

# Apprentissage, Classification & Pr evision

David Nerini

COM

premier semestre

- La description analytique de tous les processus existants et qui interagissent dans le milieu naturel est une tâche impossible. La construction de modèles mathématiques cohérents, à but explicatif, doit être effectuée sous des hypothèses souvent loin de la réalité. Cependant, leur élaboration est indispensable puisqu'ils doivent permettre de mieux *comprendre* et d'*expliquer* le fonctionnement des écosystèmes.

- La description analytique de tous les processus existants et qui interagissent dans le milieu naturel est une tâche impossible. La construction de modèles mathématiques cohérents, à but explicatif, doit être effectuée sous des hypothèses souvent loin de la réalité. Cependant, leur élaboration est indispensable puisqu'ils doivent permettre de mieux *comprendre* et d'*expliquer* le fonctionnement des écosystèmes.
- Une des grandes tâches en statistiques consiste à *décrire* le fonctionnement d'un système à partir de l'observation de variables échantillonnées sur un grand nombre d'individus. L'objectif n'est pas d'étudier les processus qui interagissent dans le système mais de *prévoir* ou d'*anticiper* des modifications de *variables cibles* liées de manière complexe avec des *variables exogènes*.

- La description analytique de tous les processus existants et qui interagissent dans le milieu naturel est une tâche impossible. La construction de modèles mathématiques cohérents, à but explicatif, doit être effectuée sous des hypothèses souvent loin de la réalité. Cependant, leur élaboration est indispensable puisqu'ils doivent permettre de mieux *comprendre* et d'*expliquer* le fonctionnement des écosystèmes.
- Une des grandes tâches en statistiques consiste à *décrire* le fonctionnement d'un système à partir de l'observation de variables échantillonnées sur un grand nombre d'individus. L'objectif n'est pas d'étudier les processus qui interagissent dans le système mais de *prévoir* ou d'*anticiper* des modifications de *variables cibles* liées de manière complexe avec des *variables exogènes*.



*expliquer*  $\neq$  *prédire*

# Position du problème

- On dispose d'un échantillon  $E = \{(y_i, \mathbf{x}_i), i = 1, \dots, n\}$  d'une *variable à prédire*  $Y$  et de *variables de prévisions* ou *explicatives*  $\mathbf{X} = (X_1, \dots, X_p)'$ . Cet échantillon doit être représentatif de la population dans laquelle il a été prélevé. Cela veut dire que si sa taille  $n$  tend vers l'infini, les caractéristiques des variables échantillonnées se rapprochent de celles de leur distribution théorique.

# Position du problème

- On dispose d'un échantillon  $E = \{(y_i, \mathbf{x}_i), i = 1, \dots, n\}$  d'une *variable à prédire*  $Y$  et de *variables de prévisions* ou *explicatives*  $\mathbf{X} = (X_1, \dots, X_p)'$ . Cet échantillon doit être représentatif de la population dans laquelle il a été prélevé. Cela veut dire que si sa taille  $n$  tend vers l'infini, les caractéristiques des variables échantillonnées se rapprochent de celles de leur distribution théorique.
- On cherche à construire un *modèle statistique* de la forme

$$y = f_{\theta}(\mathbf{x}) + \varepsilon$$

où  $f_{\theta}$  est une fonction qui décrit les relations liant la variable  $y$  aux  $x_j$ . Sa forme peut appartenir à une famille paramétrique ou ne pas être connue. La détermination de  $f_{\theta}$ , directement ou par l'intermédiaire des paramètres  $\theta$ , est effectuée en minimisant la norme des erreurs  $\varepsilon$ .

## Definition

On appelle Apprentissage Machine (*Machine Learning*) l'ensemble des méthodes statistiques relatives à l'étude des propriétés de  $E$ , de l'estimation de  $f_\theta$  et de son utilisation pour effectuer des prévisions.

- Les techniques mathématiques utilisées pour la construction des modèles et la prévision utilisent énormément de données.

## Definition

On appelle Apprentissage Machine (*Machine Learning*) l'ensemble des méthodes statistiques relatives à l'étude des propriétés de  $E$ , de l'estimation de  $f_\theta$  et de son utilisation pour effectuer des prévisions.

- Les techniques mathématiques utilisées pour la construction des modèles et la prévision utilisent énormément de données.
- L'étude de leurs propriétés et leur utilisation dans un cadre opérationnel sont souvent basées sur des algorithmes informatiques particuliers (*Machine Learning*)

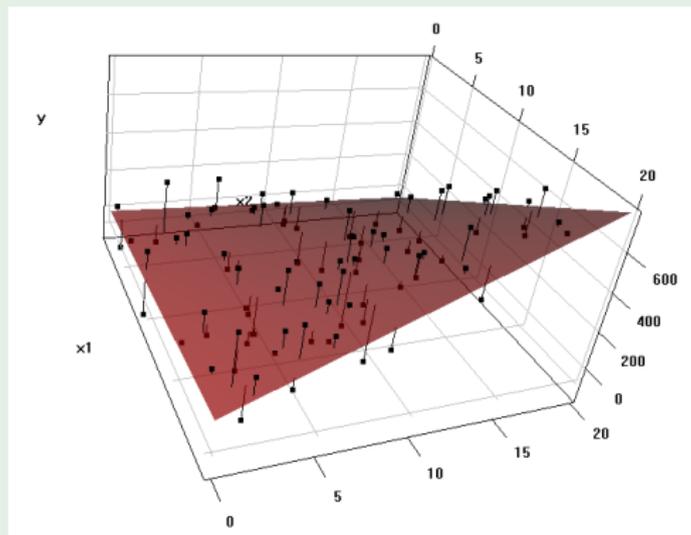
## Definition

On appelle Apprentissage Machine (*Machine Learning*) l'ensemble des méthodes statistiques relatives à l'étude des propriétés de  $E$ , de l'estimation de  $f_\theta$  et de son utilisation pour effectuer des prévisions.

- Les techniques mathématiques utilisées pour la construction des modèles et la prévision utilisent énormément de données.
- L'étude de leurs propriétés et leur utilisation dans un cadre opérationnel sont souvent basées sur des algorithmes informatiques particuliers (*Machine Learning*)
- Le choix des méthodes repose essentiellement sur le type de données avec lesquelles l'écologue travaille.

## Exemple

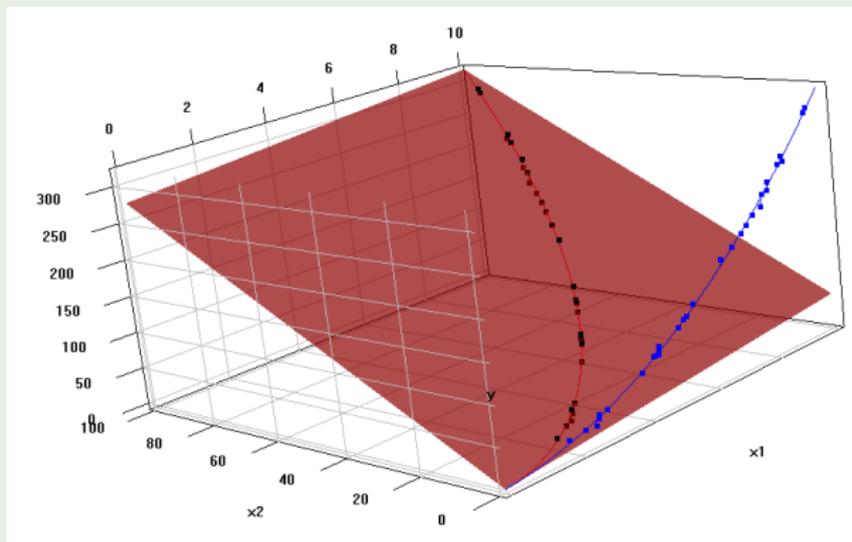
Régression linéaire multiple  $y = ax_1 + bx_2 + cx_1x_2 + \varepsilon$



- On cherche à *estimer* un modèle à partir de données qui s'ajuste au mieux à un *modèle théorique*, supposé existant.

