

# IA et apprentissage automatique

Philippe Preux

[philippe.preux@univ-lille.fr](mailto:philippe.preux@univ-lille.fr)

<https://philippe-preux.github.io/talks/IA-et-AA.DU-IA-sante.pdf>



# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.

# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.

# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.
- ▶ Concept mouvant  
du cognitivisme au comportementalisme.

# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.
- ▶ Concept mouvant  
du cognitivisme au comportementalisme.
- ▶ Pour les scientifiques, la notion évolue au fur et à mesure des progrès de la mécanisation.

# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.
- ▶ Concept mouvant  
du cognitivisme au comportementalisme.
- ▶ Pour les scientifiques, la notion évolue au fur et à mesure des progrès de la mécanisation.
- ▶  IA est actuellement utilisé à tort et à travers.

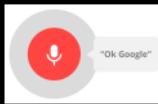
# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.
- ▶ Concept mouvant  
du cognitivisme au comportementalisme.
- ▶ Pour les scientifiques, la notion évolue au fur et à mesure des progrès de la mécanisation.
- ▶  IA est actuellement utilisé à tort et à travers.
- ▶ Avant : IA = manipulation de symboles, de connaissances : IA symbolique

# L'IA, c'est quoi ?

- ▶ non l'IA, ce n'est pas que chatGPT.
- ▶ IA = mécanisme qui donne l'impression de réaliser une tâche que l'on croit que seul l'Homme peut réaliser.
- ▶ Concept mouvant  
du cognitivisme au comportementalisme.
- ▶ Pour les scientifiques, la notion évolue au fur et à mesure des progrès de la mécanisation.
- ▶  IA est actuellement utilisé à tort et à travers.
- ▶ Avant : IA = manipulation de symboles, de connaissances : IA symbolique
- ▶ Aujourd'hui : IA = apprentissage automatique.

# Quelques applications de l'apprentissage automatique



*"man in black shirt  
is playing guitar"*



# IA symbolique vs. AA

IA symbolique	Apprentissage automatique

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation
Logique	Statistiques / Optimisation

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation
Logique	Statistiques / Optimisation
Si A est vrai et $A \Rightarrow B$ , alors B est vrai.	

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation
Logique	Statistiques / Optimisation
Si A est vrai et A => B, alors B est vrai.	$\begin{array}{l} x \quad y \\ 1,03 \quad 2 \\ -4,97 \quad -10 \\ -0,03 \quad 0 \end{array}$ $\hat{y} = 2x$

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation
Logique	Statistiques / Optimisation
Si A est vrai et A => B, alors B est vrai.	$x \quad y$ 1,03 2 -4,97 -10 -0,03 0 $\hat{y} = 2x$
Connaissance exacte Monde déterministe	Connaissance approchée Monde non déterministe

# IA symbolique vs. AA

IA symbolique	Apprentissage automatique
Déduction / Inférence	Induction
Connaissance	Données
Raisonnement	Généralisation
Logique	Statistiques / Optimisation
Si A est vrai et A => B, alors B est vrai.	$x \quad y$ 1,03 2 -4,97 -10 -0,03 0 $\hat{y} = 2x$
Connaissance exacte Monde déterministe	Connaissance approchée Monde non déterministe
Sujet de recherches actives.	

# AA

- ▶ Basé sur les données.

# AA

- ▶ Basé sur les données.
- ▶ Une donnée est un tuple de  $P$  attributs.

## AA

- ▶ Basé sur les données.
- ▶ Une donnée est un tuple de  $P$  attributs.
- ▶ On suppose que les attributs sont quantitatifs.

# AA

- ▶ Basé sur les données.
- ▶ Une donnée est un tuple de  $P$  attributs.
- ▶ On suppose que les attributs sont quantitatifs.

## Idée fondamentale 1 :

Les données vivent dans un espace géométrique à  $P$  dimensions. Elles ne sont pas réparties au hasard de manière uniforme. Au contraire, elles sont situées sur des variétés.

- ▶ Basé sur les données.
- ▶ Une donnée est un tuple de  $P$  attributs.
- ▶ On suppose que les attributs sont quantitatifs.

#### Idée fondamentale 1 :

Les données vivent dans un espace géométrique à  $P$  dimensions. Elles ne sont pas réparties au hasard de manière uniforme. Au contraire, elles sont situées sur des variétés.

#### Idée fondamentale 2 :

Une donnée particulière est un échantillon de cette variété : c'est un point tiré au hasard (pas forcément uniformément) sur cette variété.

# AA

- ▶ Basé sur les données.
- ▶ Une donnée est un tuple de P attributs.
- ▶ On suppose que les attributs sont quantitatifs.

## Idée fondamentale 1 :

Les données vivent dans un espace géométrique à P dimensions. Elles ne sont pas réparties au hasard de manière uniforme. Au contraire, elles sont situées sur des variétés.

## Idée fondamentale 2 :

Une donnée particulière est un échantillon de cette variété : c'est un point tiré au hasard (pas forcément uniformément) sur cette variété.

## Idée fondamentale 3 :

Un attribut est une quantité que l'on mesure : utile et mesurable.

- ▶ Interprétation habituelle :  
les attributs correspondent aux axes d'un repère orthogonal.

# AA

- ▶ Interprétation habituelle :  
les attributs correspondent aux axes d'un repère orthogonal.
- ▶ Attention : c'est une vision totalement fausse.  
Des axes orthogonaux correspondent à des attributs non corrélés linéairement. Or les attributs sont généralement plus ou moins corrélés.

- ▶ Interprétation habituelle :  
les attributs correspondent aux axes d'un repère orthogonal.
- ▶ Attention : c'est une vision totalement fausse.  
Des axes orthogonaux correspondent à des attributs non corrélés linéairement. Or les attributs sont généralement plus ou moins corrélés.
- ▶ Partant d'un repère non orthogonal, on peut construire un repère orthogonal (ACP)  $\Rightarrow$  mieux pour les méthodes mathématiques.  
Mais les axes de celui-ci ne correspondent plus à des attributs mesurés  $\Rightarrow$  problèmes d'interprétation / explicabilité.

