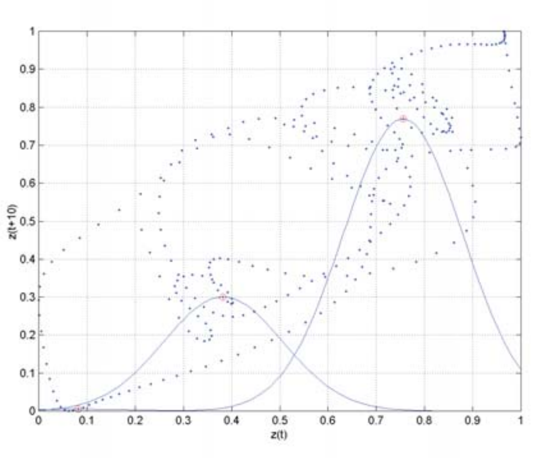
Clustering flou : Substractive clustering

Deuxième étape : Clustering



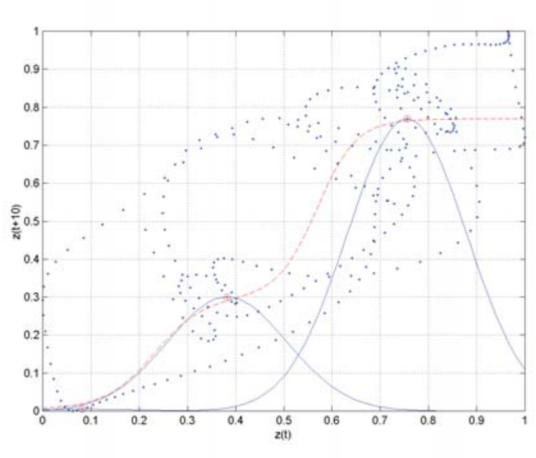
Clustering flou : Substractive clustering

Troisième étape : Modélisation



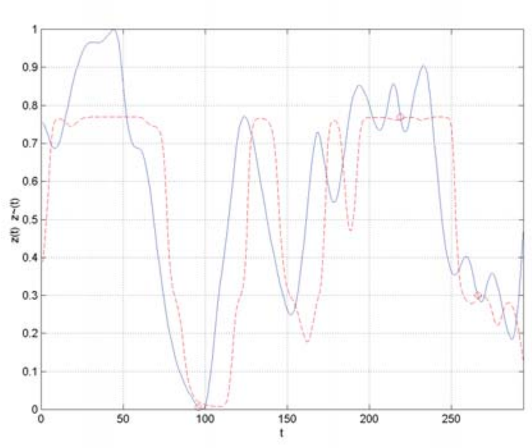
Clustering flou : Substractive clustering

Troisième étape : Modélisation



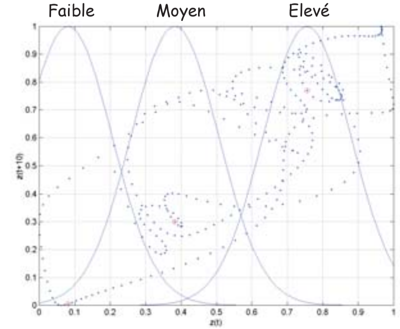
Clustering flou : Substractive clustering

Résultats



Clustering flou : Substractive clustering

Quatrième étape : identification du SIF



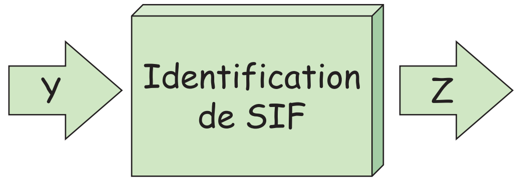
SI z(t) est Faible ALORS z(t+10) est 0
SI z(t) est Moyen ALORS z(t+10) est 0.3
SI z(t) est Elevé ALORS z(t+10) est 0.78

Clustering flou : Substractive clustering

- Algorithme efficace et peu complexe mais :
  - Les paramètres $\{r_a, r_b, \varepsilon\}$ sont difficiles à estimer
  - Les clusters sont tous de la même forme
  - Les clusters sont tous de la même taille

Clustering flou

- Méthode de Clustering
  - Méthode : « Substractive clustering » (SC)
  - Méthode : « Fuzzy C-means » (FCM)