## Exemple

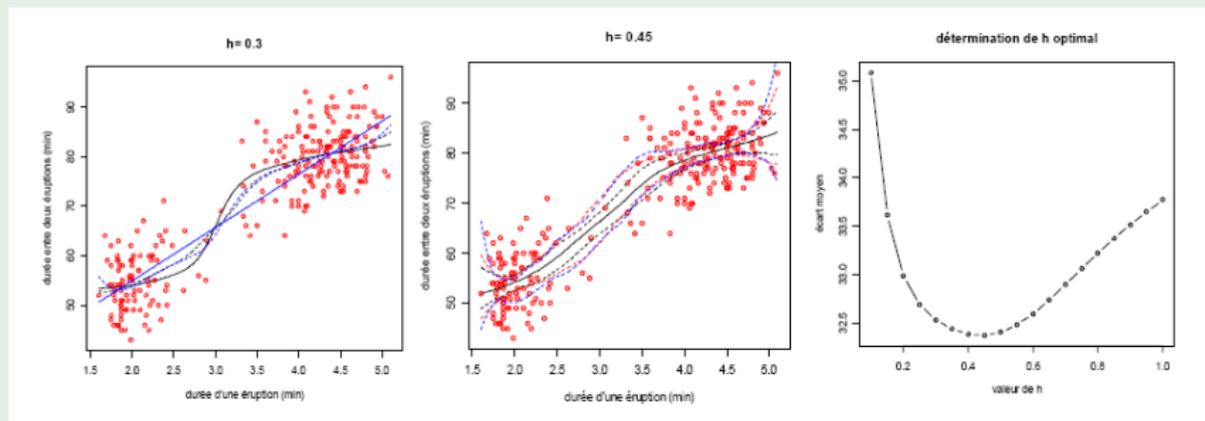
Régression polynômiale  $y = ax + bx^2 + \varepsilon$



- La forme du modèle dépend des caractéristiques des variables étudiées même si la problématique est identique au cas précédent.

## Exemple

Régression non-paramétrique  $y = f(x) + \varepsilon$



- L'estimation de la fonction est directement réalisée à l'aide des données sans passer par une *estimation de paramètres*

- Dans les trois cas précédents, l'estimation de  $f_\theta$  est effectuée par *moindres carrés* à l'aide de l'échantillon d'apprentissage  $E$ . On cherche en fait à minimiser la norme des erreurs  $\varepsilon$

$$\|\varepsilon\|^2 = \text{RSS}(f_\theta) = \sum_{i=1}^n K(x_i) (y_i - f_\theta(x_i))^2$$

- Dans les trois cas précédents, l'estimation de  $f_\theta$  est effectuée par *moindres carrés* à l'aide de l'échantillon d'apprentissage  $E$ . On cherche en fait à minimiser la norme des erreurs  $\varepsilon$

$$\|\varepsilon\|^2 = \text{RSS}(f_\theta) = \sum_{i=1}^n K(x_i) (y_i - f_\theta(x_i))^2$$

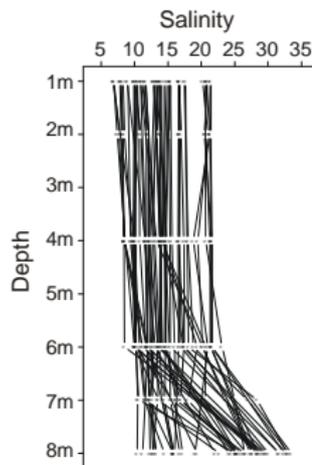
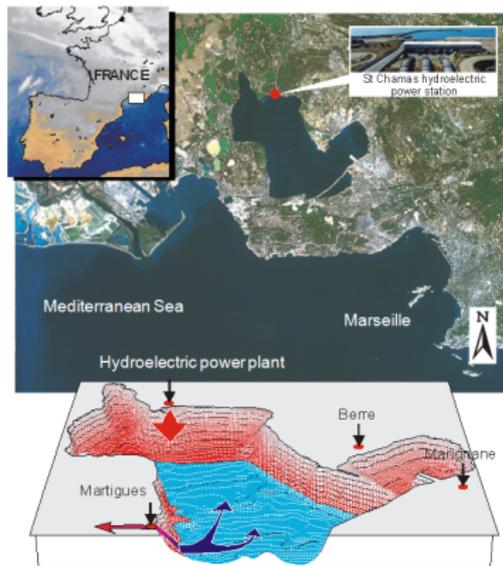
- La forme de la fonction poids  $K(x_i)$  dépend de la méthode employée

- Dans les trois cas précédents, l'estimation de  $f_\theta$  est effectuée par *moindres carrés* à l'aide de l'échantillon d'apprentissage  $E$ . On cherche en fait à minimiser la norme des erreurs  $\varepsilon$

$$\|\varepsilon\|^2 = \text{RSS}(f_\theta) = \sum_{i=1}^n K(x_i) (y_i - f_\theta(x_i))^2$$

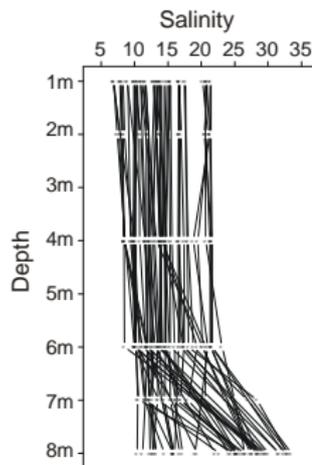
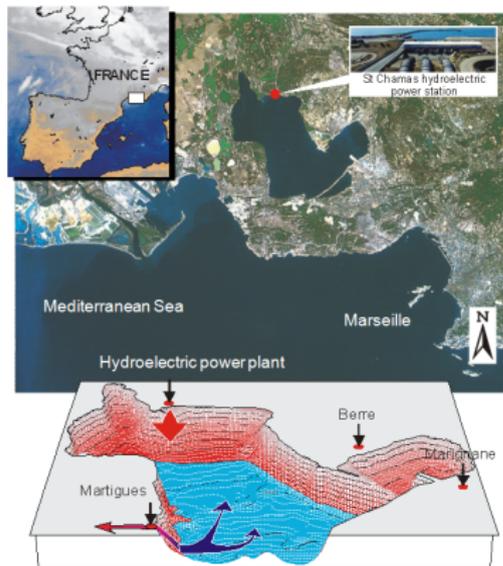
- La forme de la fonction poids  $K(x_i)$  dépend de la méthode employée
- La méthode des MC est à la base de la construction de la plupart des modèles

# Exemples : Y catégorielle et X quelconques



- On construit une *variable qualitative*  $Y \in \{e_1, \dots, e_m\}$  qui décrit les états hydrologiques de l'étang