- ▶ Interprétation habituelle :  
les attributs correspondent aux axes d'un repère orthogonal.
- ▶ Attention : c'est une vision totalement fausse.  
Des axes orthogonaux correspondent à des attributs non corrélés linéairement. Or les attributs sont généralement plus ou moins corrélés.
- ▶ Partant d'un repère non orthogonal, on peut construire un repère orthogonal (ACP)  $\Rightarrow$  mieux pour les méthodes mathématiques.  
Mais les axes de celui-ci ne correspondent plus à des attributs mesurés  $\Rightarrow$  problèmes d'interprétation / explicabilité.

#### Idée fondamentale 4 :

Toute mesure est entachée d'erreur : précision limitée + erreurs de mesure.

- ▶ Les données vivent dans un espace géométrique  
⇒ distance.

- ▶ Les données vivent dans un espace géométrique  
⇒ distance.

Hypothèse :

Des données proches sont similaires.

# AA

## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.

# AA

## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.

# AA

## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?

# AA

## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.

# AA

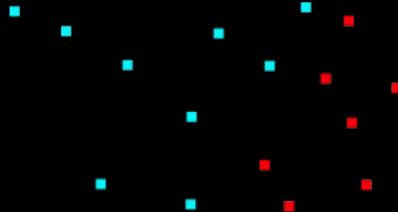
## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.

# AA

## Classification supervisée

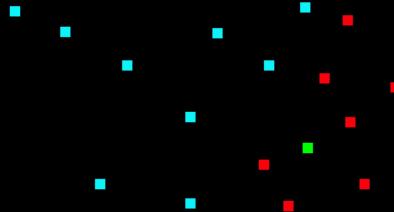
- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.



# AA

## Classification supervisée

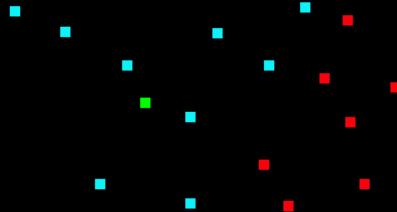
- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.



# AA

## Classification supervisée

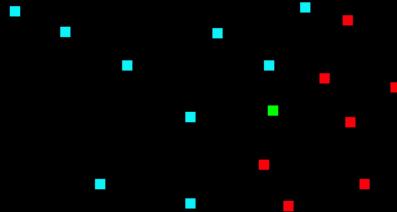
- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.



# AA

## Classification supervisée

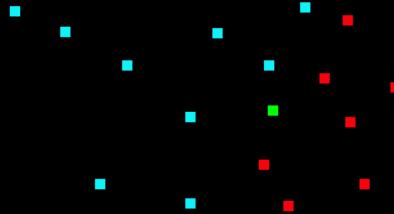
- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.



# AA

## Classification supervisée

- ▶ Chaque donnée est associée à une classe/étiquette/valeur.
- ▶ 1 exemple = 1 donnée et son étiquette.
- ▶ Problème : étant donné un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- ▶ Solution très simple : on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion.



- ▶ ↵ algorithme des plus proches voisins k-NN.

# AA

## Classification supervisée

Autres méthodes :

- ▶ si on suppose que les variétés sont des hyper-rectangles  $\Rightarrow$  arbre de décision

# AA

## Classification supervisée

Autres méthodes :

- ▶ si on suppose que les variétés sont des hyper-rectangles  $\Rightarrow$  arbre de décision
- ▶ si on suppose que l'espace des données est coupé en deux par un hyper-plan : séparateur linéaire : perceptron linéaire, séparateur à vaste marge linéaire, régression logistique, etc.

# AA

## Classification supervisée

Autres méthodes :

- ▶ si on suppose que les variétés sont des hyper-rectangles  $\Rightarrow$  arbre de décision
- ▶ si on suppose que l'espace des données est coupé en deux par un hyper-plan : séparateur linéaire : perceptron linéaire, séparateur à vaste marge linéaire, régression logistique, etc.
- ▶ si on suppose que l'espace est coupé en deux de manière non linéaire : perceptron multi-couches, séparateur à vaste marge, etc.

# Classification supervisée

## Arbre de décision : pavage de l'espace des données

- ▶ variétés = hyper-rectangles

# Classification supervisée

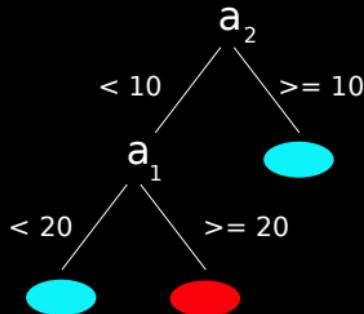
## Arbre de décision : pavage de l'espace des données

- ▶ variétés = hyper-rectangles
- ▶ exemple : données décrites par 2 attributs  $a_1$  et  $a_2$  et leur classe (bleu ou rouge)

# Classification supervisée

## Arbre de décision : pavage de l'espace des données

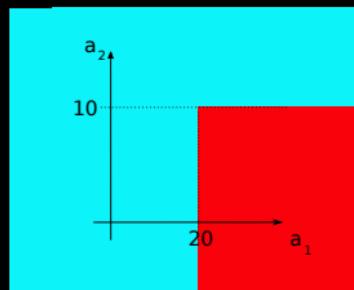
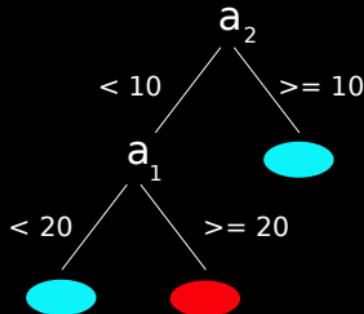
- ▶ variétés = hyper-rectangles
- ▶ exemple : données décrites par 2 attributs  $a_1$  et  $a_2$  et leur classe (bleu ou rouge)