Clustering flou : Fuzzy C-means

- Algorithme inspiré du k-means
- On considère l'espace de n points de dimension p suivant :

$$X : \begin{array}{c|ccccc|} & X^1 & \dots & X^j & \dots & X^p \\ \hline X_1 & x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_k & x_k^1 & \dots & x_k^j & \dots & x_k^p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_n & x_n^1 & \dots & x_n^j & \dots & x_n^p \\ \hline \end{array}$$

- On suppose que les n points peuvent être groupés en c clusters c < n
- Les clusters sont déterminés par leurs centres

$$V_i = [v_i^1, v_i^2, \dots, v_i^j, \dots, v_i^p], 1 \le i \le c$$

Clustering flou : Fuzzy C-means

- On considère la matrice de proximité suivante :

$$U = \begin{bmatrix} u_{11} & \dots & u_{1k} & \dots & u_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ u_{i1} & \dots & u_{ik} & \dots & u_{in} \\ \dots & \dots & \dots & \dots & \dots \\ u_{c1} & \dots & u_{ck} & \dots & u_{cn} \end{bmatrix}, \text{avec } k = 1, \dots, n \text{ et } i = 1, \dots, c$$

u_{ik} représente le degré d'appartenance du point X_k au centre V_i

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right]^{-1}, d_{ik} \text{ représente la distance entre } V_i \text{ et } X_k$$

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\|X_k - V_i\|}{\|X_k - V_j\|} \right)^{\frac{2}{m-1}} \right]^{-1}, \text{si on choisit la distance euclidienne}$$

Clustering flou : Fuzzy C-means

- Algorithme :

1) Initialiser la position des centres : $V_i = [v_i^1, v_i^2, \ldots, v_i^j, \ldots, v_i^p], 1 \le i \le c$

   l=0, initialiser la matrice : $U^{(l)}$

2) Calculer la position des centres $V_i = \dfrac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}, 1 \le i \le c$

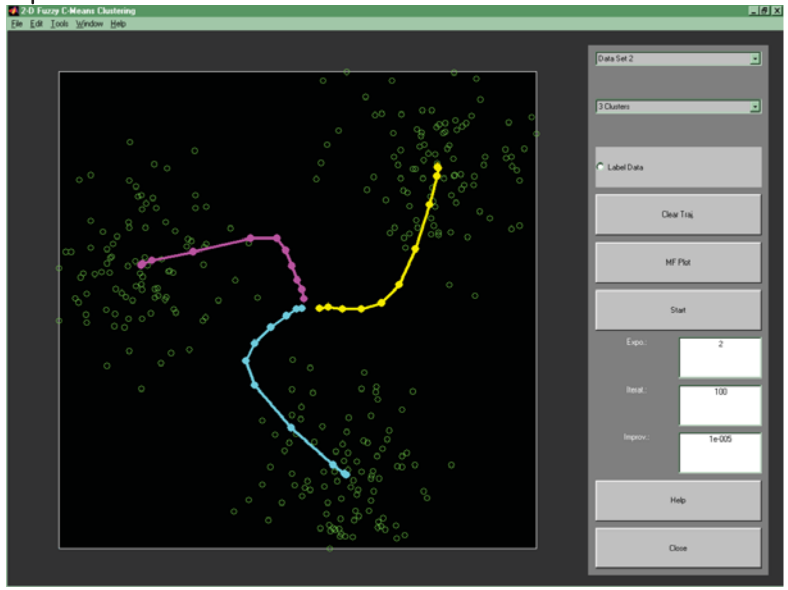
3) Calculer la matrice : $U^{(l+1)}$

4) Si $\left\| U^{(l+1)} - U^{(l)} \right\| < \varepsilon$ Arrêter l'algorithme

   Sinon l=l+1 et retourner en 2)

Clustering flou : Fuzzy C-means

- Exemple :



Clustering flou : Fuzzy C-means

- Algorithme simple mais :
  - L'initialisation des positions des centres conditionne la convergence et le résultat final
  - Le résultat dépend de la métrique utilisée pour calculer la distance
  - Le nombre de clusters c doit être fixé a priori

Applications de la logique floue

- Systèmes experts
- Modélisation de processus non-linéaires
- Contrôle/Commande de processus
- Classification
- Prévision de séries chronologiques