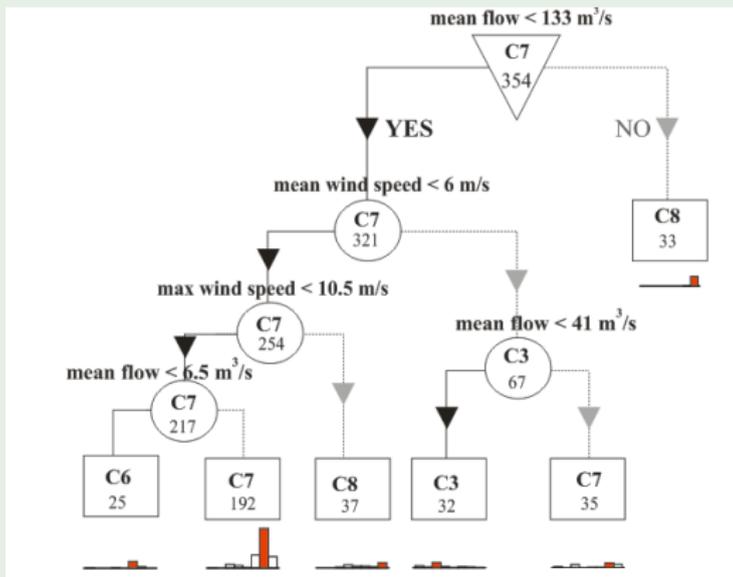
# Exemples : Y catégorielle et X quelconques



- On construit une *variable qualitative*  $Y \in \{e_1, \dots, e_m\}$  qui décrit les états hydrologiques de l'étang
- Prédiction des conditions hydrologiques qualitatives en fonction d'un forçage externe (vent et débits)

# Exemples : Y catégorielle et X quelconques

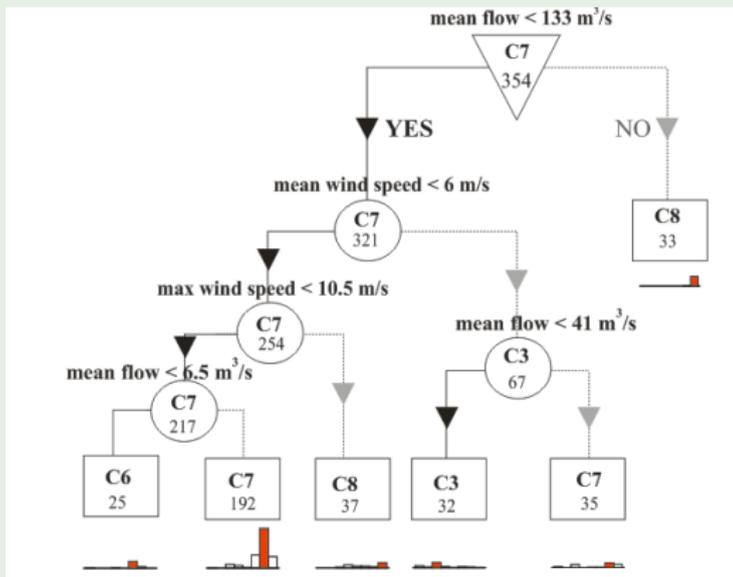
## Exemple (Arbre de classification)



- $$y = f_L(\mathbf{x}) = \sum_{j=1}^q c_j I(\mathbf{X} \in r_j)$$

# Exemples : Y catégorielle et X quelconques

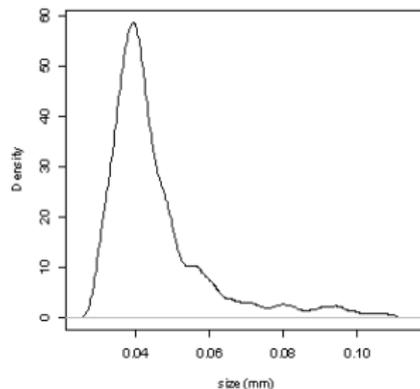
## Exemple (Arbre de classification)



- $$y = f_L(\mathbf{x}) = \sum_{j=1}^q c_j I(\mathbf{X} \in r_j)$$

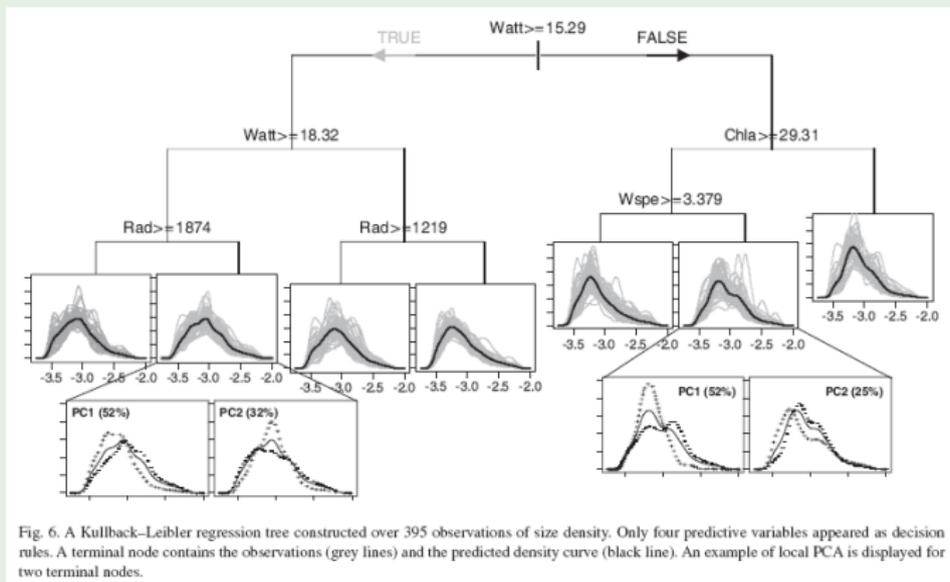
- La classe  $c_j$  est celle qui a la probabilité maximale dans un noeud

# Exemples : Y fonctionnelle et X quelconques



- Prédiction du spectre de taille du zooplancton en fonction de variables environnementales

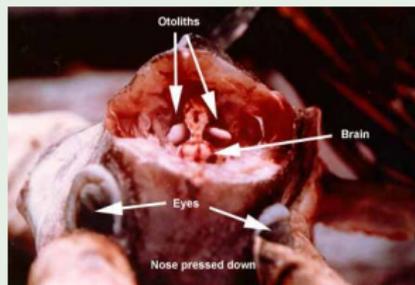
## Exemple (Arbre de régression)



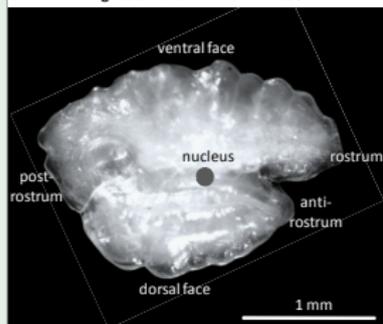
- le modèle est de la forme

$$y = f_L(\mathbf{x}) = E(Y/\mathbf{X} = \mathbf{x}) = \sum_{j=1}^q f_j I(\mathbf{X} \in r_j)$$