# Classification supervisée

## Arbre de décision : pavage de l'espace des données

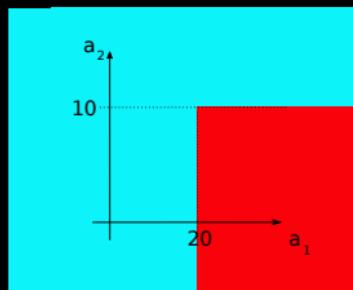
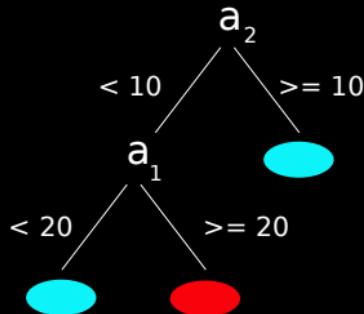
- ▶ variétés = hyper-rectangles
- ▶ exemple : données décrites par 2 attributs  $a_1$  et  $a_2$  et leur classe (bleu ou rouge)



# Classification supervisée

## Arbre de décision : pavage de l'espace des données

- ▶ variétés = hyper-rectangles
- ▶ exemple : données décrites par 2 attributs  $a_1$  et  $a_2$  et leur classe (bleu ou rouge)



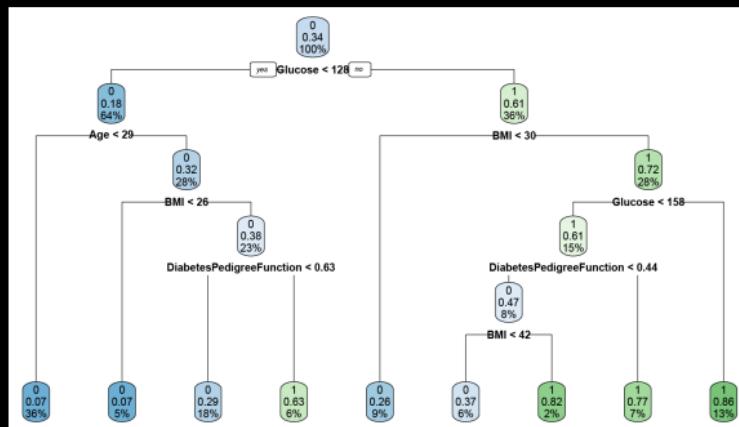
- ▶ Un arbre de décision découpe l'espace des données en hyper-parallélépipèdes.

# Classification supervisée

## Arbre de décision : pavage de l'espace des données

- exemple sur jeu de données « diabète » :

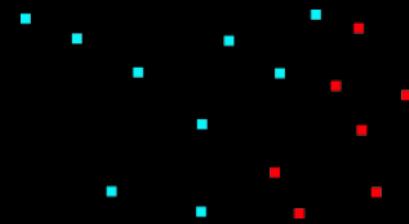
768 exemples décrits par 8 attributs et une classe : *number of times pregnant, plasma glucose concentration, diastolic blood pressure (mm Hg), triceps skin fold thickness (in mm), 2-hr serum insulin measure, body mass index, a diabetes pedigree function, and age* et la classe : patient diabétique ou non.



- avantage : modèle interprétable ; associe une importance à chaque attribut pour la décision quant à la classe de la donnée.

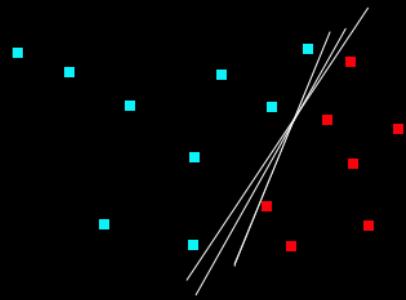
# Classification supervisée

## Séparation linéaire



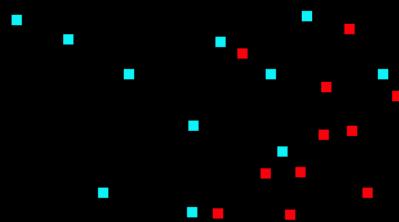
# Classification supervisée

## Séparation linéaire



# Classification supervisée

## Séparation linéaire



- ▶ On cherche une droite qui sépare au mieux.
- ▶ On coupe l'espace des données en 2 par un hyper-plan.
- ▶ Cela revient à chercher des paramètres  $\theta_0 \dots \theta_P$  tels que  $\sum_i \theta_i a_i \geq 0$  pour les bleus,  $< 0$  pour les rouges. On pose  $a_0 = 1$
- ▶ Il y a beaucoup de méthodes pour faire cela.

# Classification supervisée

Perceptron = neurone artificiel

- ▶ On cherche des paramètres  $\theta_0 \dots \theta_P$  tels que  $\sum_i \theta_i a_i \geq 0$  pour les bleus,  $< 0$  pour les rouges.

# Classification supervisée

Perceptron = neurone artificiel

- ▶ On cherche des paramètres  $\theta_0 \dots \theta_P$  tels que  $\sum_i \theta_i a_i \geq 0$  pour les bleus,  $< 0$  pour les rouges.
- ▶ Différents algorithmes permettent de calculer ces  $\theta_i$ .

# Classification supervisée

Perceptron = neurone artificiel

- ▶ On cherche des paramètres  $\theta_0 \dots \theta_P$  tels que  $\sum_i \theta_i a_i \geq 0$  pour les bleus,  $< 0$  pour les rouges.
- ▶ Différents algorithmes permettent de calculer ces  $\theta_i$ .
- ▶ Principe de l'algorithme :
  - ▶ sélectionner au hasard un sous-ensemble d'exemples et calculer pour chacun la classe prédictive par les  $\theta_i$  ;
  - ▶ pour chacun de ces exemples, si la classe prédictive est différente de la classe de l'exemple, corriger les  $\theta_i$  ;
  - ▶ recommencer ces 2 étapes un certain nombre de fois.