Conclusion sur la logique floue

- **Avantages**
  - Expression des connaissances sous forme de règles
  - Un modèle analytique n'est pas nécessaire
- **Inconvénients**
  - Difficulté d'obtention des règles
  - Risque d'explosion combinatoire

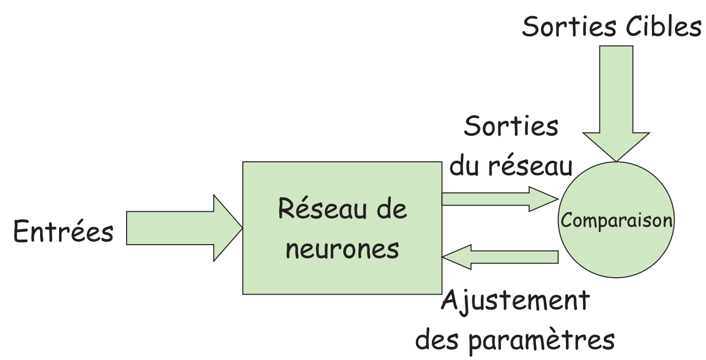
Organisation du cours

Réseaux de neurones artificiels

- Concepts de base :
  - Imiter la structure et le fonctionnement du cerveau humain
  - Trois des caractéristiques fondamentales du cerveau :
    - La connaissance est distribuée sur plusieurs neurones
    - Les neurones peuvent communiquer avec leurs voisinages
    - Le cerveau est adaptatif
  - La terminologie utilisée pour les réseaux de neurones artificiels est issue de ces caractéristiques :
    - Structure d'un neurone
    - Topologie d'un réseau
    - Règle d'adaptation ou d'apprentissage

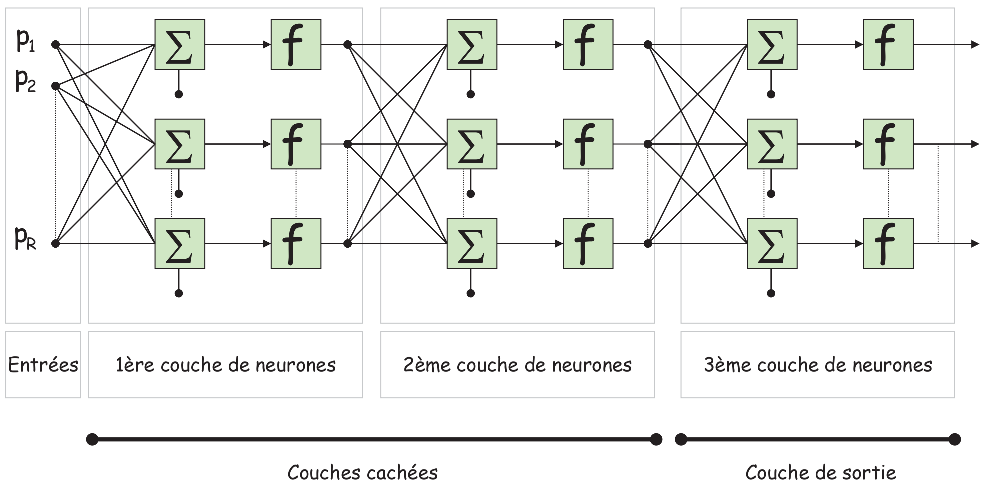
Réseaux de neurones artificiels

- Principe général



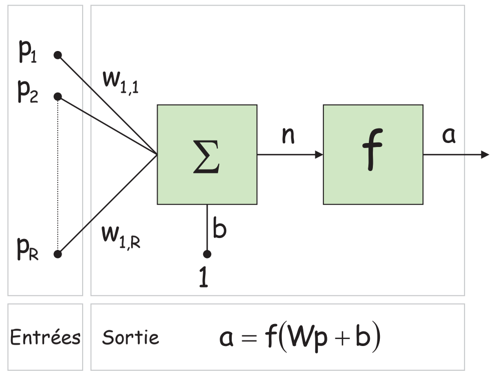
Réseaux de neurones artificiels

- Structure d'un réseau multi-couches feedforward



Réseaux de neurones artificiels

- Structure et fonctionnement d'un neurone



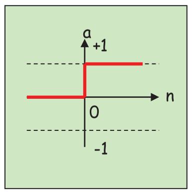
R : nombre d'entrées

$$p = \begin{bmatrix} p_1 \\ p_2 \\ \dots \\ p_R \end{bmatrix}$$

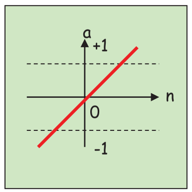
$$W = \lfloor w_{1,1}, w_{1,2}, \dots, w_{1,R} \rfloor$$