## Example (location and shape)

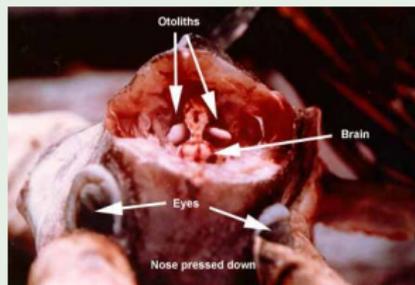


right otolith of *Mullus Barbatus*

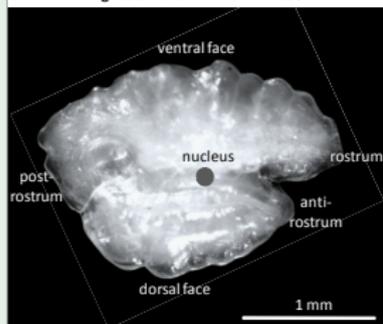


- a small piece of calcium carbonate located in the head of *Teleostean* fishes

## Example (location and shape)

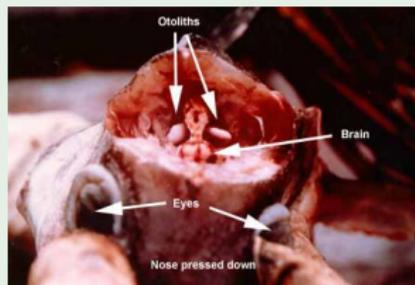


right otolith of *Mullus Barbatus*

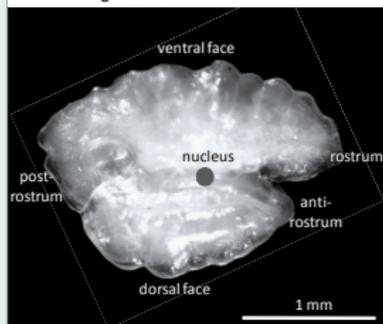


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance

## Example (location and shape)

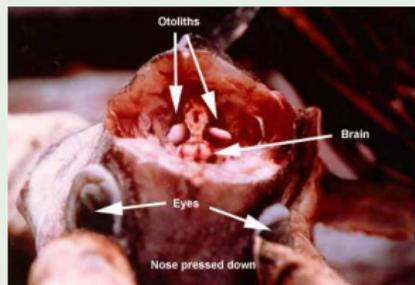


right otolith of *Mullus Barbatus*

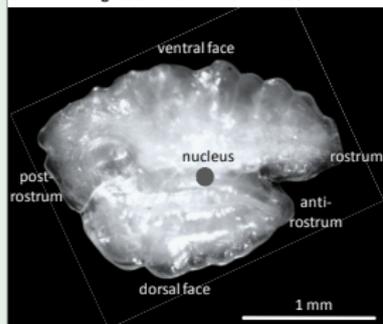


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions

## Example (location and shape)

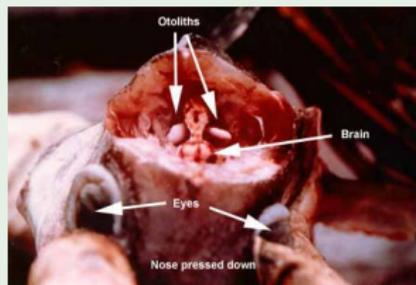


right otolith of *Mullus Barbatus*

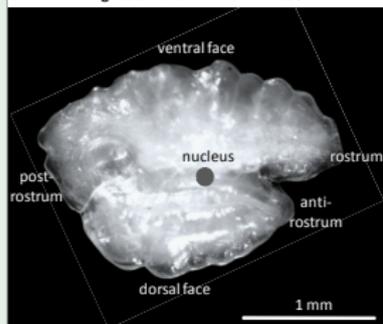


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions
- 3 pairs

## Example (location and shape)

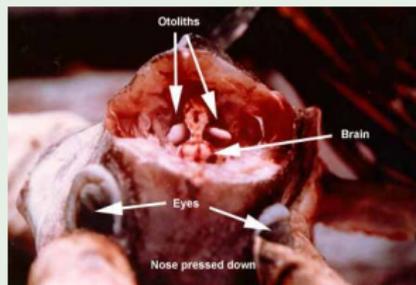


right otolith of *Mullus Barbatus*

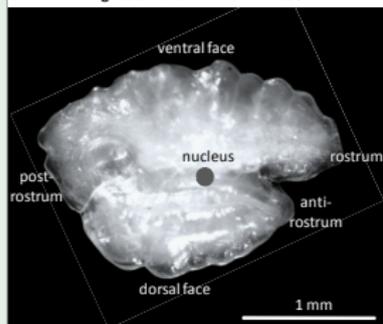


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions
- 3 pairs
- *sagitta* pair

## Example (location and shape)

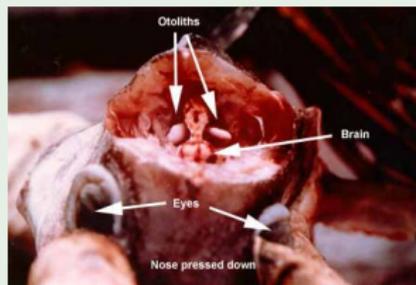


right otolith of *Mullus Barbatius*

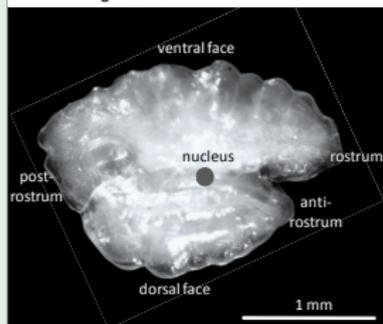


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions
- 3 pairs
- *sagitta* pair
  - the biggest one

## Example (location and shape)

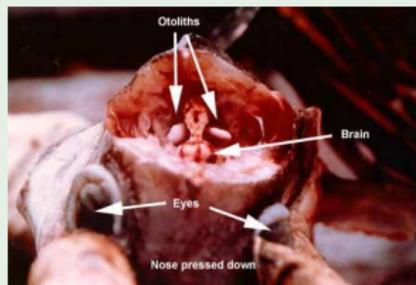


right otolith of *Mullus Barbatulus*

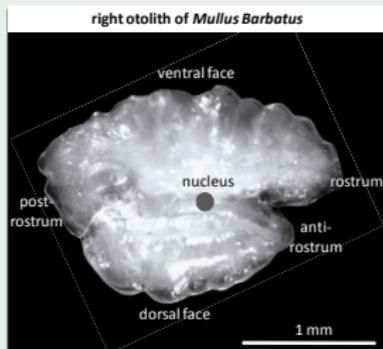


- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions
- 3 pairs
- *sagitta* pair
  - the biggest one
  - grows from the date of hatch to the time of death

## Example (location and shape)



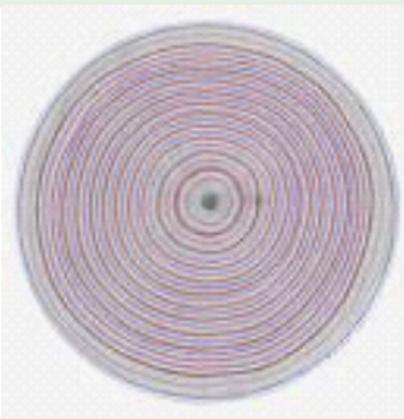
- a small piece of calcium carbonate located in the head of *Teleostean* fishes
  - acceleration perception and sense of balance
  - tridimensional perception and aid in hearing functions
- 3 pairs
- *sagitta* pair
  - the biggest one
  - grows from the date of hatch to the time of death
  - growth starts from *nucleus*



# What bring otolith studies ?

- Fish stocks, populations and species can be studied through the size, composition and shape of their otoliths
- Chemical composition and shape provide informations about fish environment (temperature, migration pathways) and dietary pattern
- Lateral section shows daily growth rings that record age and growth patterns