# Classification supervisée

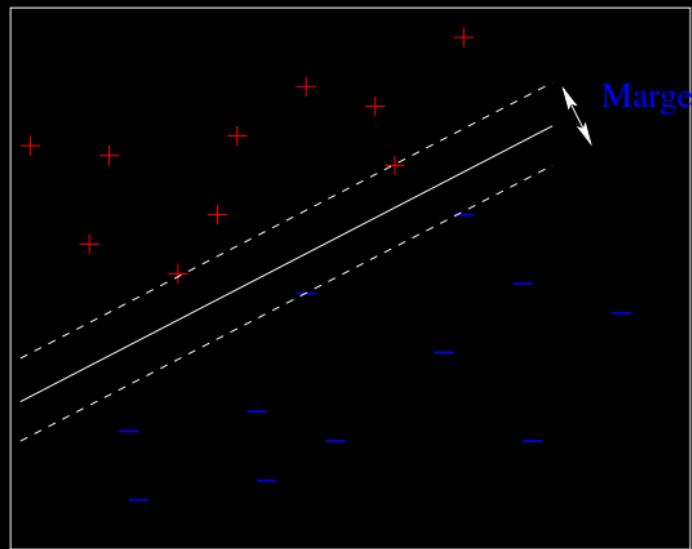
Perceptron = neurone artificiel

- ▶ On cherche des paramètres  $\theta_0 \dots \theta_P$  tels que  $\sum_i \theta_i a_i \geq 0$  pour les bleus,  $< 0$  pour les rouges.
- ▶ Différents algorithmes permettent de calculer ces  $\theta_i$ .
- ▶ Principe de l'algorithme :
  - ▶ sélectionner au hasard un sous-ensemble d'exemples et calculer pour chacun la classe prédictive par les  $\theta_i$  ;
  - ▶ pour chacun de ces exemples, si la classe prédictive est différente de la classe de l'exemple, corriger les  $\theta_i$  ;
  - ▶ recommencer ces 2 étapes un certain nombre de fois.
- ▶ Quand s'arrêter ?
  - ▶ Découper le jeu d'exemples en un jeu d'entraînement  $\mathcal{E}$  et un jeu de test  $\mathcal{T}$ .
  - ▶ Utiliser  $\mathcal{T}$  uniquement pour estimer la qualité de la prédiction.
  - ▶ Arrêter quand l'erreur de prédiction mesurée sur  $\mathcal{T}$  ne diminue plus.

# Classification supervisée

## Machines à vaste marge

- On cherche les paramètres de l'hyper-plan qui maximise la marge :



# Classification supervisée

## Machines à vaste marge

- ▶ théorème de Cover : si on ajoute des attributs qui sont des combinaisons non linéaires des attributs de base, on augmente la probabilité que les données soient linéairement séparables.
- ▶ Pulvérisation des données dans un espace de grande dimension.
- ▶ Idée naïve : on ajoute au hasard de tels attributs et on espère que ça marche.
- ▶ Idée non naïve : on utilise une méthode pour cela avec des fondements mathématiques. Approche basée sur une notion de distance entre les données.
- ▶ Géométriquement : les données pulvérisées sont linéairement séparables.

## Réseaux de neurones : perceptron multi-couches

- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.

## Réseaux de neurones : perceptron multi-couches

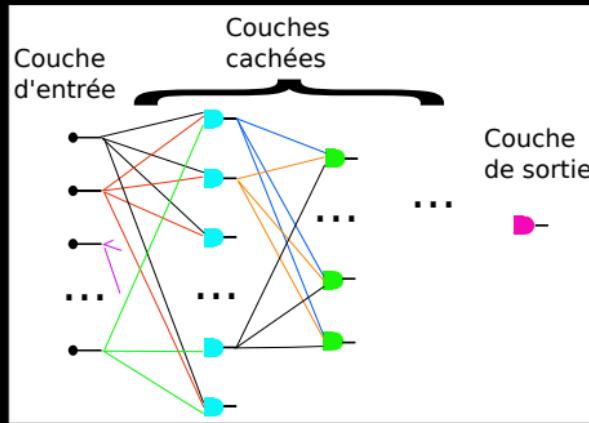
- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.
- ▶ un neurone  $\Rightarrow$  séparatrice = droite

## Réseaux de neurones : perceptron multi-couches

- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.
- ▶ un neurone  $\Rightarrow$  séparatrice = droite
- ▶ plusieurs neurones  $\Rightarrow$  séparatrice de forme arbitraire (continue)

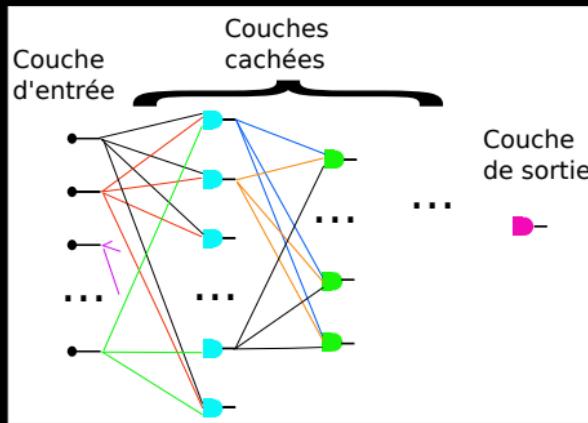
# Réseaux de neurones : perceptron multi-couches

- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.
- ▶ un neurone  $\Rightarrow$  séparatrice = droite
- ▶ plusieurs neurones  $\Rightarrow$  séparatrice de forme arbitraire (continue)



