Institut
Mines-Télécom

SI221 Bases de l'Apprentissage
Introduction

G. Varni (giovanna.varni@telecom-paristech.fr)

5 février 2019

Equipe pédagogique

- **Coordinatrices:**

  Laurence Likforman, Chloé Clavel, Giovanna Varni (IDS)

- **Assistant:**

  Emile Chapuis (doctorant IDS)

Contacts

Giovanna Varni
giovanna.varni@telecom-paristech.fr
Bureau: B602

Emile Chapuis
emile.chapuis@telecom-paristech.fr
Bureau: B505-3

A propos du cours…(1/6)

- **Apprentissage actif!**

« Dis-moi et j'oublierai. Montre-moi, et je ne me souviendrai peut-être pas. Implique-moi, et je comprendrai » - Proverbe amérindien

Donc…un apprentissage centré sur l' étudiant!

Conditions de succès

S'engager à : parler / écouter ; lire ; réfléchir

A propos du cours…(2/6)

- **Comment?**

  Par la classe inversée!

A propos du cours…(3/6)

- **Concrètement**

  Dans l' emploi du temps:

  - Etude individuel à la maison (slides, poly, vidéos…)
  - TP/TD en classe avec superviseur et en autonomie
  - Cours de restructuration / temps étude individuel ou en groupe

Important!

Présence obligatoires aux TPs!

TP en MATLAB, pas de binômes

Note finale: TPs (5 points) + Contrôle écrit (15 points)

A propos du cours…(4/6)

Emploi du temps

05/02/2019	mardi	13:30	15:00		Leçon	Introduction
05/02/2019	mardi	15:15	16:45		Leçon	Introduction
12/02/2019	mardi	13:30	15:00		Travaux pratiques	TP Bayes
12/02/2019	mardi	15:15	16:45		Travaux pratiques	TP Bayes
19/02/2019	mardi	13:30	15:00		Leçon	Cours restructuration Bayes / étude individuel
19/02/2019	mardi	15:15	16:45		Travaux dirigés	TD Bayes
05/03/2019	mardi	13:30	15:00		Travaux pratiques	TP Neurones
05/03/2019	mardi	15:15	16:45		Travaux pratiques	TP Neurones
12/03/2019	mardi	13:30	15:00		Leçon	Cours restructuration Neurones/ étude individuel
12/03/2019	mardi	15:15	16:45		Travaux pratiques	TP Kppv
26/03/2019	mardi	13:30	15:00		Travaux pratiques	TP Classification automatique
26/03/2019	mardi	15:15	16:45		Travaux pratiques	TP Classification automatique
02/04/2019	mardi	13:30	15:00		Leçon	Cours restructuration classification automatique / étude individuel
02/04/2019	mardi	15:15	16:45		Travaux pratiques	TP Markov
09/04/2019	mardi	13:30	15:00		Leçon	Cours restructuration Markov / étude individuel
09/04/2019	mardi	15:15	16:45		Contrôle de connaissance	

A propos du cours…(5/6)

- **En particulier**

TPs: consigne sur le site pédagogique (avant minuit)

Donc:

24/02 TP Bayes

17/03 TP Neurones

22/03 TP Kppv

6/04 TP Classification automatique

14/04 TP Markov

Cours de restructuration: pas de questions, pas de cours.

-> Temps pour l' étude individuel ou en groupe.

Questions à envoyer par email au moins **4 jours** avant le cours

A propos du cours…(6/6)

- **Matériel**

Site pédagogique du cours:

https://sitepedago.telecom-paristech.fr/front/frontoffice.php?SP_ID=2576&#R2059

Ressources:

- polycopié (version imprimé disponible)
- **slides**
- exercices corrigés
- annales
- vidéos :

https://drive.google.com/drive/folders/13oc7nopTw8ECIBwZ0aIambc2ciPkbfXn?usp=sharing

Concepts de base

- **Reconnaissance de formes**

  En anglais:

  machine learning / pattern recognition / pattern classification

  Reconnaitre : le fait d'identifier un objet, un être comme tel

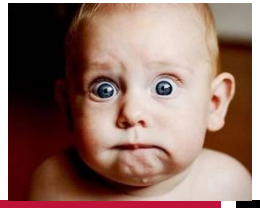
  Qu'est ce qu' une forme?

  Forme n. f. : A. Apparence, aspect visible.

  1) ... 2) apparence extérieure donnant à un objet ou à un être sa spécificité

  Exemples de formes :

  image , parole , texte, empreintes digitales…



Reconnaissance de formes (RdF)

- **Reconnaissance de formes (RdF)**

  Discipline scientifique qui concerne la déscription et la reconnaissance (classification) de formes

Définitions de l'etat de l'art:

« The assignement of a physical object or event to one of several pre-specified categories » - Duda and Hart

« A problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories or classes » - Fukunaga

« The process of giving names to observations X » - Schalkoff

« Pattern recognition is concerned with answering the question: what is this ? » - Morse

Objectif

- **Objectif:** doter les machines des capacités de l'homme (la plus parfaite machine de reconnaissance!) à reconnaitre des caractères, des objets, des sons…

La RdF consiste donc à étudier comment une machine peut:

- apprendre à **extraire** des structures d' intérêt ;
- **prendre de décisions** en observant un environnement ;
- **reconnaitre, décrire ou classifier** des formes

Au début, la RdF était plutôt traitement du signal:

- présence /absence d' un signal
- identification de sources multiples

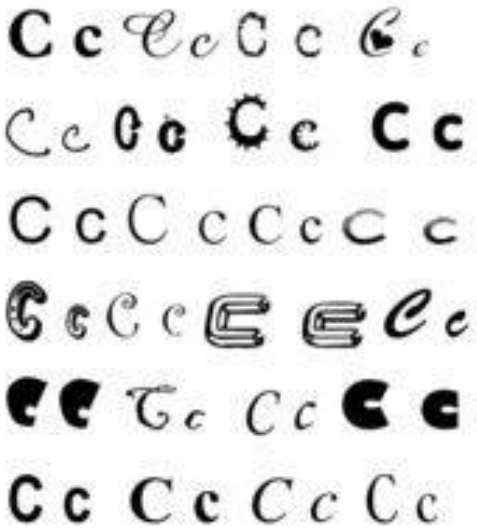
progressivement…

- reconnaissance de visage, sons, objets…

…independamment

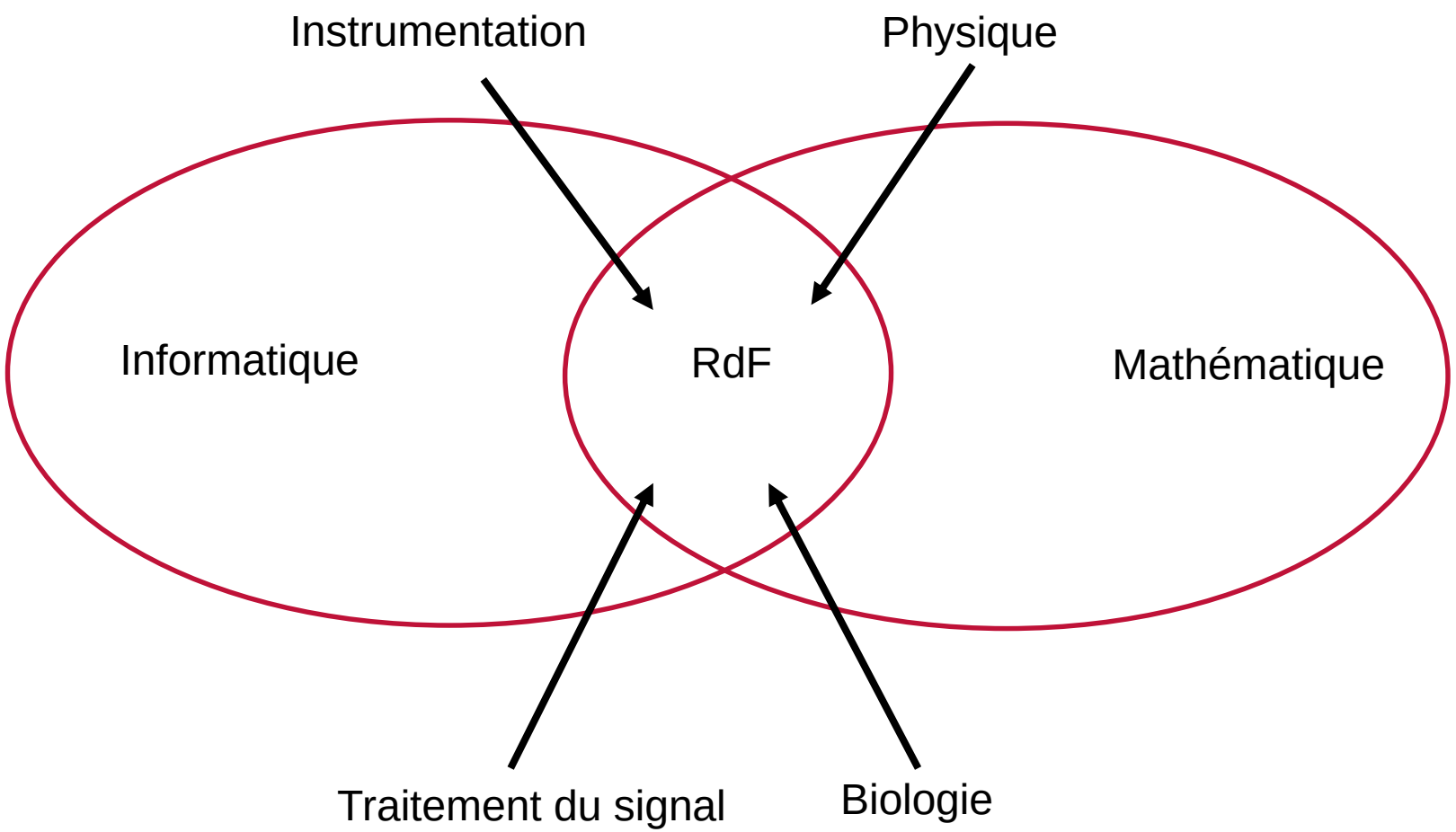
Des conditions d'observation

De leur variabilite'



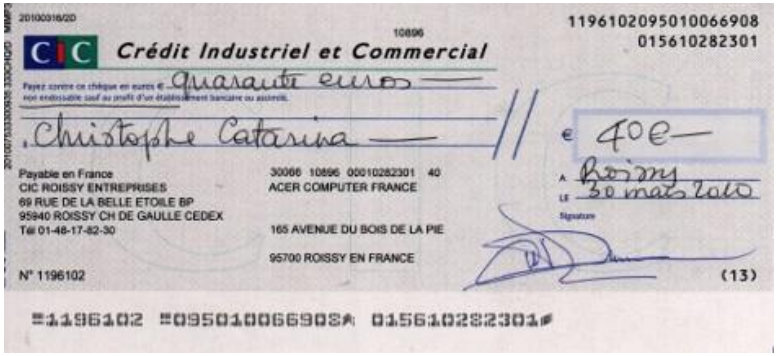
Du point de vue sous lesquels on les observe