Réseaux de neurones artificiels

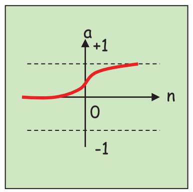
- Fonctions d'activation classiques :



Hard limit



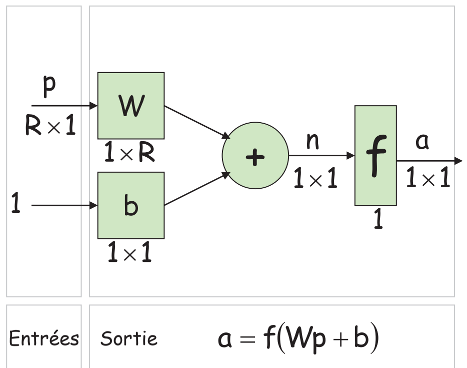
Linéaire



Log-Sigmoid

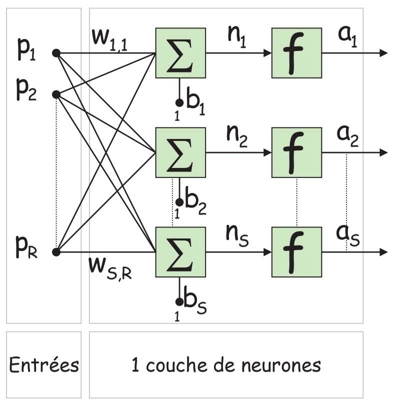
Réseaux de neurones artificiels

- Notation



Réseaux de neurones artificiels

- Structure d'une couche de neurones

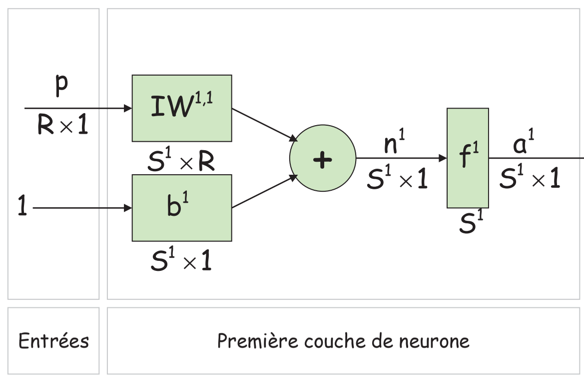


$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & & w_{1,R} \\ w_{2,1} & w_{2,2} & & w_{2,R} \\ & & & \\ w_{S,1} & w_{S,2} & & w_{S,R} \end{bmatrix}$$