# Réseaux de neurones : perceptron multi-couches

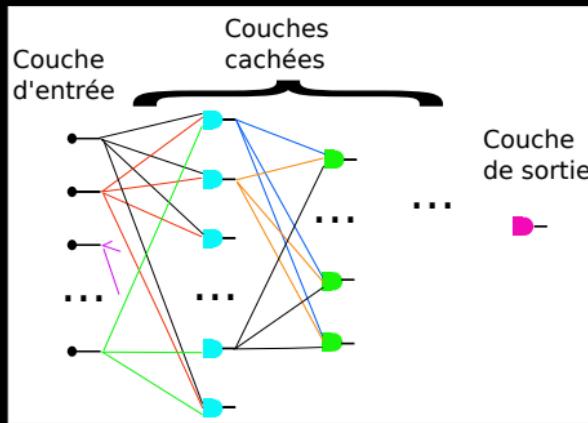
- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.
- ▶ un neurone  $\Rightarrow$  séparatrice = droite
- ▶ plusieurs neurones  $\Rightarrow$  séparatrice de forme arbitraire (continue)



- ▶ un paramètre  $\theta_i$  par connexion

# Réseaux de neurones : perceptron multi-couches

- ▶ géométriquement : la séparation entre les 2 classes peut avoir à peu près n'importe quelle forme.
- ▶ un neurone  $\Rightarrow$  séparatrice = droite
- ▶ plusieurs neurones  $\Rightarrow$  séparatrice de forme arbitraire (continue)



- ▶ un paramètre  $\theta_i$  par connexion
- ▶ pour calculer ces points, même principe que pour le perceptron

## Réseaux de neurones profonds

- Même principe mais avec des dizaines, voire des centaines de couches cachées.

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.
- ▶ Typiquement, le réseau est composé (implicitement) de deux parties :

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.
  
- ▶ Typiquement, le réseau est composé (implicitement) de deux parties :
  - ▶ les premières couches calculent une représentation des données en combinant de manière non linéaire les attributs des données

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
  - ▶ Réseaux récurrents.
  - ▶ Réseaux à convolution.
  
  - ▶ Typiquement, le réseau est composé (implicitement) de deux parties :
    - ▶ les premières couches calculent une représentation des données en combinant de manière non linéaire les attributs des données
    - ▶ les dernières couches déterminent la classe en fonction de cette représentation
- typiquement par combinaison linéaire

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.
  
- ▶ Typiquement, le réseau est composé (implicitement) de deux parties :
  - ▶ les premières couches calculent une représentation des données en combinant de manière non linéaire les attributs des données
  - ▶ les dernières couches déterminent la classe en fonction de cette représentation  
typiquement par combinaison linéaire
- ▶ Thème récurrent en analyse de donnée, apprentissage automatique : utiliser une bonne représentation, adaptée à la tâche à résoudre, est essentiel pour le succès.

# Réseaux de neurones profonds

- ▶ Même principe mais avec des dizaines, voire des centaines de couches cachées.
- ▶ Réseaux récurrents.
- ▶ Réseaux à convolution.
- ▶ Typiquement, le réseau est composé (implicitement) de deux parties :
  - ▶ les premières couches calculent une représentation des données en combinant de manière non linéaire les attributs des données
  - ▶ les dernières couches déterminent la classe en fonction de cette représentation

typiquement par combinaison linéaire
- ▶ Thème récurrent en analyse de donnée, apprentissage automatique : utiliser une bonne représentation, adaptée à la tâche à résoudre, est essentiel pour le succès.
- ▶ Les réseaux profonds nécessitent beaucoup d'exemples.

# Classification supervisée

## Régression logistique

- ▶ Soit  $p$  la probabilité d'être bleu
- ▶ On cherche des paramètres  $\theta_0 \dots \theta_P$  tels que  $\log \frac{p}{1-p} = \sum_i \theta_i a_i$

# Classification supervisée

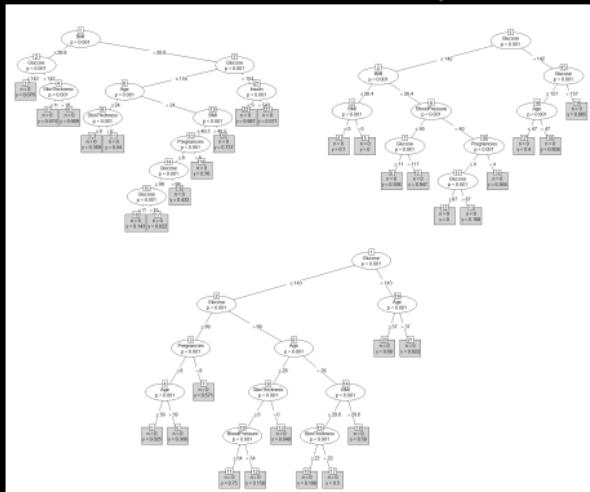
## Forêts

- ▶ Forêt = ensemble d'arbres (de décision)

## Classification supervisée

# Forêts

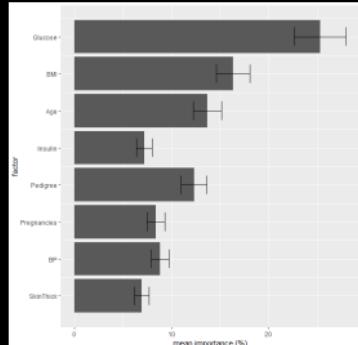
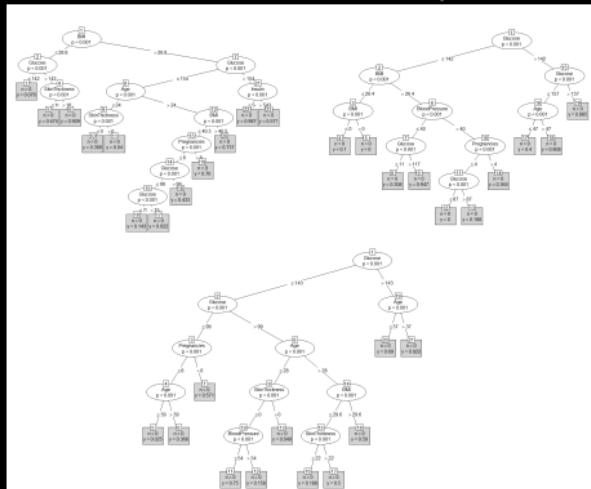
- Forêt = ensemble d'arbres (de décision)