Domaines connexes



Applications : reconnaissance de textes

Tri postal

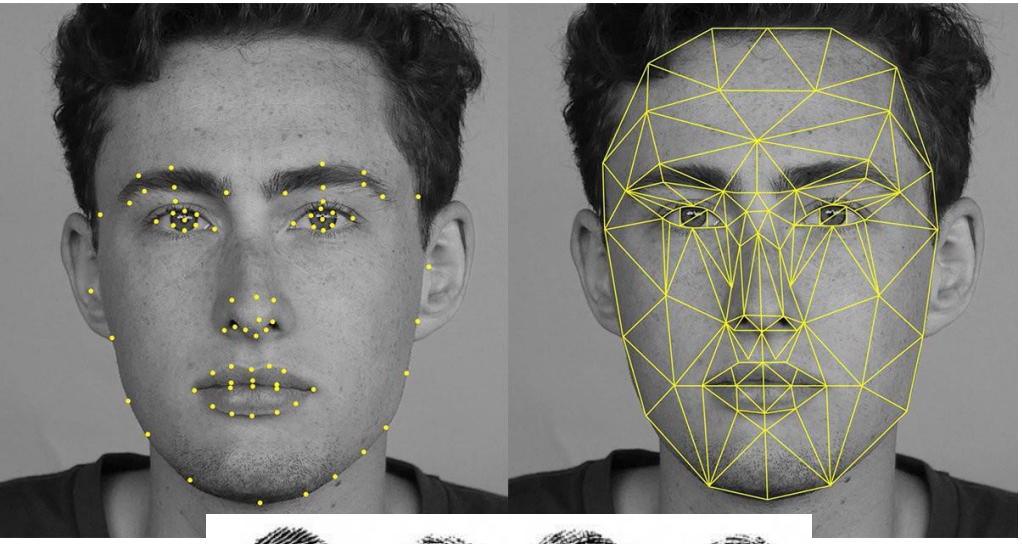


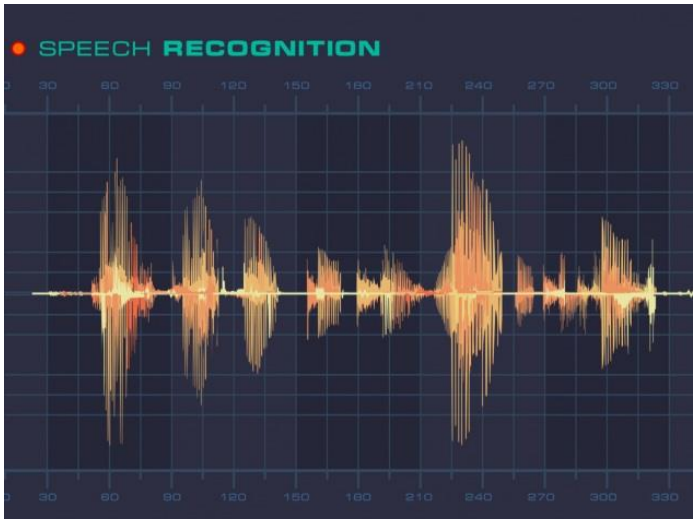
Lecture de chèques

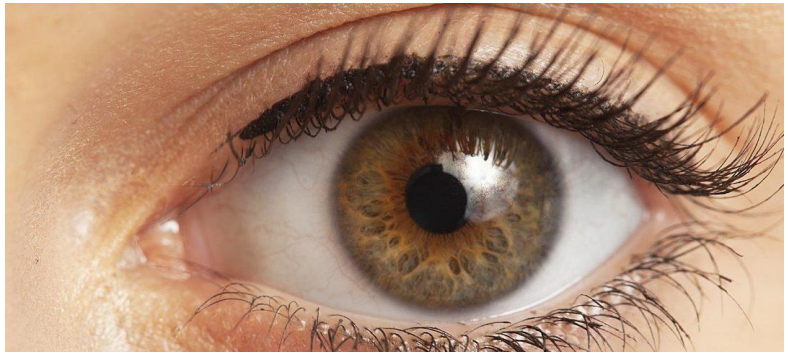
Lecture panneaux

Stylo multilanguage

Applications : biométrie

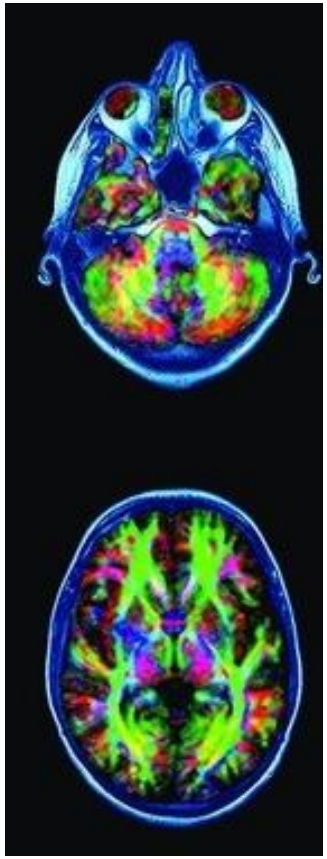




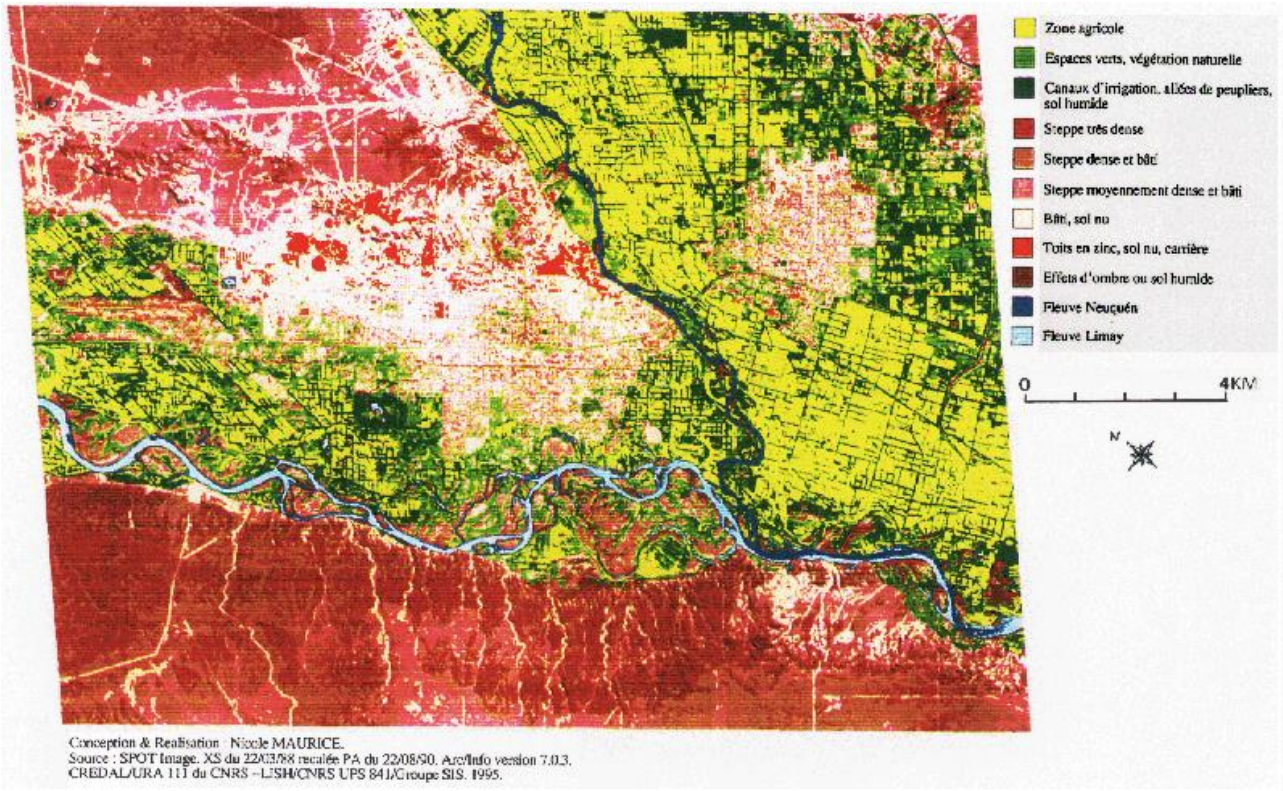


Applications : imagérie

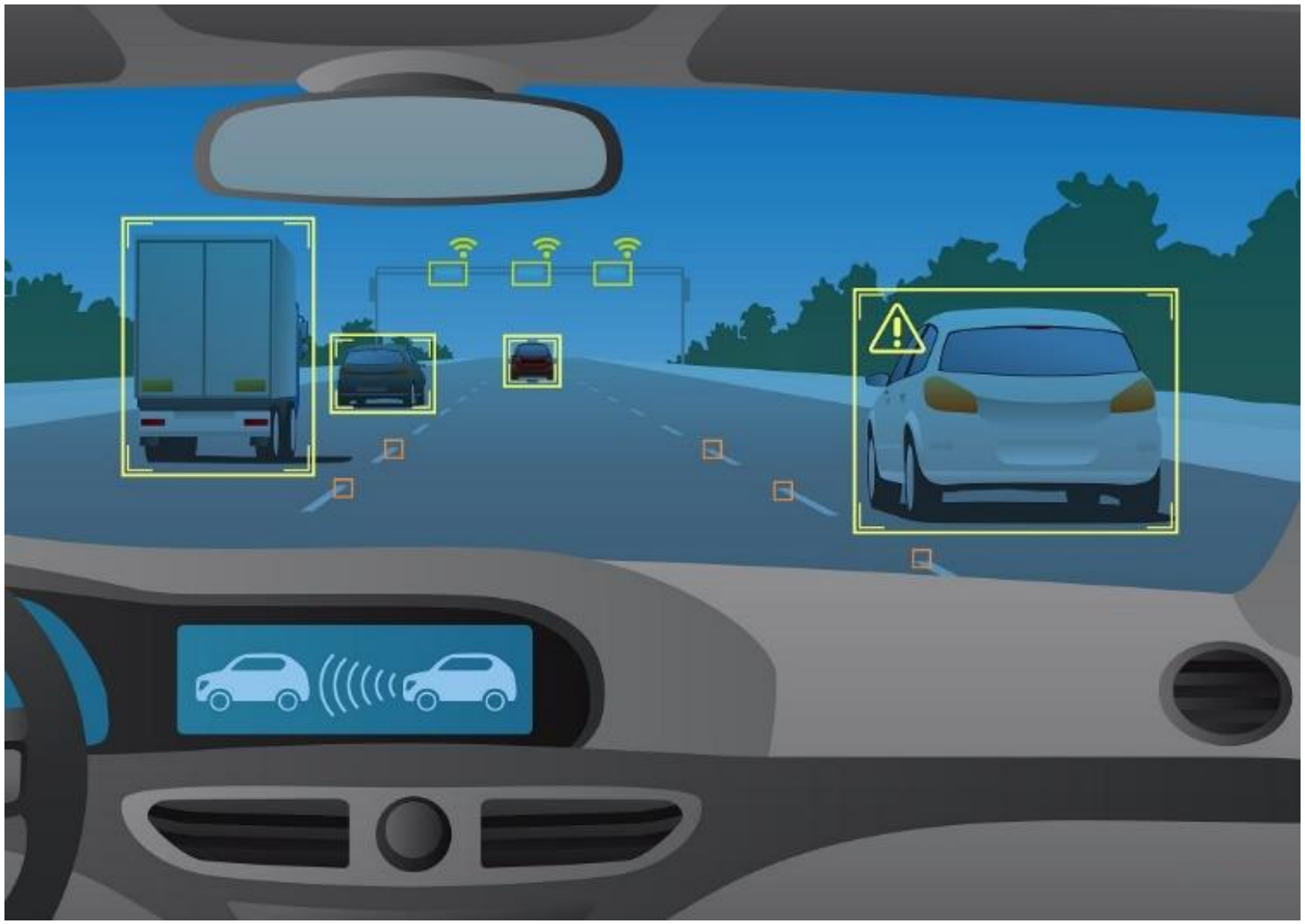
Medicale



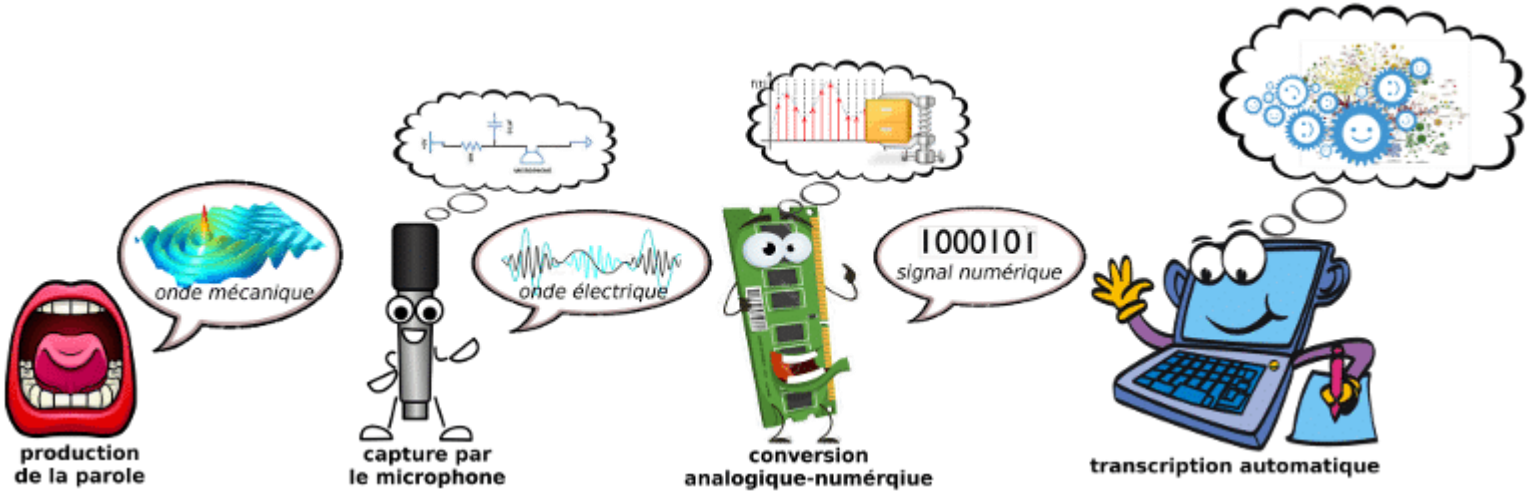
Satellitaire



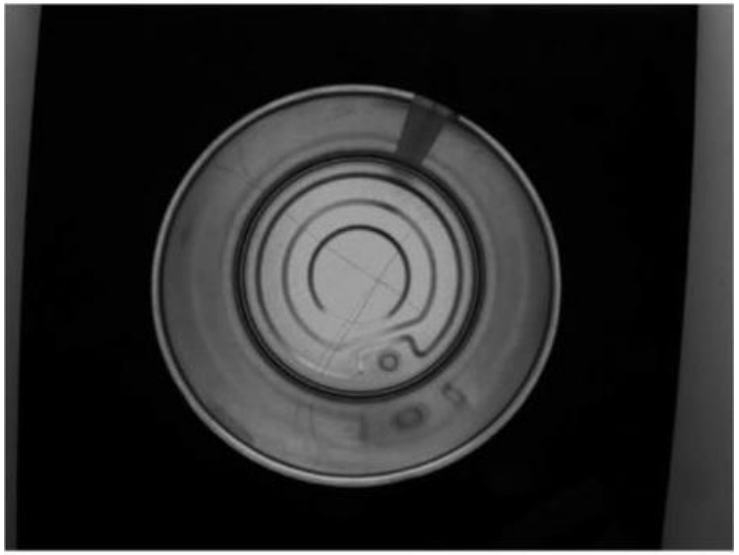
Applications : automotive



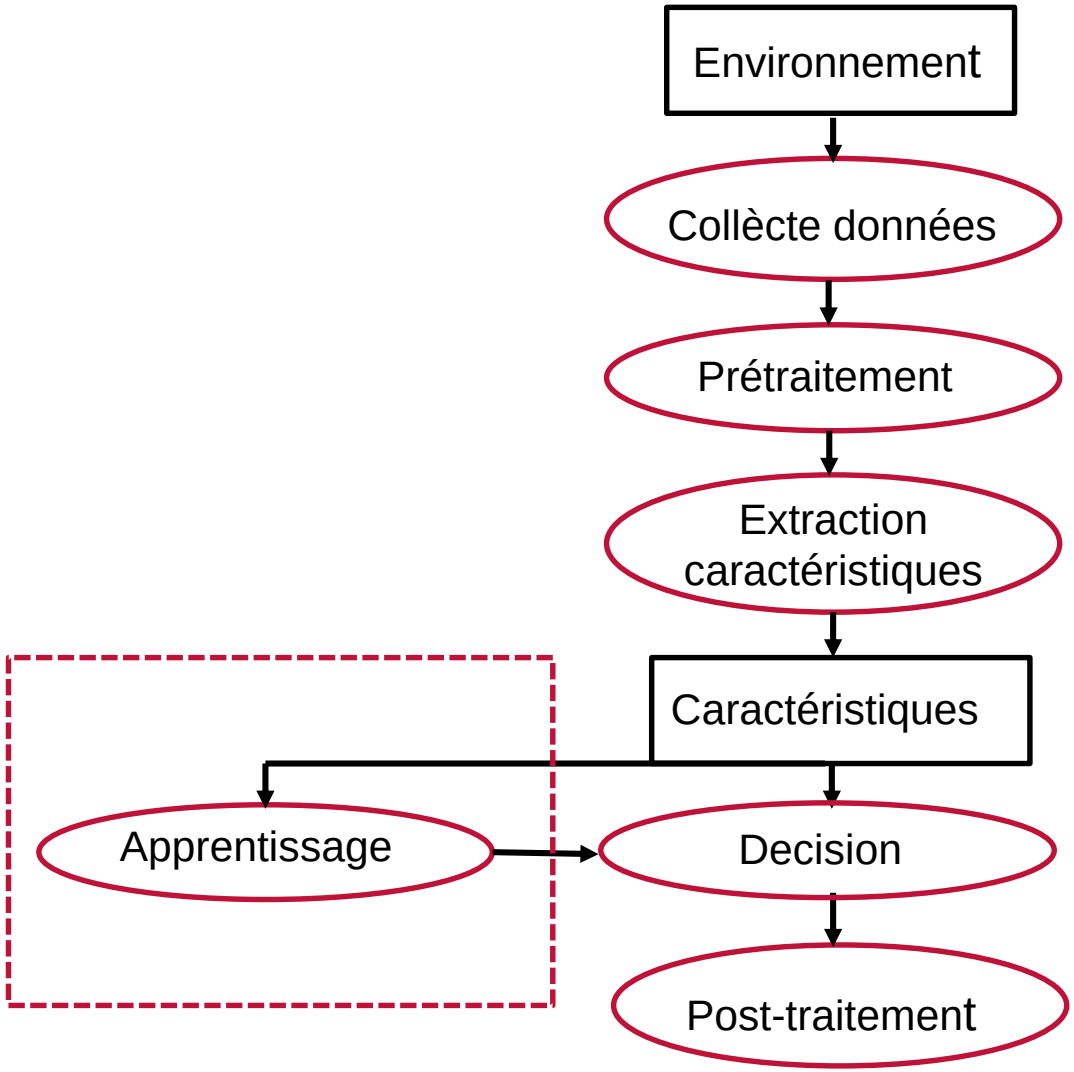
Applications : reconnaissance parole



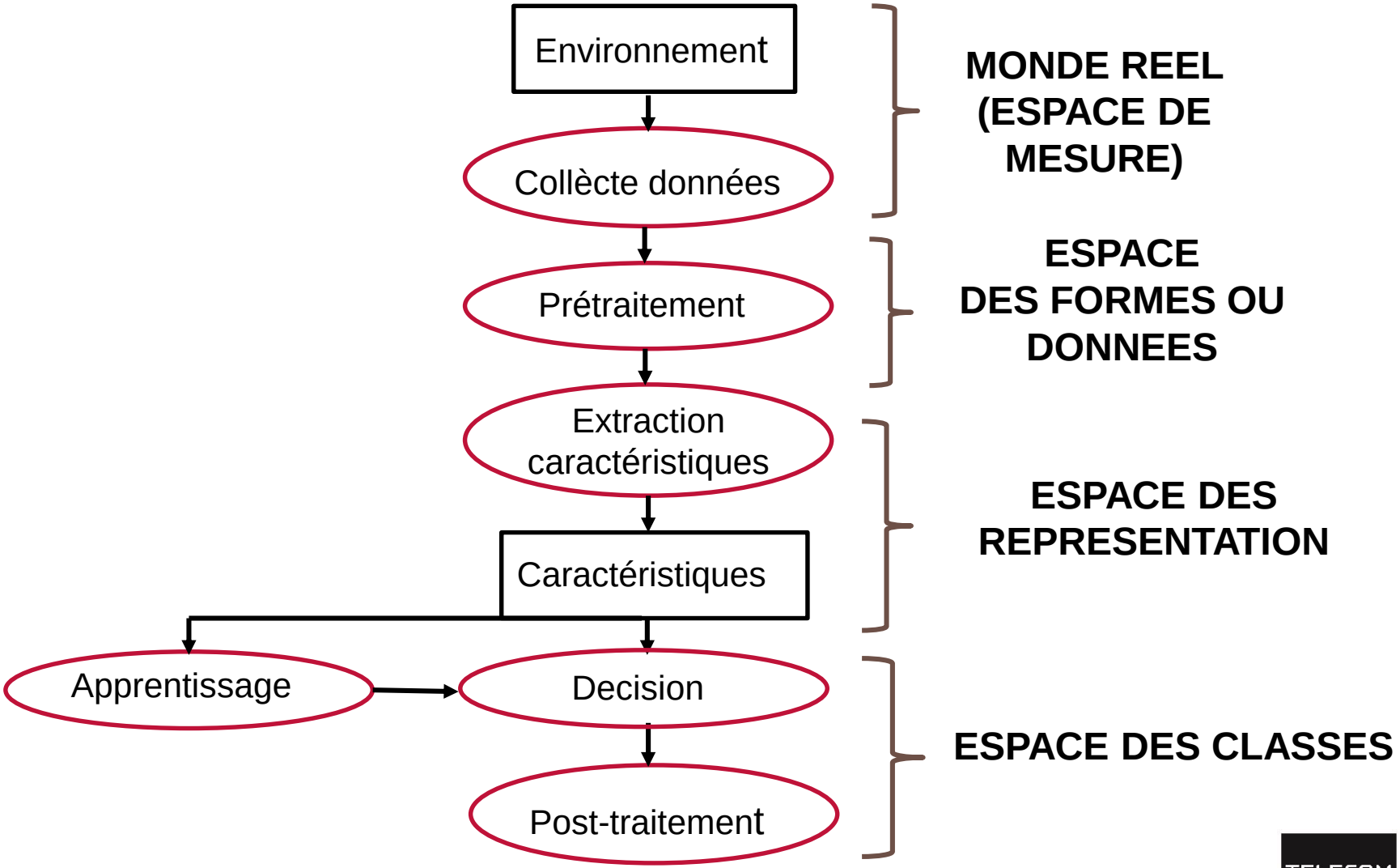