R : nombre d'entrées
S: nombre de neurones

Réseaux de neurones artificiels

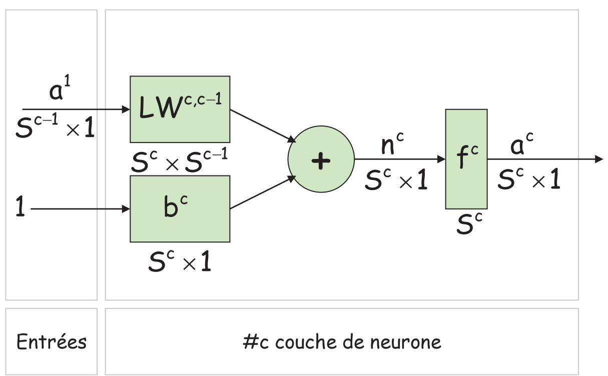
- Fonctionnement de la première couche



$$a^1 = f^1\left(IW^{1,1}p + b^1\right)$$

Réseaux de neurones artificiels

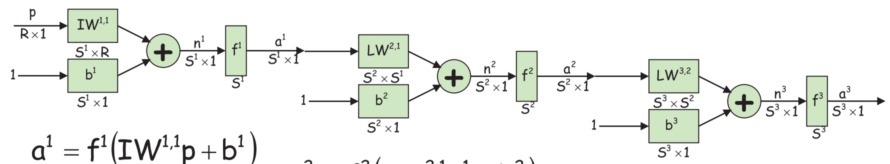
- Fonctionnement d'une couche c



$$a^c = f^c\left(LW^{c,c-1}a^{c-1} + b^c\right)$$

Réseaux de neurones artificiels

- Fonctionnement d'un réseau multi-couches feedforward



$$a^1 = f^1\left(IW^{1,1}p + b^1\right)$$

$$a^2 = f^2\left(LW^{2,1}a^1 + b^2\right)$$

$$a^3 = f^3\left(LW^{3,2}a^2 + b^3\right)$$

$$y = a^3 = f^3\left(IW^{3,2}f^2\left(LW^{2,1}f^1\left(IW^{1,1}p + b^1\right) + b^2\right) + b^3\right)$$

Réseaux de neurones artificiels

- Algorithme d'apprentissage : « Backpropagation »
  - Principes
    - Appliquer un vecteur d'entrée
    - Calculer la sortie
    - Modifier les poids des neurones pour minimiser l'erreur en partant des neurones de sortie (Backpropagation)

$$E = \frac{1}{2}\sum_i (y_i - o_i)^2$$

y_i : sorties cibles

o_i : sorties calculées

Réseaux de neurones artificiels

- Algorithme d'apprentissage : « Backpropagation »

$w_{j,i} = w_{j,i} + \Delta w_{j,i}$

$\Delta w_{j,i} = -\alpha\, \partial_j\, o_i$

α : learning rate

$$\partial_j = \begin{cases} f'(e_j) \times (t_j - s_j) & \text{si l'unité j est une unité de sortie} \\ f'(e_j) \times \left(\sum_k \delta_k w_{k,j}\right) & \text{si l'unité j est une unité cachée} \end{cases}$$

j : indice de l'unité courante

i : indice de l'unité précédente de j avec un poids $w_{j,i}$ de i à j

k : indice de l'unité suivante de j avec un poids $w_{k,j}$ de j à k

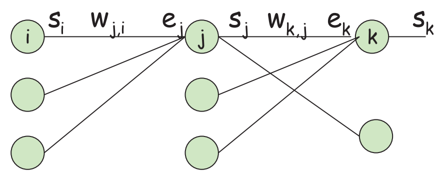
s_j : sortie de l'unité j

e_j : entrée de l'unité j

δ_j : erreur de l'unité j

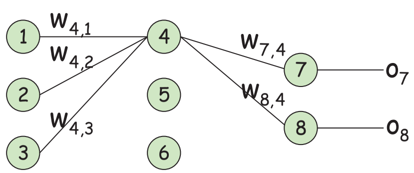
α : facteur d'apprentissage

t_j : valeur cible de l'unité de sortie j



Réseaux de neurones artificiels

- Algorithme d'apprentissage « Backpropagation » : exemple



$\Delta w_{7,4} = \alpha \delta_7 s_4$ $\quad$ $\Delta w_{8,4} = \alpha \delta_8 s_4$

$e_7 = s_4 w_{7,4}$ $\quad$ $e_8 = s_4 w_{8,4}$

$\delta_7 = f'(e_7) \times (t_7 - s_7)$ $\quad$ $\delta_8 = f'(e_8) \times (t_8 - s_8)$

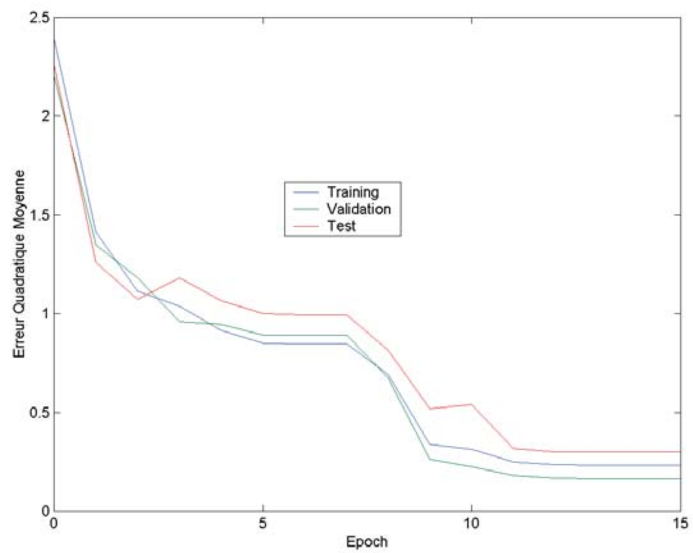
$$e_4 = s_1 w_{4,1} + s_2 w_{4,2} + s_3 w_{4,3}$$