## Example (Otolith and tree rings)



# Shape analysis

- We aim to compare shape of a collection of *Mullus barbatus* otoliths

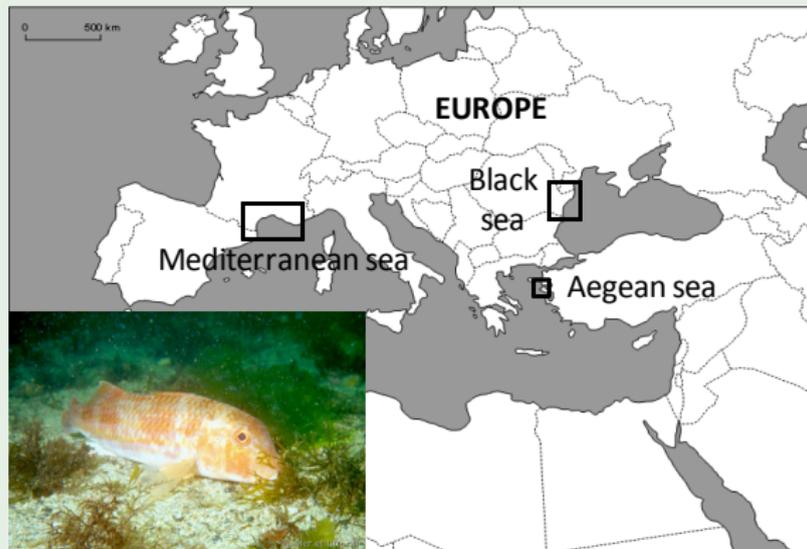
# Shape analysis

- We aim to compare shape of a collection of *Mullus barbatus* otoliths
  - **objective** : discriminate individuals belonging to same species but sampled at different places

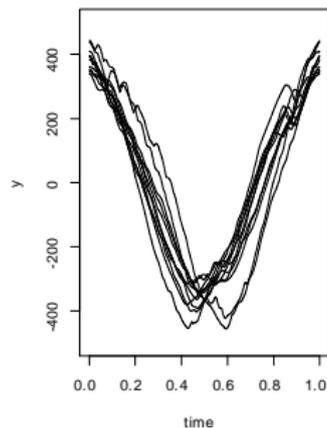
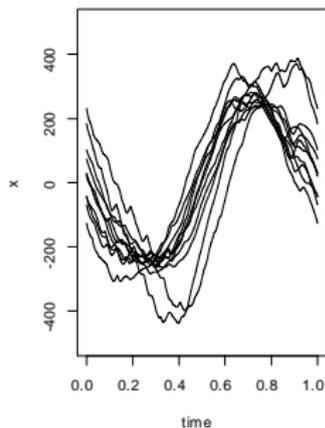
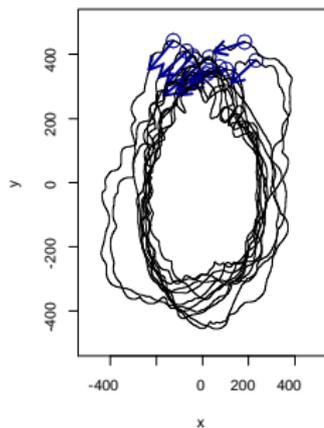
# Shape analysis

- We aim to compare shape of a collection of *Mullus barbatus* otoliths
  - **objective** : discriminate individuals belonging to same species but sampled at different places

## Example (Sampling location & *Mullus barbatus*)

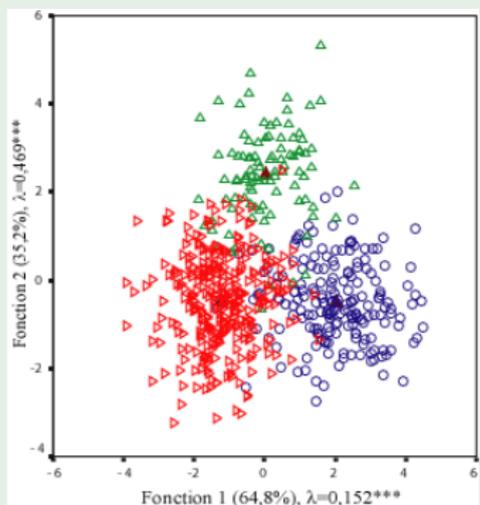


# Exemples : Y catégorielle et X fonctionnelle bivariée



- Problème de recalage (analyse Procuste), objets bidimensionnels
- Peut-on utiliser leurs formes pour connaître leur provenance ?

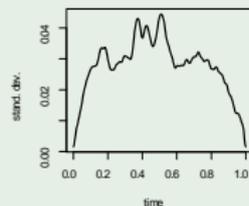
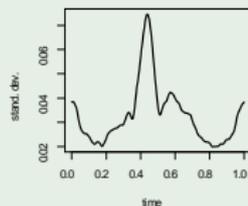
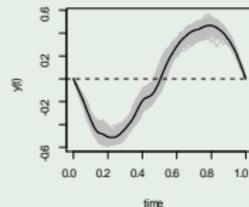
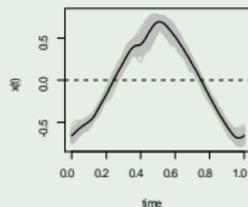
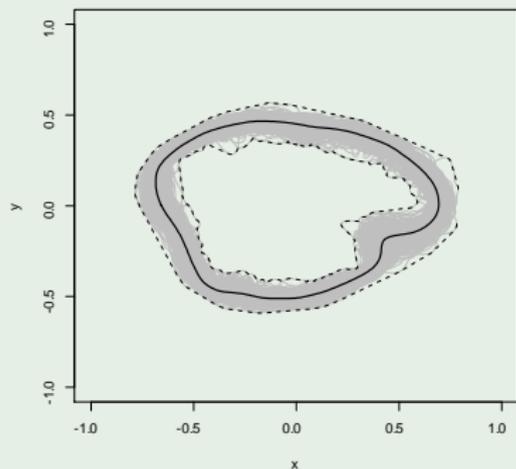
## Exemple (Analyse Discriminante)



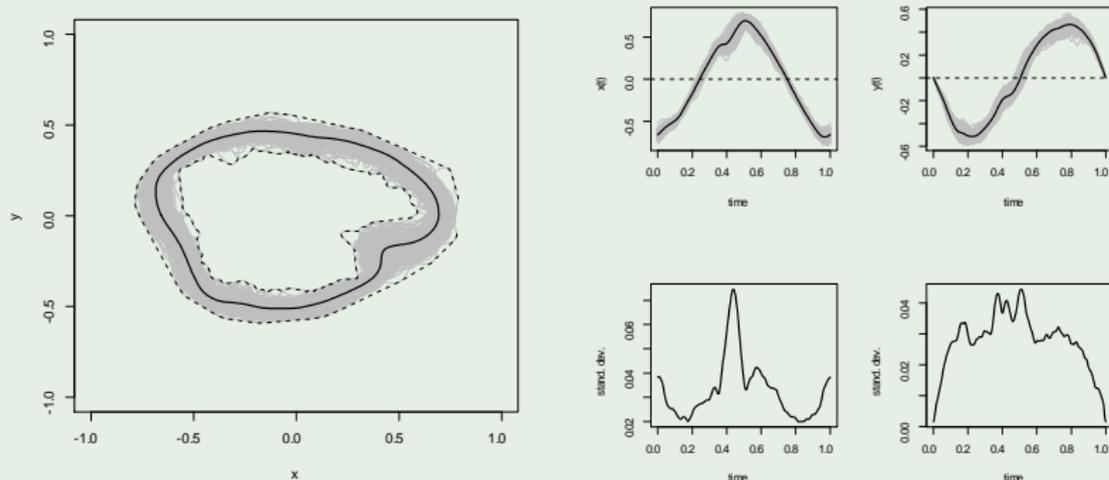
- Séparer au mieux les  $k$  classes a priori sur l'ensemble des individus échantillonnés à partir des  $p$  prédicteurs