Applications : contrôle de qualité



Un système de reconnaissance de formes



Un système de reconnaissance de formes



Un exemple jouet : reconnaissance de poissons

- **Reconnaitre les truites de saumons**

Un exemple jouet : collecte de données et prétraitement

- **Utiliser une camera pour collecter des images**
- **Extraire la forme (le poisson) de l'images**

Un exemple jouet : extraction des caractéristiques

- **Codage: extraire de la forme des caractéristiques**

  ou **mesures**

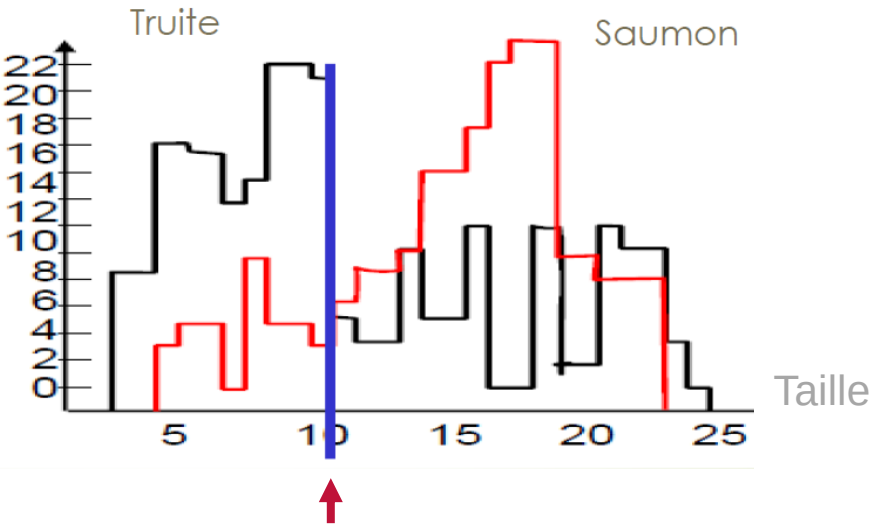
  ou **features**

E.g. dans 1 dimension : la taille

la largeur

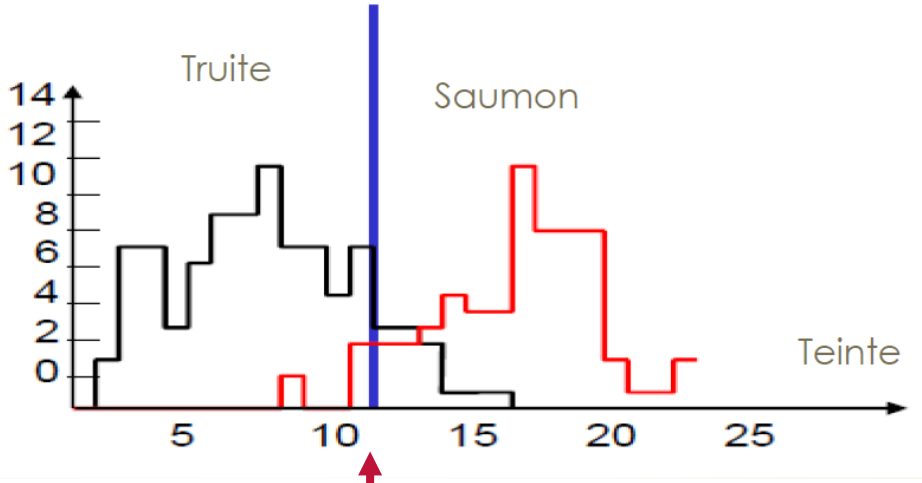
nombre et forme de nageoires