# Classification supervisée

## Forêts

- Forêt = ensemble d'arbres (de décision)

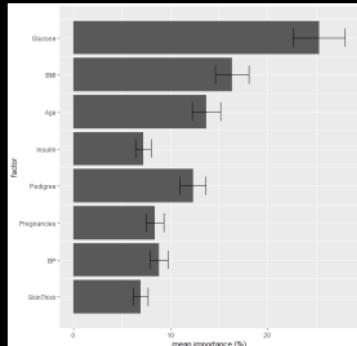
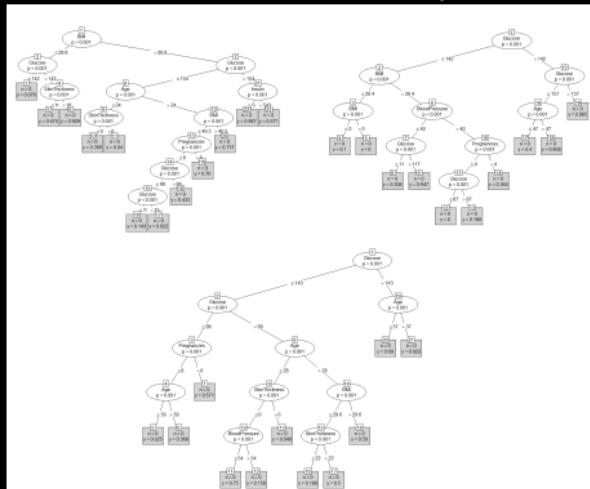


- Différents arbres sélectionnent différents attributs.

# Classification supervisée

## Forêts

- Forêt = ensemble d'arbres (de décision)



- Différents arbres sélectionnent différents attributs.
- Fonctionnent avec un nombre réduit d'exemples.

# AA

## Classification supervisé

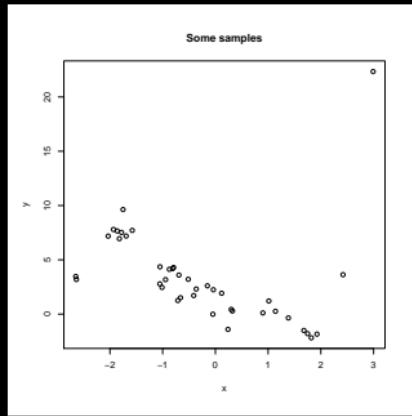
- ▶ Tout modèle possède des paramètres : << apprendre >> = ajuster ces paramètres aux données.

# AA

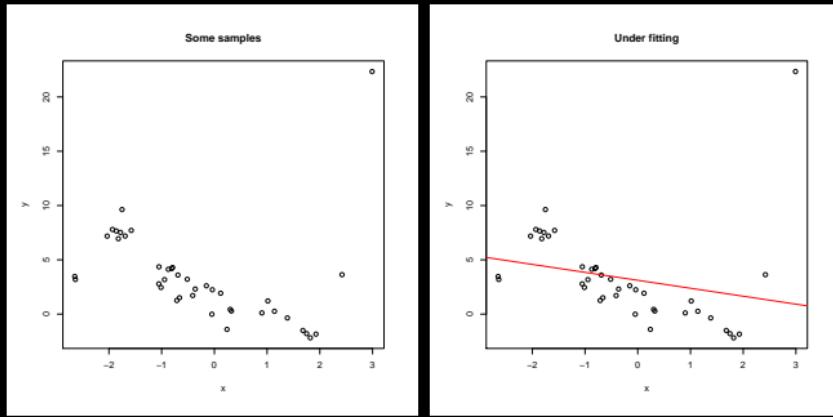
## Classification supervisé

- ▶ Tout modèle possède des paramètres : « apprendre » = ajuster ces paramètres aux données.
- ▶ Le « bon » nombre de paramètres dépend du nombre de données et surtout, de la complexité des variétés sur lesquelles vivent les données.