# Exemples : Y catégorielle et X fonctionnelle bivariée

## Exemple

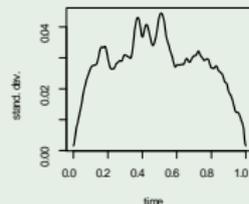
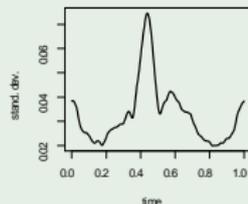
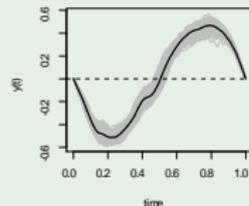
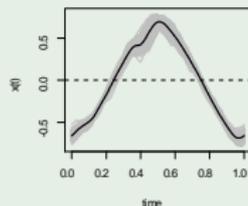
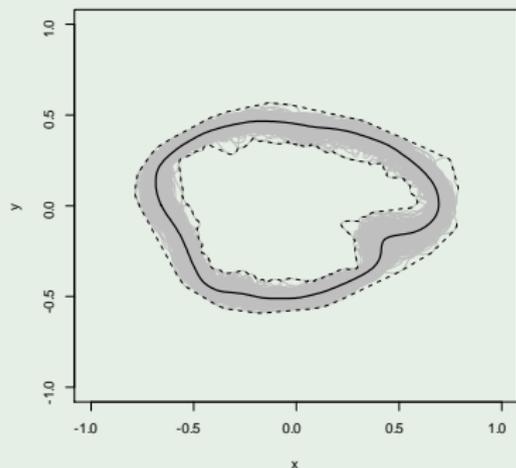


## Example



- The average otolith is defined as  $\{(\bar{x}(t), \bar{y}(t)), t \in [0; 1]\}$

## Example



- The average otolith is defined as  $\{(\bar{x}(t), \bar{y}(t)), t \in [0; 1]\}$
- The variance can be computed and allows to identify variability zones

## Definition

L'ensemble des méthodes qui cherchent à reproduire une variable  $Y$  étant données des observations simultanées de  $\mathbf{X}$  font partie des méthodes d'*apprentissage supervisé*. Il s'agit de construire un modèle

$$y = f_{\theta}(\mathbf{x}) + \varepsilon$$

à partir d'un échantillon d'apprentissage  $E$ .

- Et s'il n'y a pas d'information a priori ?

## Definition

L'ensemble des méthodes qui cherchent à reproduire une variable  $Y$  étant données des observations simultanées de  $\mathbf{X}$  font partie des méthodes d'*apprentissage supervisé*. Il s'agit de construire un modèle

$$y = f_{\theta}(\mathbf{x}) + \varepsilon$$

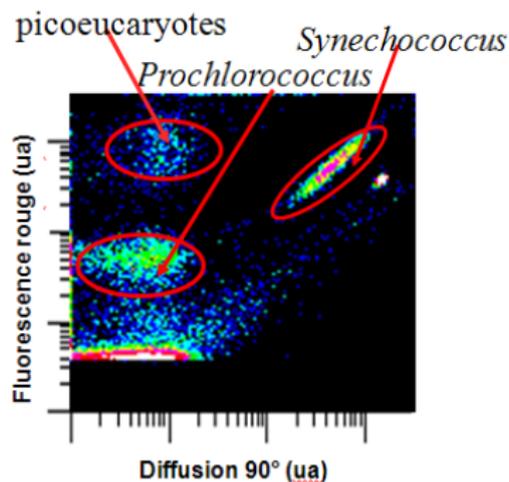
à partir d'un échantillon d'apprentissage  $E$ .

- Et s'il n'y a pas d'information a priori ?

## Definition

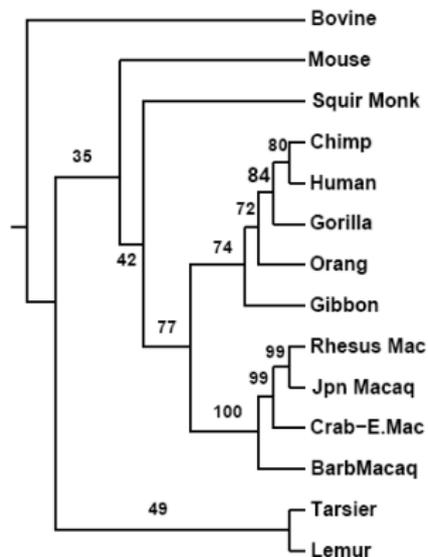
Lorsqu'on cherche à regrouper des objets qui se ressemblent sur la base d'un échantillon des seules variables  $\mathbf{X}$ , on parlera d'*apprentissage non supervisé*. On ne connaît pas a priori l'appartenance d'un individu à une classe qui serait donnée par  $Y$  : il faut déterminer un nombre de classes optimal

## Exemple (cytométrie en flux)



- On utilisera dans ce cas des *méthodes de classification*

# Exemple : X qualitatifs



An example of bootstrap sampling of trees  
232 nucleotide, 14-species mitochondrial D-loop data set  
Analyzed by parsimony, 100 bootstrap replicates

- Un arbre de classification ascendante est construit à partir d'un échantillon de "mots"

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédictives* notées sous la forme d'un vecteur  $\mathbf{X}$ .

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédictives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
- ①  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédictives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
- ①  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
- ②  $\mathbf{X}_{\Theta}$  toutes qualitatives,

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédictives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
- ①  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
- ②  $\mathbf{X}_{\Theta}$  toutes qualitatives,
- ③  $\mathbf{X}_{\mathbb{R} \cup \Theta}$  un mélange de qualitatives et quantitatives.

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédicatives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
- ①  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
- ②  $\mathbf{X}_{\Theta}$  toutes qualitatives,
- ③  $\mathbf{X}_{\mathbb{R} \cup \Theta}$  un mélange de qualitatives et quantitatives.
- La variable à expliquer (à prédire, cible) peut être :

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédicatives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
  - 1  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
  - 2  $\mathbf{X}_{\Theta}$  toutes qualitatives,
  - 3  $\mathbf{X}_{\mathbb{R} \cup \Theta}$  un mélange de qualitatives et quantitatives.
- La variable à expliquer (à prédire, cible) peut être :
  - 1  $Y \in \mathbb{R}$  quantitative,

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédicatives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
  - 1  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
  - 2  $\mathbf{X}_{\Theta}$  toutes qualitatives,
  - 3  $\mathbf{X}_{\mathbb{R} \cup \Theta}$  un mélange de qualitatives et quantitatives.
- La variable à expliquer (à prédire, cible) peut être :
  - 1  $Y \in \mathbb{R}$  quantitative,
  - 2  $Z \in \{0, 1\}$  qualitative à 2 modalités,

- L'apprentissage statistique suppose l'existence d'un *échantillon de construction*  $E$ . Il est constitué de  $p$  *variables explicatives* ou *prédicatives* notées sous la forme d'un vecteur  $\mathbf{X}$ .
  - 1  $\mathbf{X}_{\mathbb{R}}$  toutes quantitatives,
  - 2  $\mathbf{X}_{\Theta}$  toutes qualitatives,
  - 3  $\mathbf{X}_{\mathbb{R} \cup \Theta}$  un mélange de qualitatives et quantitatives.
- La variable à expliquer (à prédire, cible) peut être :
  - 1  $Y \in \mathbb{R}$  quantitative,
  - 2  $Z \in \{0, 1\}$  qualitative à 2 modalités,
  - 3  $T \in \Theta$  qualitative.