Un exemple jouet : la reconnaissance



Seuille de decision

Pas mal de chevauchement…
Decision pas robuste, que faire?

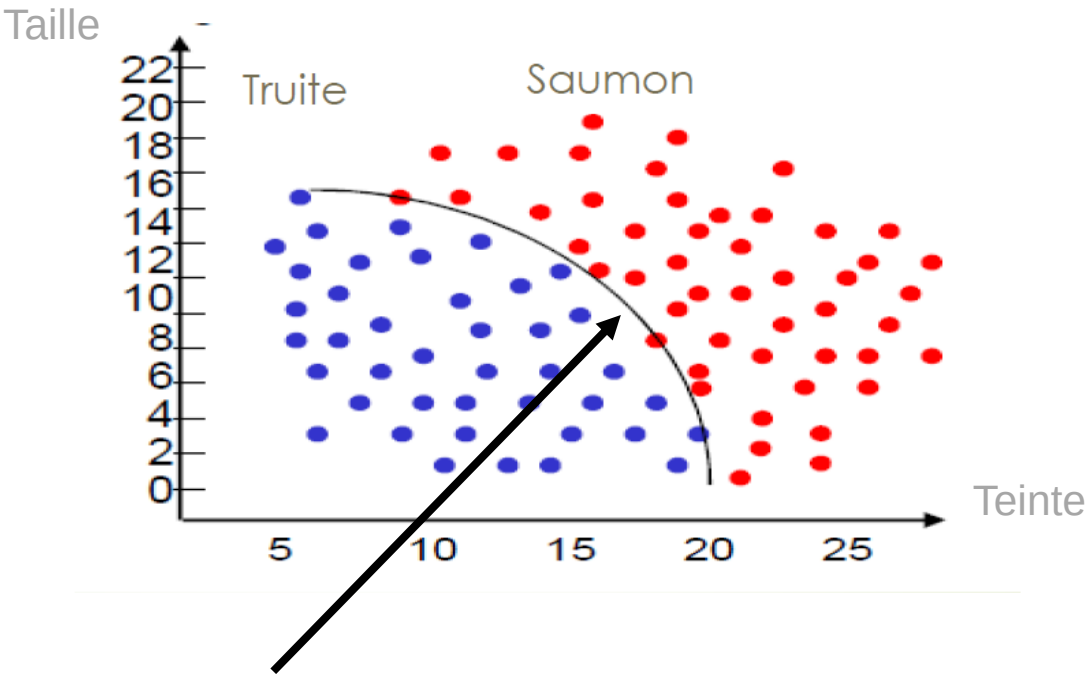


Seuille de decision

On choisi une autre caracteristique, par exemple la teinte.
Decision encore pas robuste…

Un exemple jouet : la reconnaissance

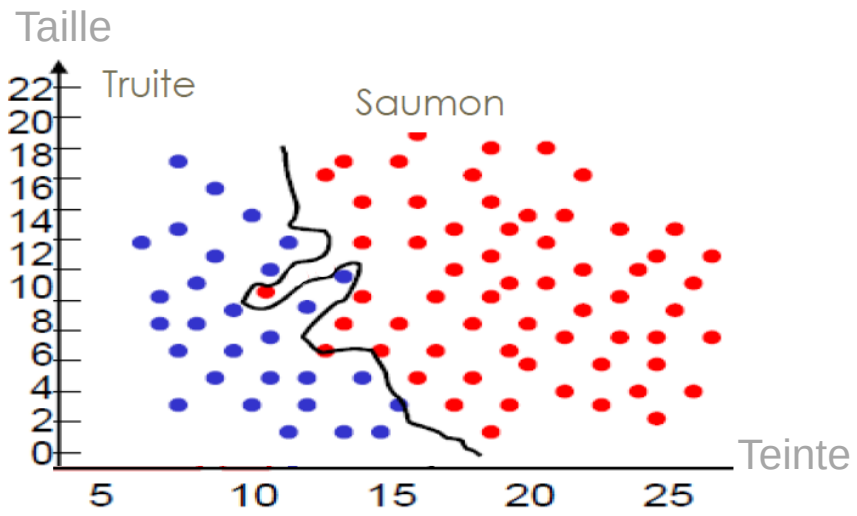
On considère le taille et la teinte ensemble, donc un vecteur en 2D

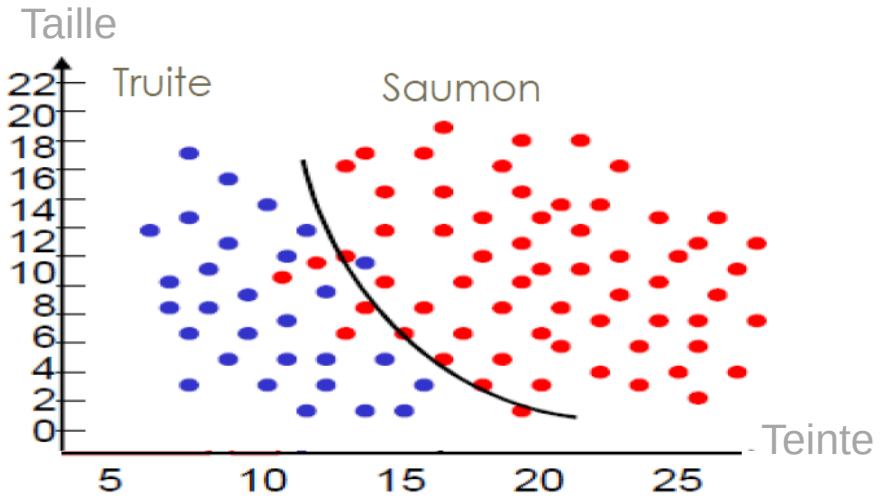


La seuille deviant une courbe! Frontière de decision

Un exemple jouet : la classification

On considère le taille et la teinte ensemble, donc un vecteur en 2D





C’est mieux une frontière plus complèxe?