$$\delta_4 = f'(e_4) \times (\delta_7 w_{7,4} + \delta_8 w_{8,4})$$

$$\Delta w_{4,1} = \alpha \delta_4 s_1$$

Réseaux de neurones artificiels

- Un exemple d'apprentissage sur la base des iris



Applications des réseaux de neurones artificiels

- Reconnaissance de formes
- Modélisation de processus non-linéaires
- Contrôle/Commande de processus
- Classification
- Prévision de séries chronologiques

Conclusion sur les réseaux de neurones artificiels

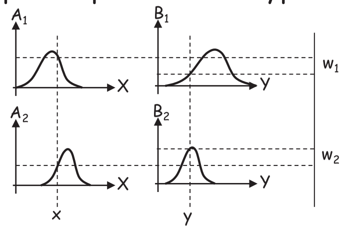
- **Avantages**
  - Approximateur universel
  - Modèle non-linéaire
  - Champ d'application étendu
- **Inconvénients**
  - Difficultés de mise en œuvre
  - Problème de convergence
  - Approche boîte noire

Une méthode hybride : ANFIS

- ANFIS : « Adaptive-Network-Based Fuzzy Inference System »
  - Combinaison des approches logique floue et réseaux de neurones
  - L'objectif est d'optimiser un ensemble de règles floues par apprentissage
  - L'idée principale est de considérer un FIS comme un réseau de neurones spécialisé et d'appliquer des règles classiques d'apprentissage

Une méthode hybride : ANFIS

- Un exemple simple : FIS de type SUGENO avec 2 règles

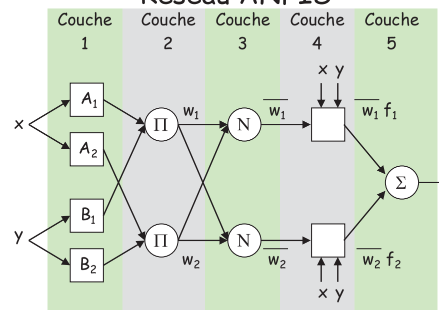


$f_1 = p_1 x + q_1 y + r_1$

$f_2 = p_2 x + q_2 y + r_2$

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$$

Réseau ANFIS



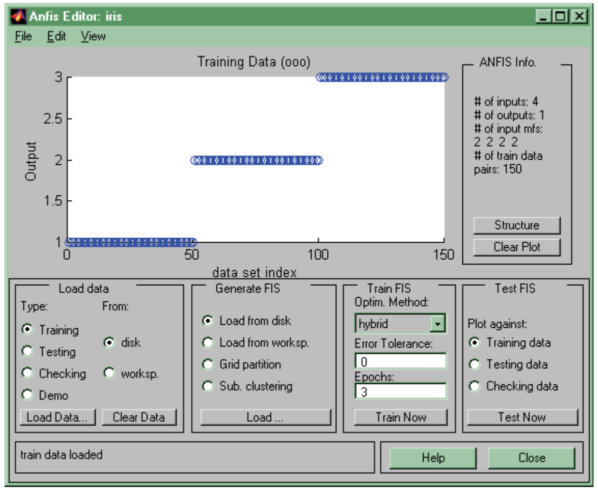
Couche 1 : application de la fonction d'appartenance
Couche 2 : Application du ET flou (multiplication)
Couche 3 : Moyenne des sortie du ET
Couche 4 : Calcul de la conclusion de chaque règle (type SUGENO)
Couche 5 : Agrégation des règles

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$$

Par une méthode d'apprentissage les paramètres : w_1, w_2, p_1, p_2, q_1, q_2, r_1, r_2 peuvent être modifiés pour s'adapter à un jeu de données