- Trois objectifs principaux sont poursuivis dans les applications classiques d'apprentissage :

- Trois objectifs principaux sont poursuivis dans les applications classiques d'apprentissage :
- ① **Exploration multidimensionnelle ou réduction de dimension** : production de graphes, d'un sous-ensemble de variables représentatives des données initiales ou d'un ensemble de composantes préalable à l'utilisation d'une autre technique (ex. Berre)

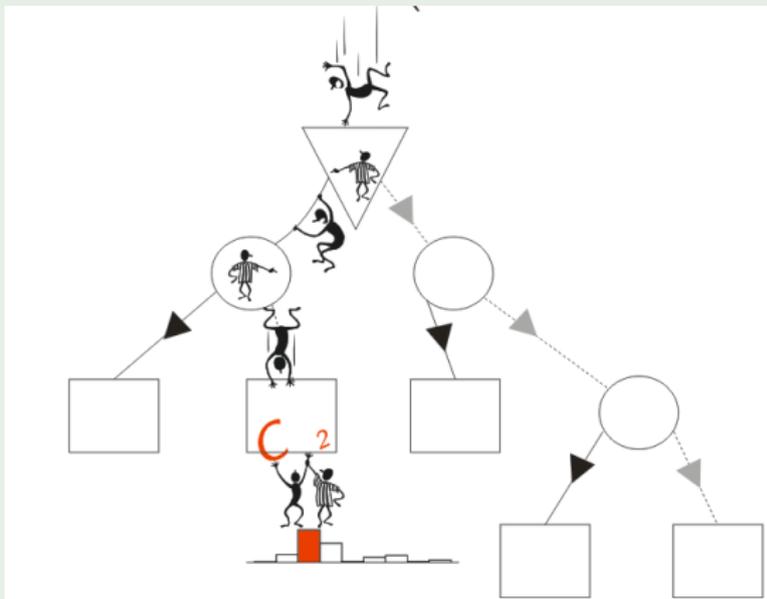
- Trois objectifs principaux sont poursuivis dans les applications classiques d'apprentissage :
- ① **Exploration multidimensionnelle ou réduction de dimension** : production de graphes, d'un sous-ensemble de variables représentatives des données initiales ou d'un ensemble de composantes préalable à l'utilisation d'une autre technique (ex. Berre)
- ② **Classification (clustering) ou segmentation** : production d'une variable qualitative à partir de données quantitatives ou qualitatives.

- Trois objectifs principaux sont poursuivis dans les applications classiques d'apprentissage :
- ① **Exploration multidimensionnelle ou réduction de dimension** : production de graphes, d'un sous-ensemble de variables représentatives des données initiales ou d'un ensemble de composantes préalable à l'utilisation d'une autre technique (ex. Berre)
- ② **Classification (clustering) ou segmentation** : production d'une variable qualitative à partir de données quantitatives ou qualitatives.
- ③ **Modélisation** ( $Y$  ou  $Z$ ) ou **Discrimination** ( $Z$  ou  $T$ ) : production d'un *modèle de prévision* de  $Y$  (resp.  $Z$ ,  $T$ )

- Trois objectifs principaux sont poursuivis dans les applications classiques d'apprentissage :
- ① **Exploration multidimensionnelle ou réduction de dimension** : production de graphes, d'un sous-ensemble de variables représentatives des données initiales ou d'un ensemble de composantes préalable à l'utilisation d'une autre technique (ex. Berre)
- ② **Classification (clustering) ou segmentation** : production d'une variable qualitative à partir de données quantitatives ou qualitatives.
- ③ **Modélisation** ( $Y$  ou  $Z$ ) ou **Discrimination** ( $Z$  ou  $T$ ) : production d'un *modèle de prévision* de  $Y$  (resp.  $Z$ ,  $T$ )
- Une fois ces étapes déterminées, on peut passer à la *sélection* et à l'*utilisation* des modèles construits dans un but prévisionnel.

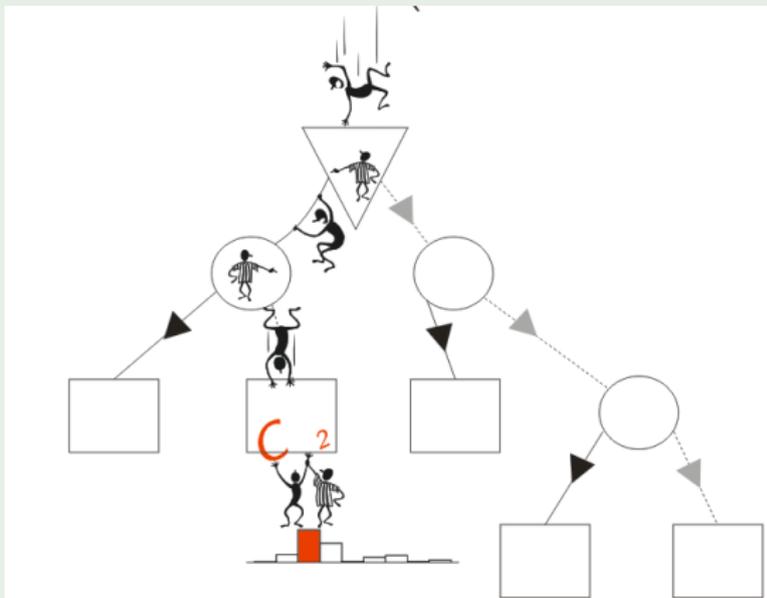
- L'ensemble des méthodes utilisées peuvent servir à faire des prévisions

## Exemple (Arbre de classification)



- L'ensemble des méthodes utilisées peuvent servir à faire des prévisions

## Exemple (Arbre de classification)

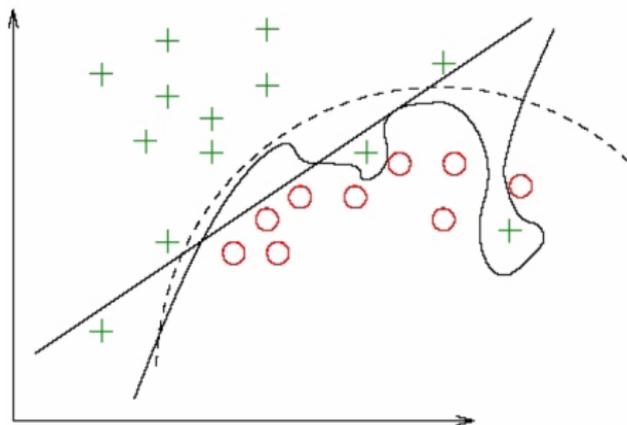


- Comment estimer le bon comportement du modèle?

- Les performances d'un modèle sont basées sur des critères de *qualité de prévision* qui visent la recherche de *modèles parcimonieux*. Ces modèles fournissent un compromis entre une complexité limitée (nombre de paramètres ou flexibilité) et une bonne capacité de généralisation. L'interprétabilité passe au deuxième plan

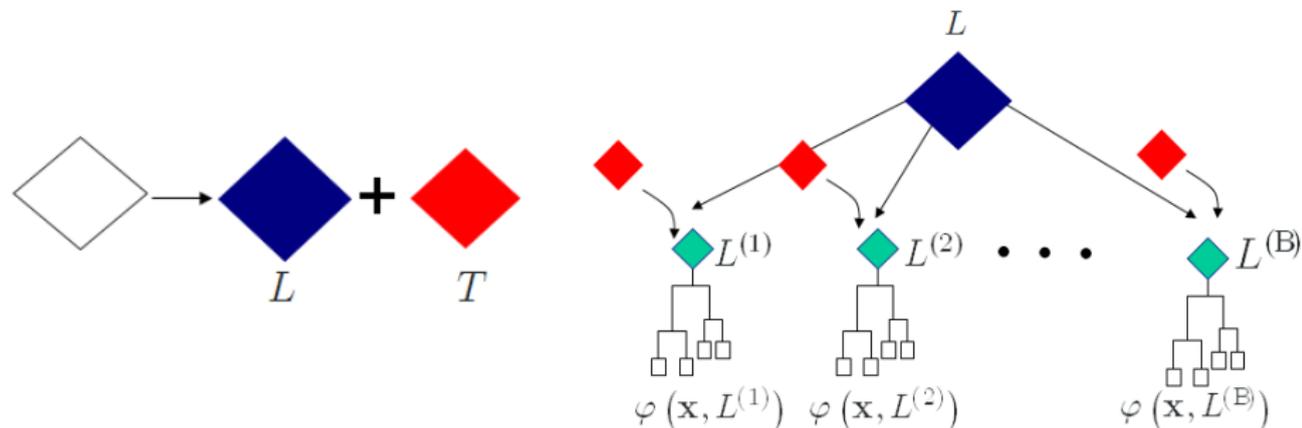
- Les performances d'un modèle sont basées sur des critères de *qualité de prévision* qui visent la recherche de *modèles parcimonieux*. Ces modèles fournissent un compromis entre une complexité limitée (nombre de paramètres ou flexibilité) et une bonne capacité de généralisation. L'interprétabilité passe au deuxième plan