Trouver une frontière moins specifique qui **generalise** sur des données inconnues

On revient sur le codage

- **Un bon codage doit avoir :**
  - **Pouvoir discriminant :**
    Forte variance inter-classes
  - **Pouvoir unifiant :**
    Faible variance intra-classes
  - **Stabilité / Invariance :**
    Insensibilité au bruit
    Invariance en translation , rotation , changement d’ échelle
  - **Faible dimension :**
    Codage de faible dimension -> temps de calcul faibles
    Malédiction de grandes dimensions

Le codage

- **Codage statistique**

  On code toute la forme sans extraire des éléments spécifiques.

  On extrait un vecteur de caractéristiques qui consiste en différentes mesures sur la forme analysée.

  Pour reconnaitre un poisson : taille et teinte

- **Codage structurel**

  On extrait des éléments spécifiques de la forme et leur relation

  On décompose la forme en primitives

  Pour reconnaitre un « L » : segment d'abord vertical puis horizontal

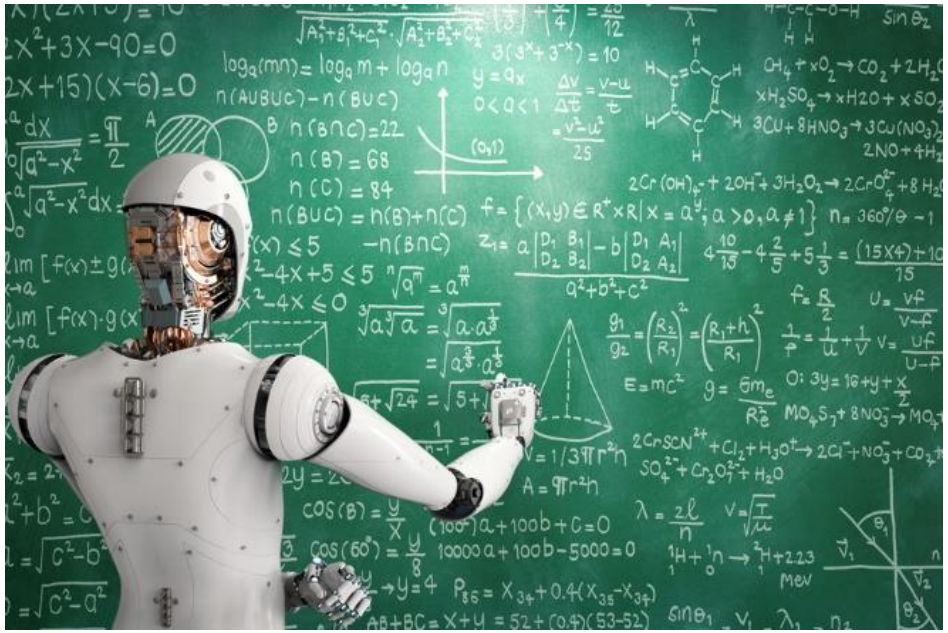
Le codage

- **Comment trouver un bon codage?**
  - **Méthodes empiriques :**
    Choisir les déscripteurs les plus pertinentes
  - **Méthodes statistiques pour réduire la dimension des données :**
    Analyse en composantes principales (ACP)
    Séléction / extraction de caractéristiques
    - filter
    - embedded
    - wrapper

…

Apprentissage automatique

« On dit qu'un programme apprend de l'expérience E en se référant à certaines classes de tâches T et à la mesure de performance P, si sa performance dans la tâche T, telle que mesurée par P, s'améliore avec l'expérience E. » - T. Mitchell



Apprentissage automatique

Supervisé et non supervisé

Apprentissage supervisé :
Apprendre une fonction qui met en correspondance une entrée et une sortie sur la base de paires d'exemples d'entrées-sorties.
Il déduit une fonction à partir d' un ensemble de données d' apprentissage étiquetées (ensemble d'exemples).

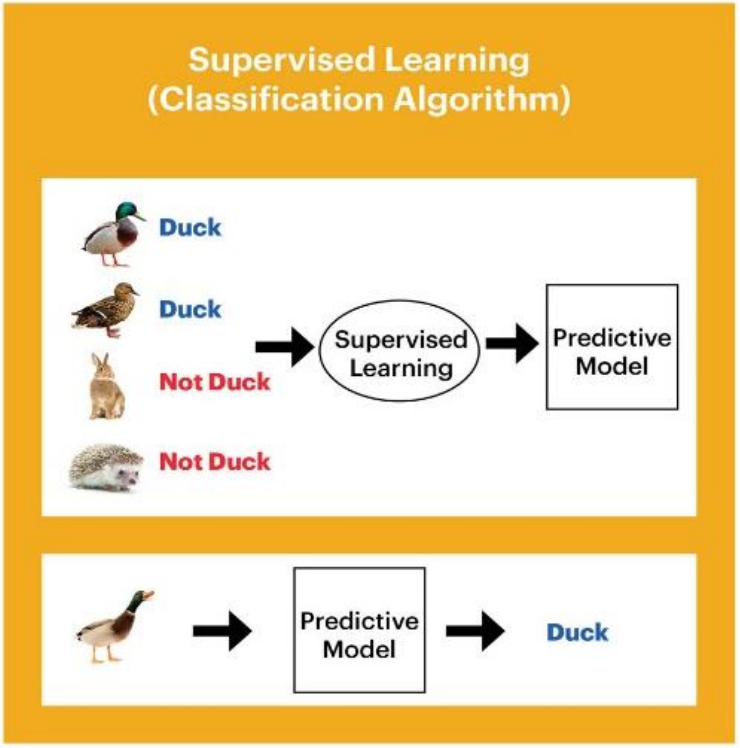
Sorties discrètes/catégories -> **classification**
Sorties continues -> **régression**

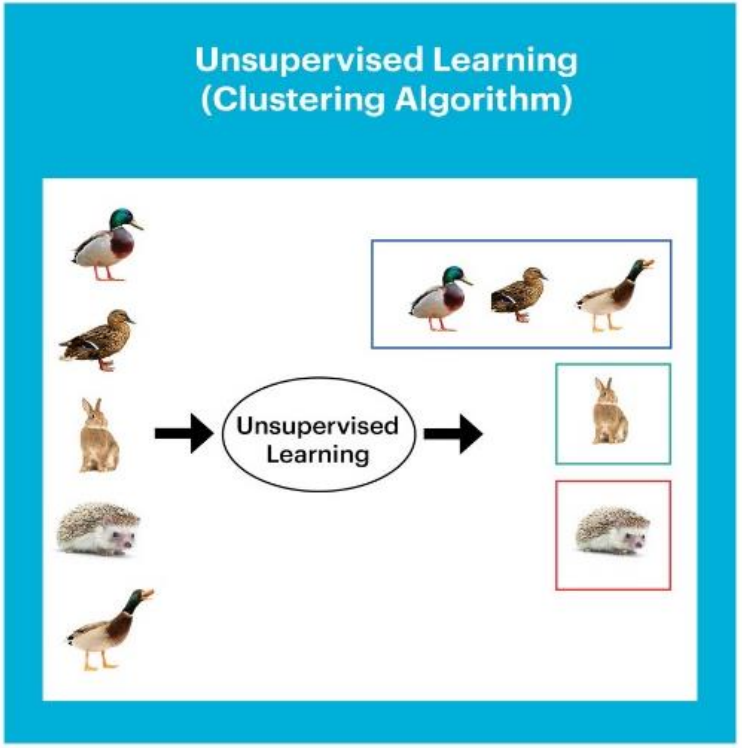
Apprentissage non supervisé :
Rechercher des représentations des données

Objectives :
- Découvrir des groupes d'exemples similaires dans les données (**clustering**)
- Déterminer la distribution des données dans l'espace des caract. (**density estimation**)
- Projeter les donnes dans un espace de dimensions plus petite (**visualization**)

Apprentissage automatique

- **Exemples**





Apprentissage automatique

- **Exemples**

On a une image binaire et on doit l'associer à une chiffre.

On connait à l'avance que il y a 10 classes possibles.

La base d'apprentissage est étiquetée avec ces 10 classes

Un client vient d'acheter un disque jazz sur Internet.

Qui sont les personnes avec les mêmes gouts ?