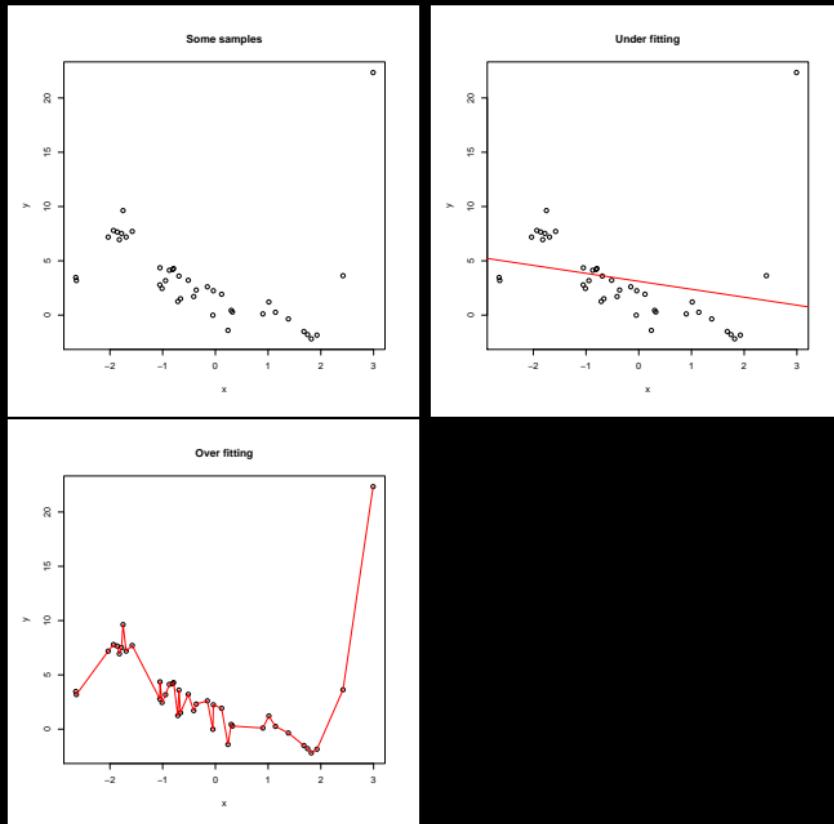
# Apprentissage supervisé



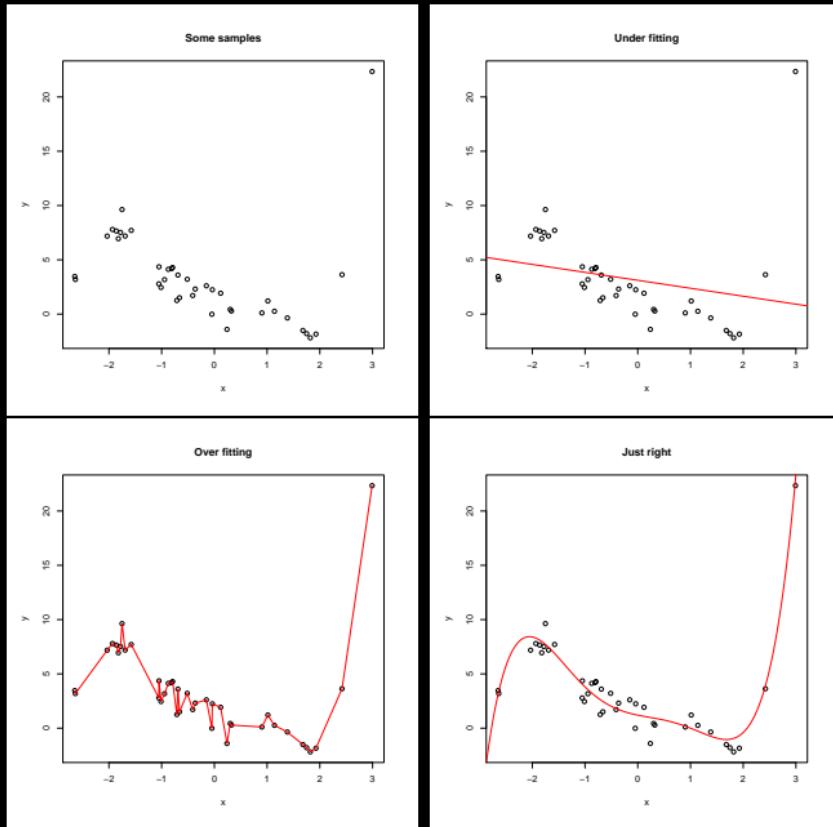
# Apprentissage supervisé



# Apprentissage supervisé



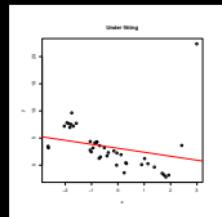
# Apprentissage supervisé



# Apprentissage supervisé

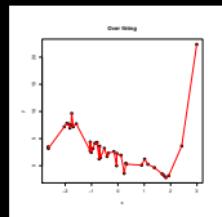
Sous-ajustement/apprentissage (*under-fitting*) : modèle trop simple pour  $y$ .

Le modèle n'a pas assez de paramètres pour s'ajuster à la complexité des données.

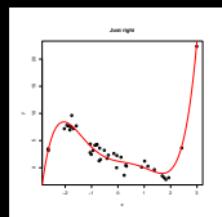


Sur-ajustement/apprentissage (*over-fitting*) : modèle trop complexe pour  $y$ .

Le modèle a trop de paramètres pour la complexité du jeu de données.



Le bon modèle retire juste le bruit.



2 projets collaboratifs lillois récents.

# Projet RAID

## Réseaux de neurones profonds appliqués à la radiologie

- ▶ Application à la radiologie.
- ▶ Projet 2019–2020, Université de Lille, CHU Lille/Service de Radiologie et Imagerie Musculosquelettique, Inria/CRIStAL Scool.
- ▶ Développement d'un outil basé sur un réseau de neurones qui détecte une certaine pathologie du coude chez l'enfant.

# Projet RAID

## Réseaux de neurones profonds appliqués à la radiologie

- ▶ Application à la radiologie.
- ▶ Projet 2019–2020, Université de Lille, CHU Lille/Service de Radiologie et Imagerie Musculosquelettique, Inria/CRIStAL Scool.
- ▶ Développement d'un outil basé sur un réseau de neurones qui détecte une certaine pathologie du coude chez l'enfant.



Radio

# Projet RAID

## Réseaux de neurones profonds appliqués à la radiologie

- ▶ Application à la radiologie.
- ▶ Projet 2019–2020, Université de Lille, CHU Lille/Service de Radiologie et Imagerie Musculosquelettique, Inria/CRIStAL Scool.
- ▶ Développement d'un outil basé sur un réseau de neurones qui détecte une certaine pathologie du coude chez l'enfant.



Radio



Explication/interprétation du diagnostic

# Projet RAID

## Réseaux de neurones profonds appliqués à la radiologie

- ▶ Application à la radiologie.
- ▶ Projet 2019–2020, Université de Lille, CHU Lille/Service de Radiologie et Imagerie Musculosquelettique, Inria/CRIStAL Scool.
- ▶ Développement d'un outil basé sur un réseau de neurones qui détecte une certaine pathologie du coude chez l'enfant.