## Exemple



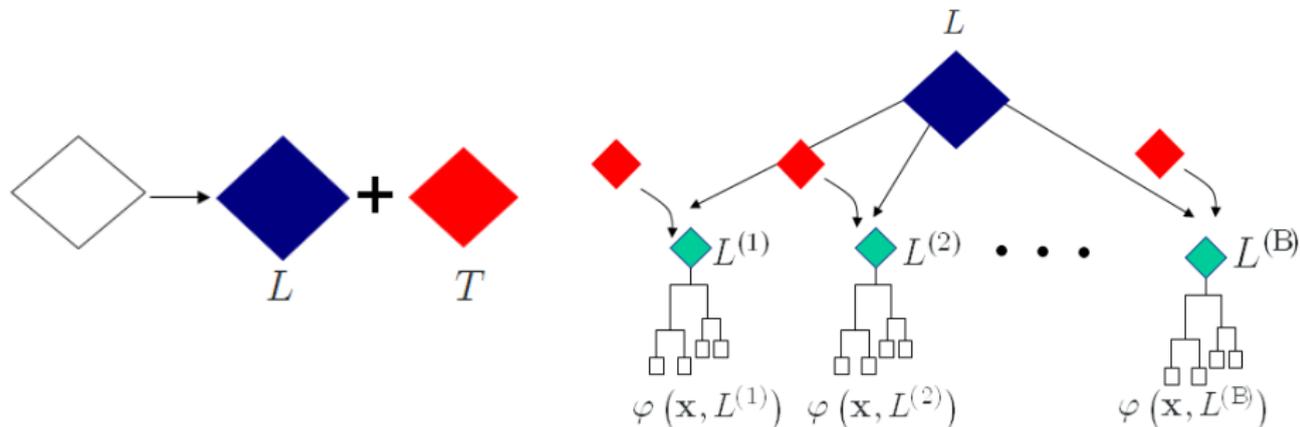
# Modèles agrégés (bagging)

- Il existe un moyen efficace de tester la *capacité de prévision* d'un modèle grâce à la technique du Bootstrap



# Modèles agrégés (bagging)

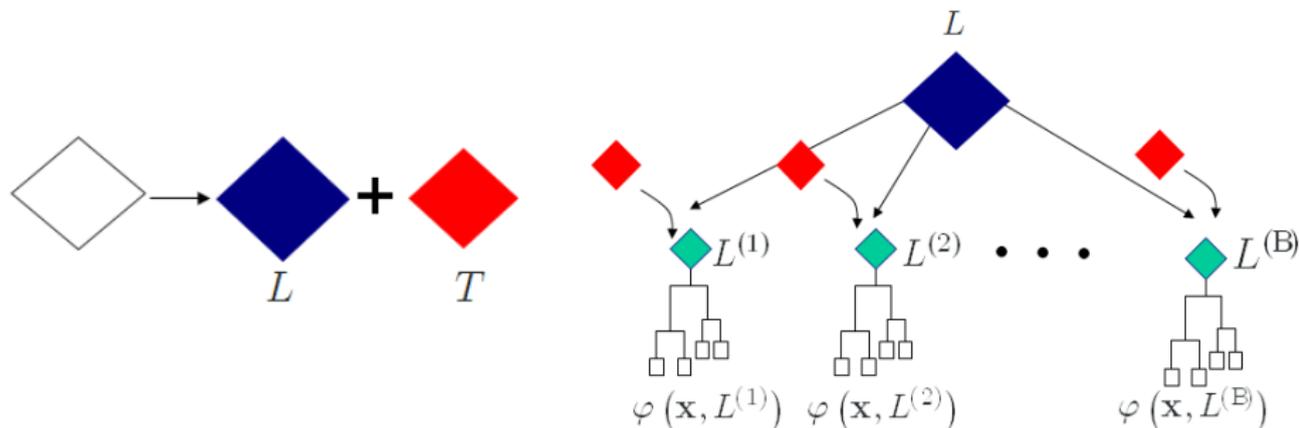
- Il existe un moyen efficace de tester la *capacité de prévision* d'un modèle grâce à la technique du Bootstrap



- La prévision est donnée par  $\varphi_B(\mathbf{x}) = moy_i \left( \varphi(\mathbf{x}, L^{(i)}) \right)$

# Modèles agrégés (bagging)

- Il existe un moyen efficace de tester la *capacité de prévision* d'un modèle grâce à la technique du Bootstrap



- La prévision est donnée par  $\varphi_B(\mathbf{x}) = moy_i \left( \varphi(\mathbf{x}, L^{(i)}) \right)$
- On peut soit utiliser  $B$  fois le même modèle soit en utiliser plusieurs : la stratégie du choix du modèle n'est pas figée

- Les objectifs (ambitieux...) de ce cours sont :

- Les objectifs (ambitieux...) de ce cours sont :
- ① Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent

- Les objectifs (ambitieux...) de ce cours sont :
- 1 Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent
  - 2 Donner les **bases géométriques** nécessaires à la **construction des modèles** et à la compréhension de certaines notions en statistiques

- Les objectifs (ambitieux...) de ce cours sont :
- 1 Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent
  - 2 Donner les **bases géométriques** nécessaires à la **construction des modèles** et à la compréhension de certaines notions en statistiques
  - 3 Aborder quelques méthodes de **classification, segmentation** et d'**aide à la décision** (AFD)

- Les objectifs (ambitieux...) de ce cours sont :
- ① Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent
- ② Donner les **bases géométriques** nécessaires à la **construction des modèles** et à la compréhension de certaines notions en statistiques
- ③ Aborder quelques méthodes de **classification, segmentation** et d'**aide à la décision** (AFD)
- ④ Donner les bases de la construction de modèles agrégés par **bootstrap**

- Les objectifs (ambitieux...) de ce cours sont :
- ① Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent
- ② Donner les **bases géométriques** nécessaires à la **construction des modèles** et à la compréhension de certaines notions en statistiques
- ③ Aborder quelques méthodes de **classification, segmentation** et d'**aide à la décision** (AFD)
- ④ Donner les bases de la construction de modèles agrégés par **bootstrap**
- ⑤ Appliquer quelques unes de ces méthodes sur ordinateur

- Les objectifs (ambitieux...) de ce cours sont :
- ① Comprendre ce qu'est un **tableau de données** et les différentes **variables** qui le composent
- ② Donner les **bases géométriques** nécessaires à la **construction des modèles** et à la compréhension de certaines notions en statistiques
- ③ Aborder quelques méthodes de **classification, segmentation** et d'**aide à la décision** (AFD)
- ④ Donner les bases de la construction de modèles agrégés par **bootstrap**
- ⑤ Appliquer quelques unes de ces méthodes sur ordinateur
- Mon (piètre) objectif est que vous puissiez ouvrir ensuite un livre qui traite de ce sujet sans prendre peur !