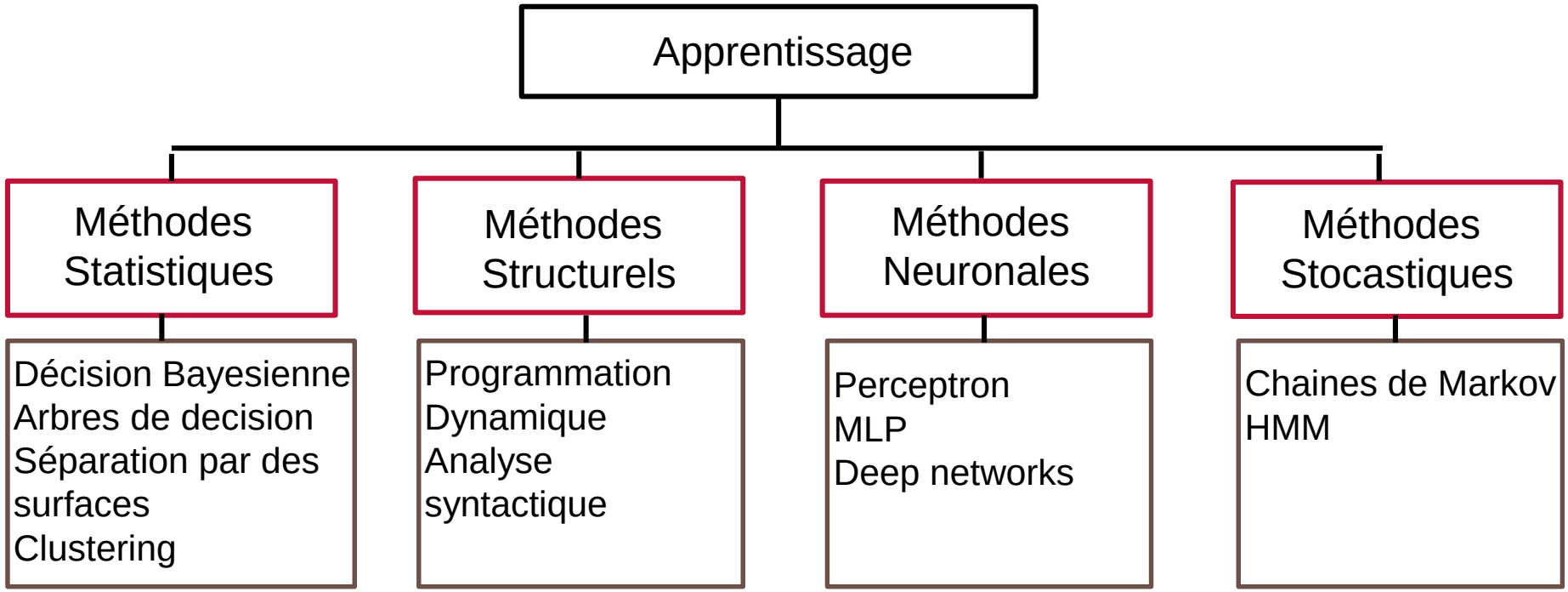
On ne connait pas à l'avance les classes possibles.

Base d'apprentissage : clients ayant fait des achats sur le même site .

On souhaite regrouper les clients qui écoutent du jazz dans une classe

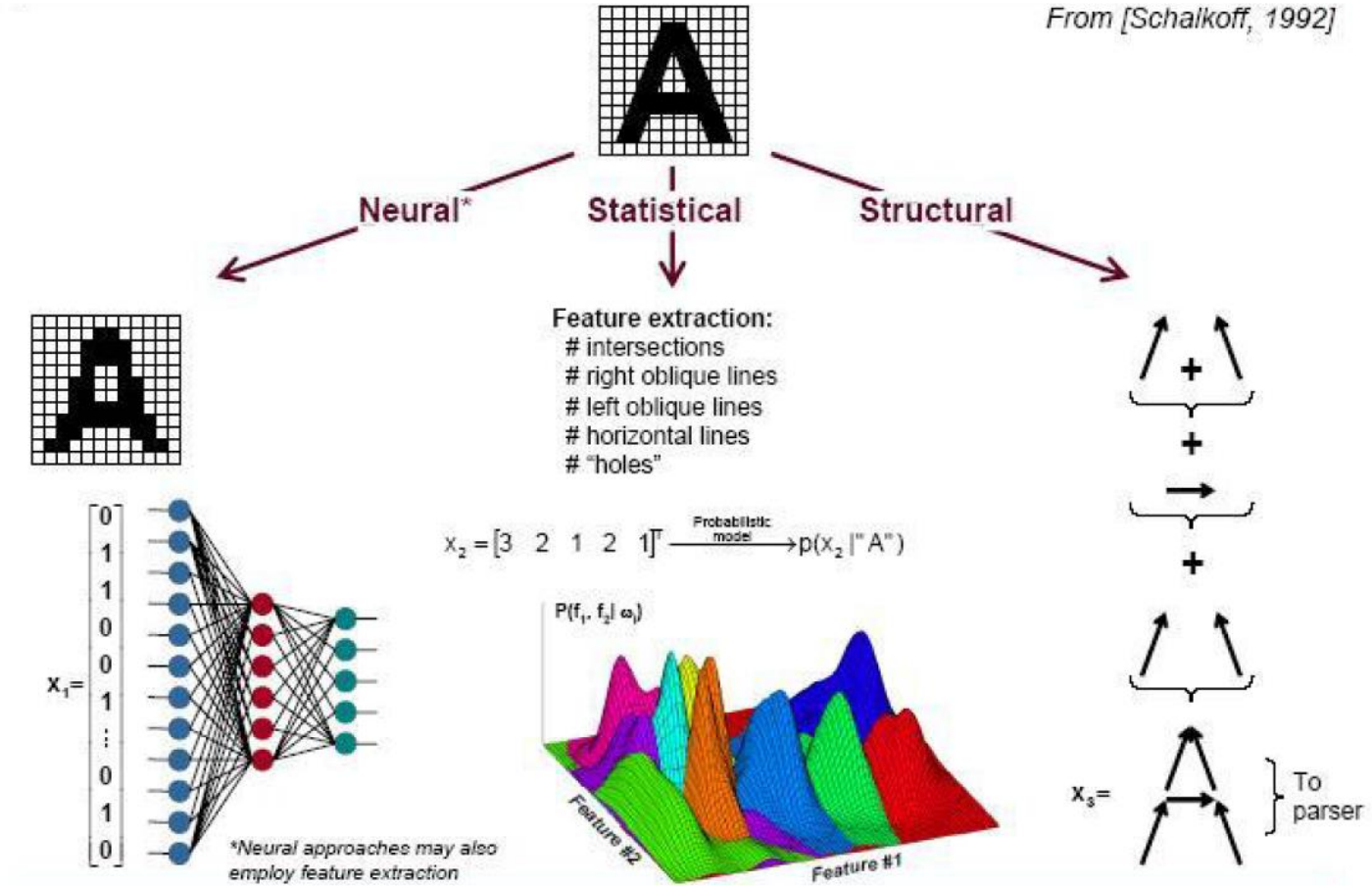
Apprentissage

- **Différentes méthodes d'apprendre**



Apprentissage

- **Exemple**



Conception de systèmes

- **Collecte de données**

  Quels sont les bonnes données? Taille et dimension raisonnables

- **Choix des caractéristiques**

  Dépendance forte du problème. Simple à extraire, invariantes, insensibles au bruit