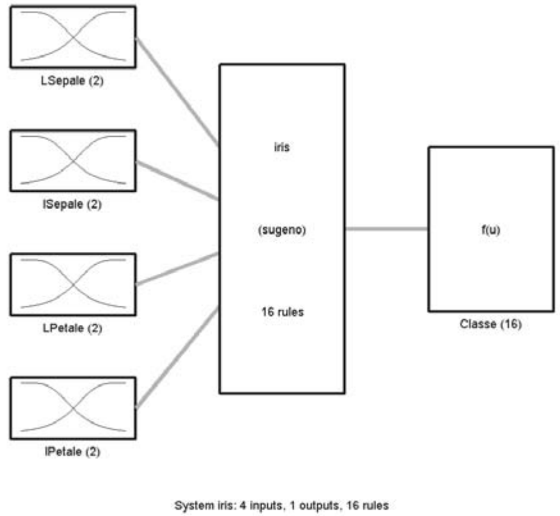
Une méthode hybride : ANFIS

- Un exemple : classification des iris



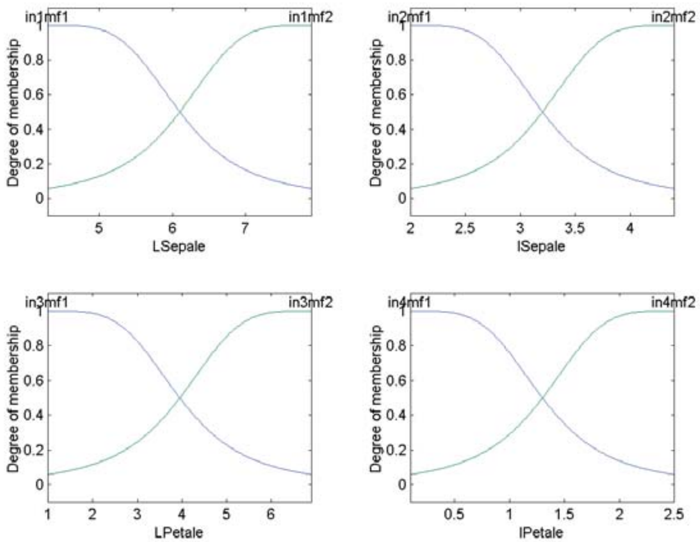
Une méthode hybride : ANFIS

- Un exemple : classification des iris



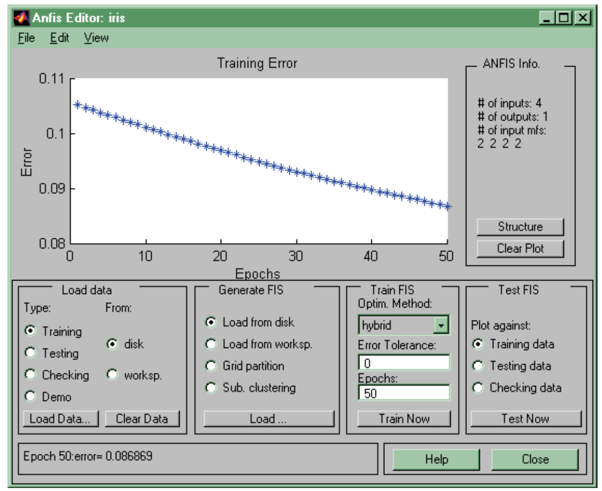
Une méthode hybride : ANFIS

- Un exemple : classification des iris



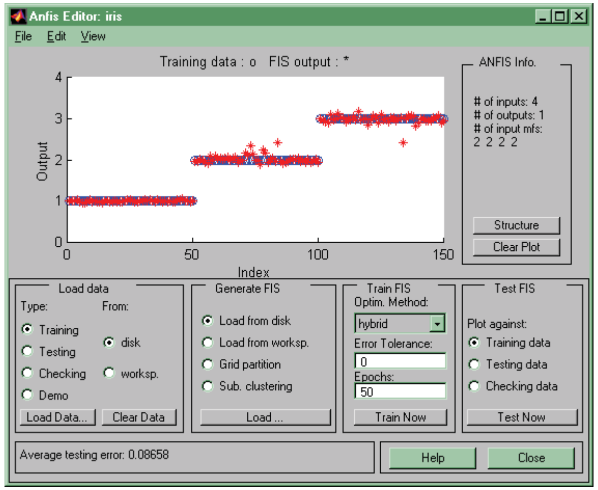
Une méthode hybride : ANFIS

- Un exemple : classification des iris



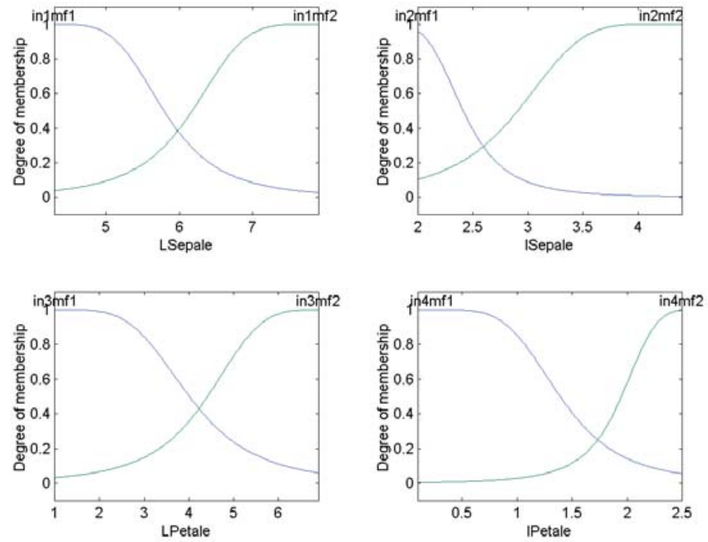
Une méthode hybride : ANFIS

- Un exemple : classification des iris



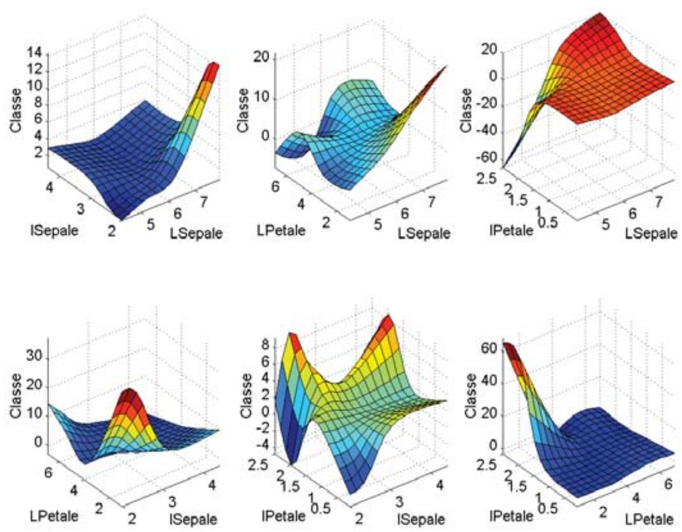
Une méthode hybride : ANFIS

- Un exemple : classification des iris



Une méthode hybride : ANFIS

- Un exemple : classification des iris



Organisation du cours

Introduction
Bases du KDD et du DM

Bases de l'analyse des données
Analyse factorielle
Classification
Clustering
Logique floue
Réseaux de neurones artificiels

Conclusion

Sources

- Denis, François et Gilleron, Rémi.
  http://www.grappa.univ-lille3.fr/polys/apprentissage/index.html.

- Tommassi, Marc et Gilleron, Rémi.
  http://www.grappa.univ-lille3.fr/polys/fouille/index.html.