- **Choix de modèle et Apprentissage**

  Dépendance forte du problème et des données

- **Evaluation**

  Taux d'erreur pour : différents jeux de données, différentes caractéristiques

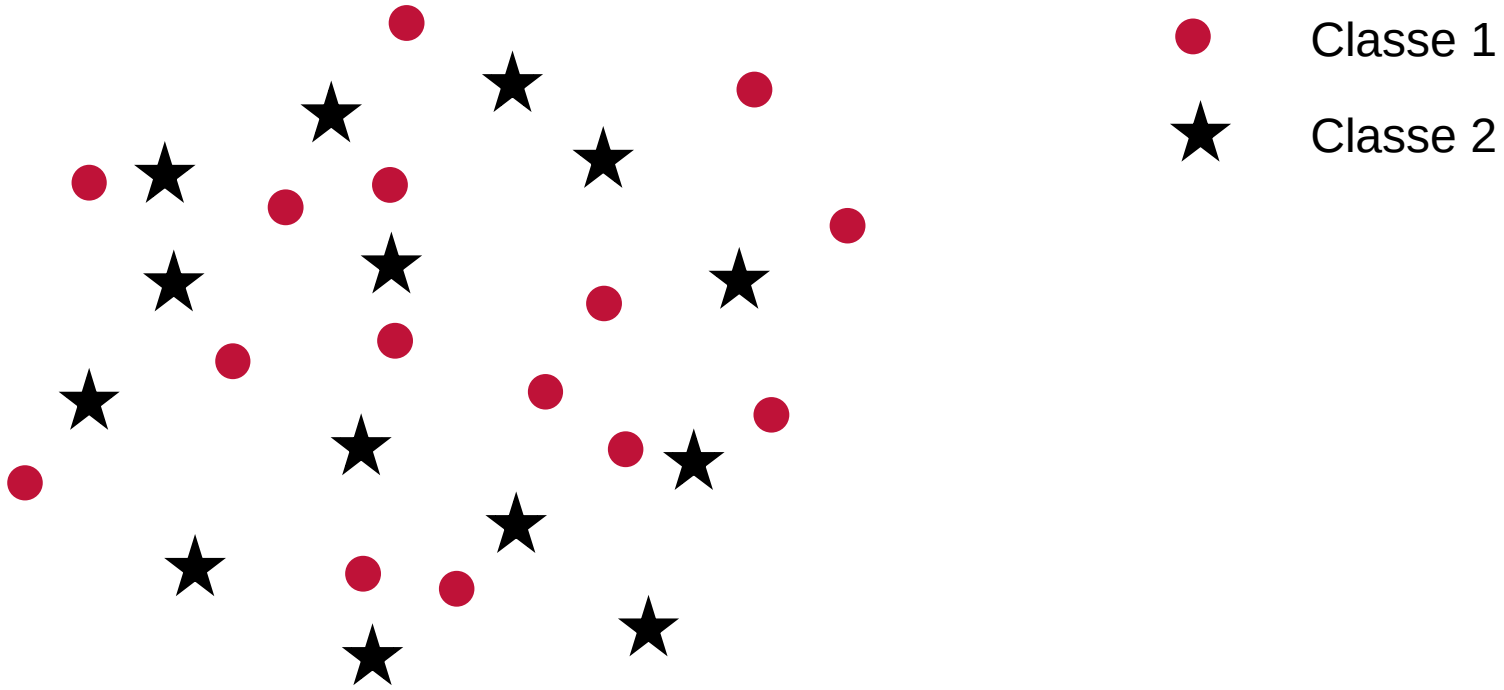
- **Complexité Computationnelle**

  Bon *trade-off* entre simplicité computationnelle et bonnes performances

  Comment un algo s'échelonne en fonction du nombre de caractéristiques, classes…

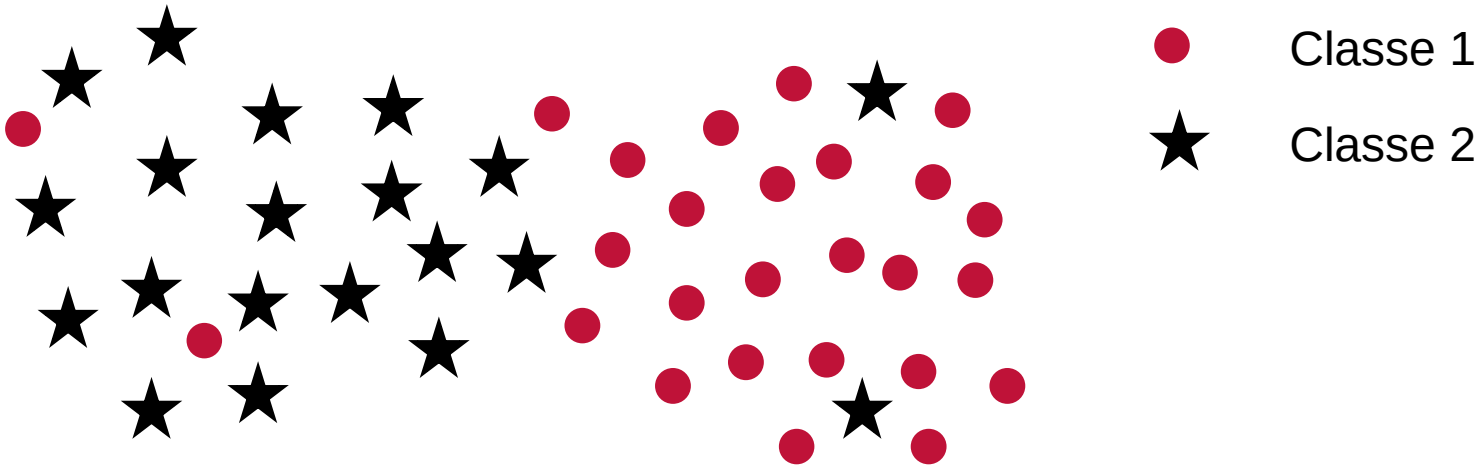
Qualité de la base de données

- **Données inadaptées :** aucune cohérence n'apparait



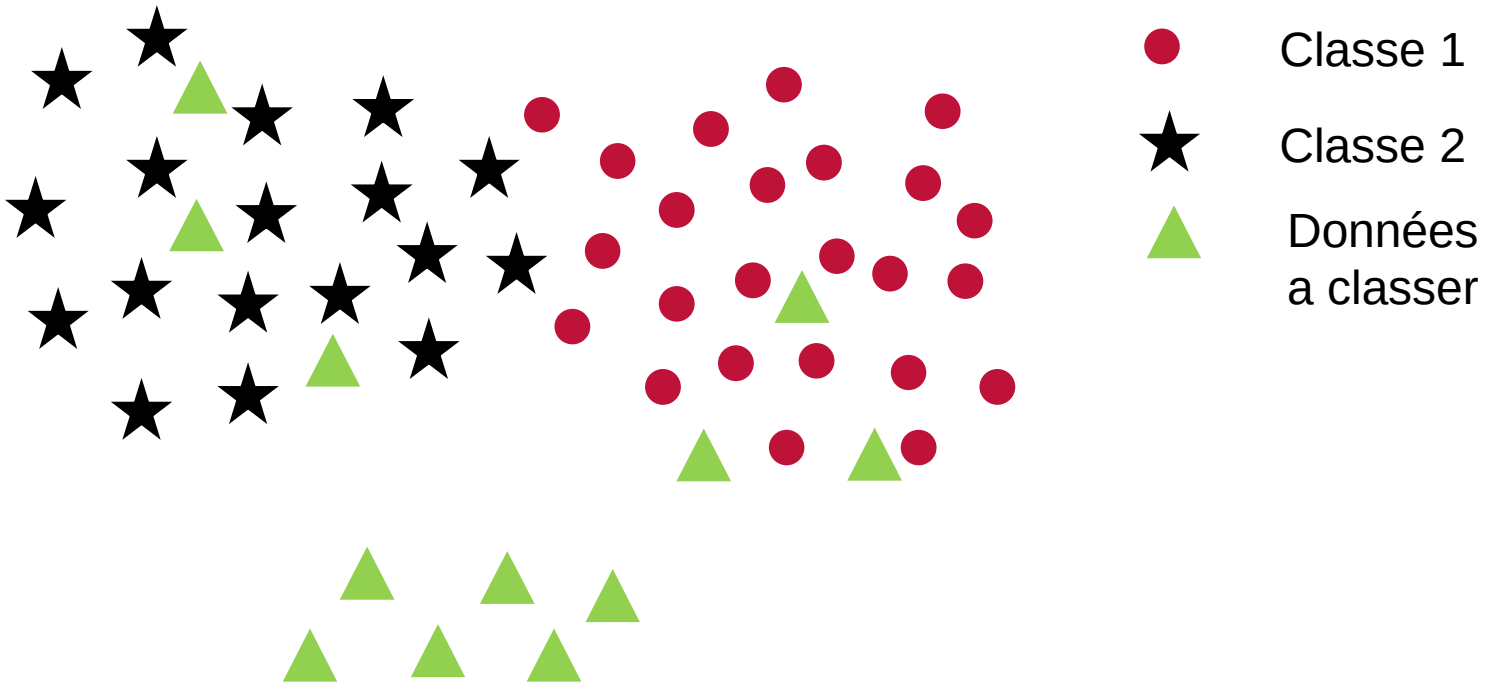
Qualité de la base de données

- **Données aberrantes**



Qualité de la base de données

- **Données manquantes**



Evaluation

- Processus qui consiste à donner objectivement la mesure dans laquelle les modèles d'apprentissage accomplissent les tâches spécifiques pour lesquelles ils ont été conçus

- Diffèrent approches pour apprentissage supervisé et non supervisé

Evaluation : apprentissage supervise - métriques

- **Classification**

Il y a plusieurs métriques : Matrice de confusion, AU-ROC…

Matrice de confusion

Etiquette \ Décision	1	2
1	Nb d'exemples réellement 1 étiquetés 1	Nb d'exemples réellement 1 étiquetés 2
2	Nb d'exemples réellement 2 étiquetés 1	Nb d'exemples réellement 2 étiquetés 2

Definie aussi dans les cas multi-classes

Evaluation : apprentissage supervise - métriques

- On peut calculer le **taux de bonne classification**

Sans rejet

Taux de bonne classification (Tb_s) $$Tb_s = \frac{\text{Nb d'exemples bien classés}}{\text{Nb d'exemples}}$$

Taux d'erreur (Te_s) $Te_s = 1 - Tb_s$

Avec rejet (si le système peut ne pas prendre une décision)

Taux de bonne classification (Tb_a) $$Tb_a = \frac{\text{Nb d'exemples bien classés}}{\text{Nb d'exemples}}$$

Taux de rejet (Tr) $$Tr = \frac{\text{Nb d'exemples non classés}}{\text{Nb d'exemples}}$$

Taux d'erreur (Te_a) $Te_a = 1 - Tr - Tb_a$

Evaluation : apprentissage supervise - métriques

- **Taux de bonne classification :**

  mesure « faible » ne tient pas compte la distribution des classes

Exemple

En diagnostic médical, très peu de personnes sont malades (5%).

On a donc des taux très bons en disant que la personne est saine.

Or, ce que l'on souhaite, c'est ne pas rater ces 5% et donc, associer un mauvais taux au classificateur qui dirait toujours 'personne saine'.

Exemple sur 100 personnes

	malade	sain
malade	0	5
sain	0	95

$Tb_s = 0.95$

Evaluation : apprentissage supervise - métriques

- **Taux de bonne classification :**

  ne tient pas compte la distribution des classes

Solution : matrice de confusion normalisée

Décision / Etiquette	1	2
1	Nb d'exemples réellement 1 étiquetés 1 **/ N1**	Nb d'exemples réellement 1 étiquetés 2 **/ N1**
2	Nb d'exemples réellement 2 étiquetés 1 **/ N2**	Nb d'exemples réellement 2 étiquetés 2 **/N2**

N1 = nb exemples de la classe 1 ; N2 = nb exemples de la classe 2

Evaluation : apprentissage supervise - métriques

- **On définie un nouveau taux de bonne classification :**

$$Tb = \frac{1}{Nc} \sum_{k=1}^{Nc} \frac{Nb\ d'exemples\ réellement\ k\ étiquetés\ k}{Nk}$$

	malade	sain
malade	0	1
sain	0	1

Tb = 0.5 Te = 0.5 (Te = 1 - Tb – Tr)

Evaluation : apprentissage supervisé - métriques

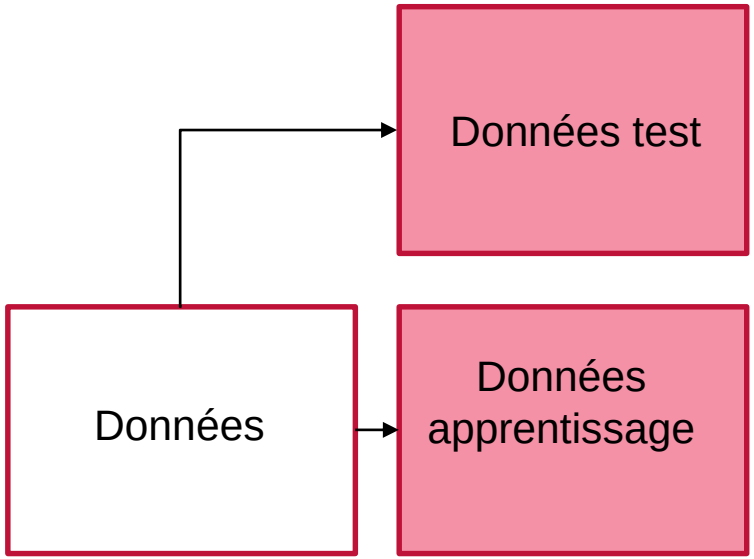
- **Régression**

  il y a plusieurs métriques :

  - Erreur absolue moyenne
  - Erreur quadratique moyenne
  - R ^2

Evaluation : apprentissage supervise - méthodes

- **Comment évaluer un modèle ?**
  - Scenario 1 : apprentissage – test

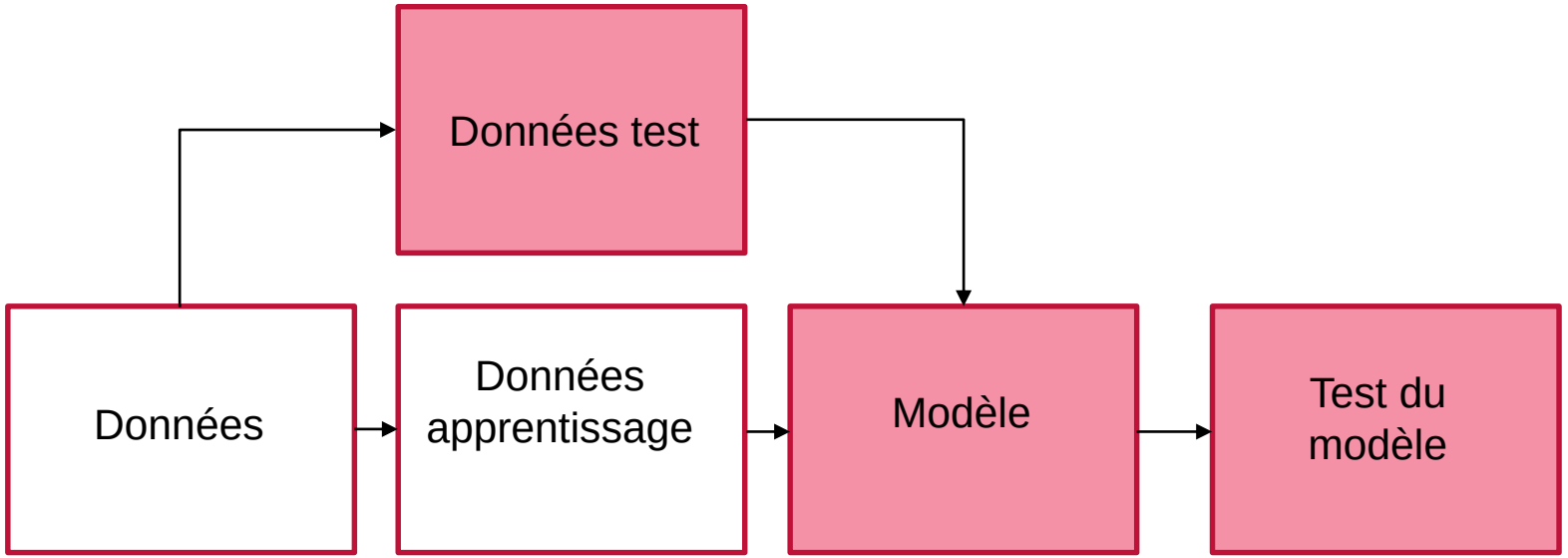


Evaluation : apprentissage supervise - méthodes

- **Comment évaluer un modèle ?**
  - Scenario 1 : apprentissage – test

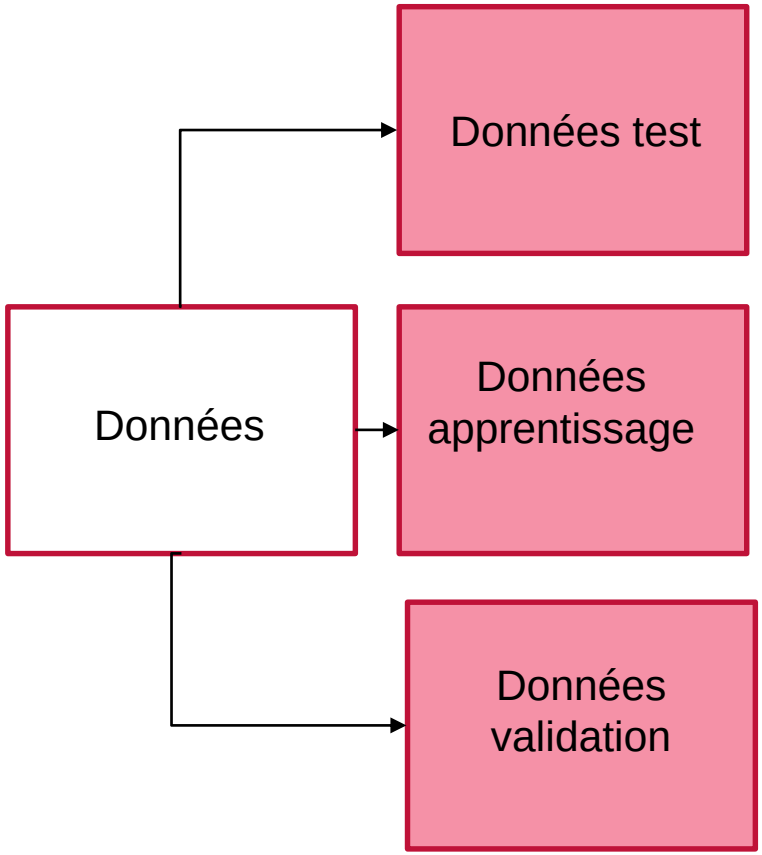
Evaluation : apprentissage supervise - méthodes

- **Comment évaluer un modèle ?**
  - Scenario 1 : apprentissage – test

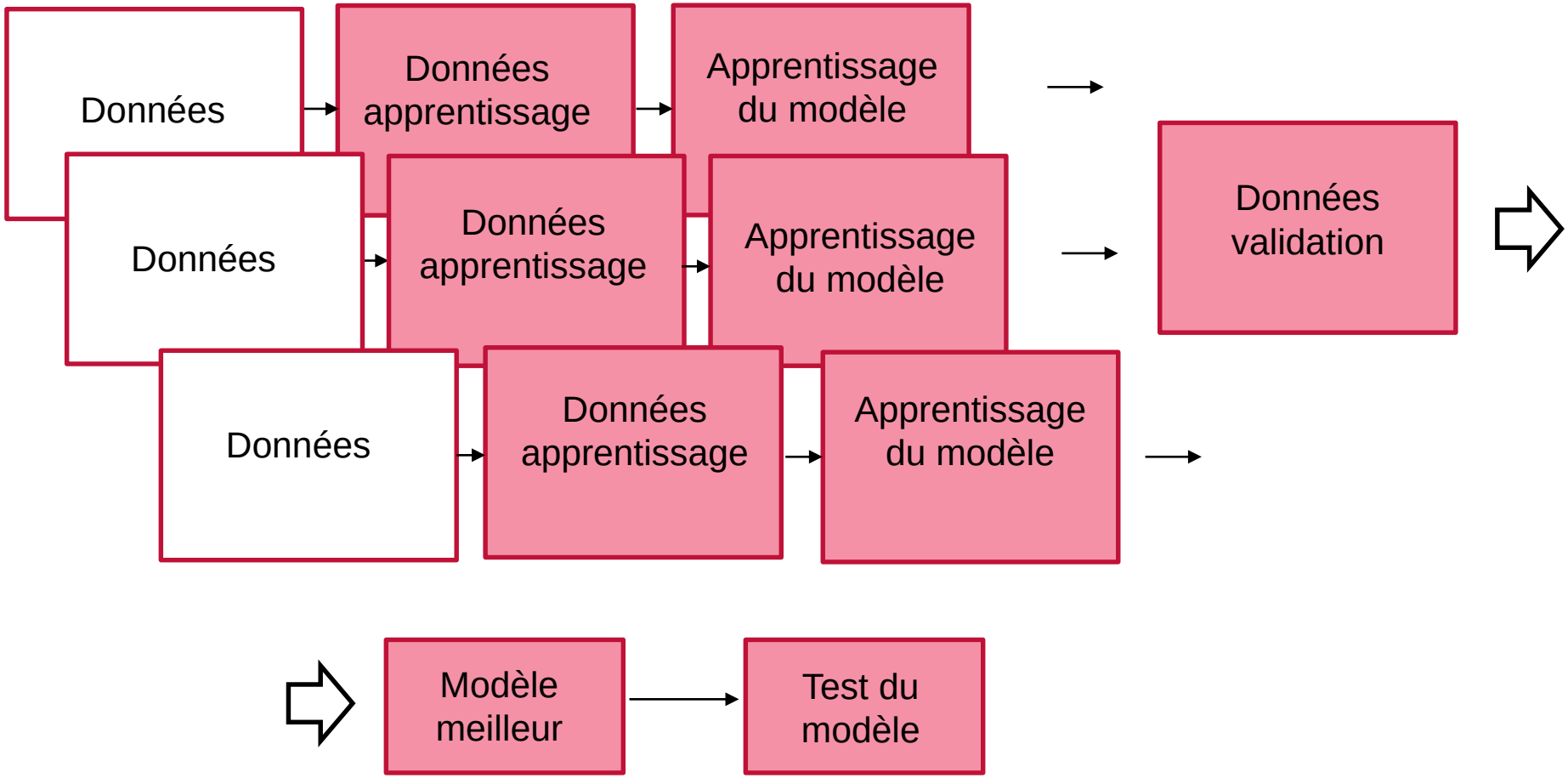


Evaluation : apprentissage supervise - méthodes

- **Comment évaluer un modèle ?**
  - Scenario 2 : apprentissage – validation – test

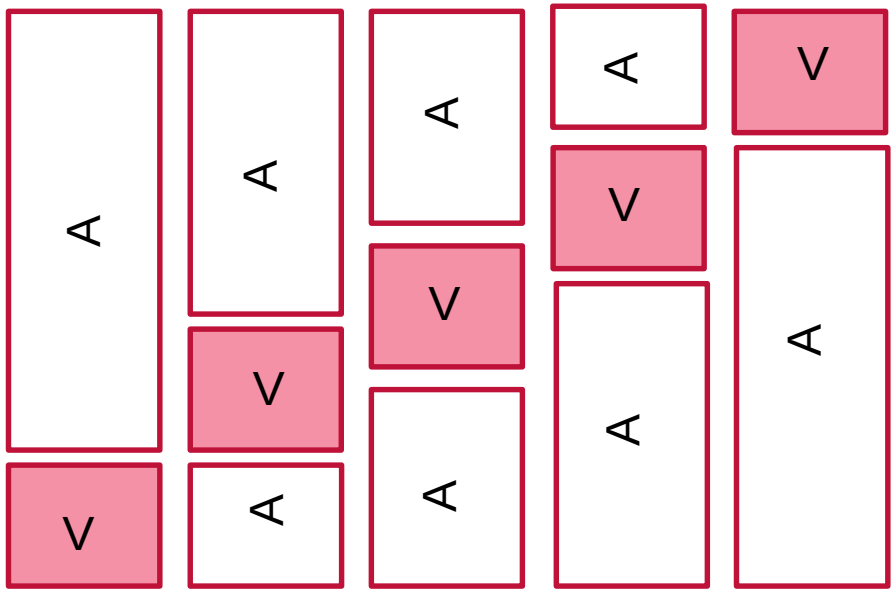


Evaluation : apprentissage supervise - méthodes



Evaluation : apprentissage supervise - méthodes

- **Comment évaluer un modèle ?**
  - Scenario 2 : validation croisée



validation croisée à **5-plis**

Evaluation : apprentissage non supervisé

- **Clustering**

Evaluation problématique…

Deux classes de évaluation :

- *Validation interne :* utilise des informations intrinsèques aux seules données

  Reflète : compacité, séparation, connectivité

$$\text{Indices} = (\alpha \times \text{Separation}) / (\beta \times \text{Compactness}) \qquad \alpha, \beta \text{ poids}$$

- *Validation externe :* utilise les connaissances antérieures sur les données

  Connaissances antérieures : gold standard

  On cherche dans quelle mesure le regroupement correspond au gold standard

  Indices : pureté, entropie, Adjusted Random Index, …