Bibliographie

- Statistiques et analyses de données
  - « Probabilité, analyse des données et statistique », G. Saporta, éditions Technip
  - « Analyse des données », M. Volle, Economica
  - « Le data mining », René Lefébure et Gilles Venturi, Eurolles
  - Cornuéjols A. & Miclet L. (2010) Apprentissage artificiel Concepts et algorithmes. - Editeur : Eyrolles - Collection : Algorithmes - 03/06/2010 (2e édition) - EAN13 : 9782212124712

- Logique Floue et réseaux de neurones
  - « Fuzzy Logic & Neural Network Handbook », C.H. Chen McGraw-Hill
  - « La logique floue et ses applications », Bernadette Bouchon Meunier Addison-Wesley
  - « La logique floue », Bernadette Bouchon Meunier, Que sais-je ?
  - « Réglage par logique floue », Hansruedi Bühler Presses Polytechniques et Universitaires Romandes
  - « ANFIS:Adaptive-Network-Based Fuzzy Inference system », Jyh-Shing Roger Jang IEEE Trans. On Systems, Man, and Cybernetics, Vol. 23, N°3, May/June 1993
  - « Fuzzy Model Identification Based on Cluster Estimation », Stephen L. Chiu, Journal of Intelligent and Fuzzy Systems, Vol 2, p. 267-278, 1994