Radio



Explication/interprétation du diagnostic

- ▶ Étude de l'interaction outil–radiologue.

# Projet RAID

## Étude de l'interaction outil–radiologue

- ▶ Comment l'outil est utilisé par le praticien ?
- ▶ Comment la décision du praticien est-elle influencée par cet outil ?
  
- ▶ Tous les outils sont imparfaits et commettent des erreurs. (Les humains aussi.)
  
- ▶ Thèse de médecine, C. Rozwag, soutenue le 1/7/2021, UdL : pas de conclusion simple quant à l'aide que procure ce type d'outils à un praticien.
  
- ▶ Voir C. Rozwag, F. Valentini, A. Cotten, X. Demondion, Ph. Preux, Th. Jacques, Elbow trauma in children : development and evaluation of radiological artificial intelligence models, *Research in Diagnostic and Interventional Imaging*, **6**, avril 2023, hal-04244410.

# Prédiction de la courbe de poids post-chirurgie bariatrique

Collaboration Inserm U. 1190/CHU de Lille : équipe de F. Pattou  
et  
Inria/CRIStAL Scool (UdL).

# Prédiction de la courbe de poids post-chirurgie bariatrique

Question générale : prédire la courbe de poids à partir des données de la cohorte ABOS : environ 1500 patients, suivis pendant 5 ans. Données recueillies lors de visite à t0 puis 1, 3, 12, 24, 60 mois. 434 attributs.

- ▶ Comment caractériser le poids ? une courbe de poids ?
- ▶ Quelle précision de prédiction peut-on obtenir en utilisant les données ABOS ?
- ▶ Quel modèle utilisé ?
- ▶ Modèle interprétable ?
- ▶ Utilité et utilisation d'un tel modèle ?
- ▶ Validation externe ?

# Prédiction de la courbe de poids post-chirurgie bariatrique

- ▶ Comment caractériser le poids ? une courbe de poids ?
- ▶ Quel modèle utilisé ?
- ▶ Utilité et utilisation d'un tel modèle ?
- ▶ Validation externe ?

# Prédiction de la courbe de poids post-chirurgie bariatrique

- ▶ Comment caractériser le poids ? une courbe de poids ?  
% perte de poids à M1, M3, M12, M24 et M60. + Interpolation
- ▶ Quel modèle utilisé ?
- ▶ Utilité et utilisation d'un tel modèle ?
- ▶ Validation externe ?

# Prédiction de la courbe de poids post-chirurgie bariatrique

- ▶ Comment caractériser le poids ? une courbe de poids ?  
% perte de poids à M1, M3, M12, M24 et M60. + Interpolation
- ▶ Quel modèle utilisé ?  
Arbre de décision : connus pour bien fonctionner dans ce contexte  
(assez peu de données, attributs quantitatifs et nominaux) + résultat interprétable.
- ▶ Utilité et utilisation d'un tel modèle ?
- ▶ Validation externe ?

# Prédiction de la courbe de poids post-chirurgie bariatrique

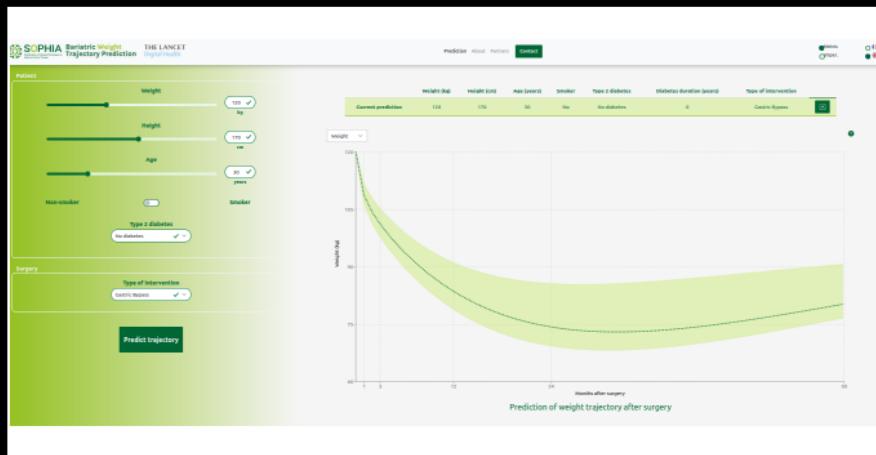
- ▶ Comment caractériser le poids ? une courbe de poids ?  
% perte de poids à M1, M3, M12, M24 et M60. + Interpolation
- ▶ Quel modèle utilisé ?  
Arbre de décision : connus pour bien fonctionner dans ce contexte  
(assez peu de données, attributs quantitatifs et nominaux) + résultat interprétable.
- ▶ Utilité et utilisation d'un tel modèle ?  
Pré-intervention : information du patient ; aide au choix du type d'intervention.  
Pendant le suivi : lors des consultations, informer le patient, lui montrer comment il se situe → conseils, etc.
- ▶ Validation externe ?

# Prédiction de la courbe de poids post-chirurgie bariatrique

- ▶ Comment caractériser le poids ? une courbe de poids ?  
% perte de poids à M1, M3, M12, M24 et M60. + Interpolation
- ▶ Quel modèle utilisé ?  
Arbre de décision : connus pour bien fonctionner dans ce contexte  
(assez peu de données, attributs quantitatifs et nominaux) + résultat interprétable.
- ▶ Utilité et utilisation d'un tel modèle ?  
Pré-intervention : information du patient ; aide au choix du type d'intervention.  
Pendant le suivi : lors des consultations, informer le patient, lui montrer comment il se situe → conseils, etc.
- ▶ Validation externe ?  
Possible si les attributs utilisés par l'arbre sont utilisées dans d'autres cohortes. I.e., ce sont des attributs « simples » que tout le monde mesure.

# Prédiction de la courbe de poids post-chirurgie bariatrique

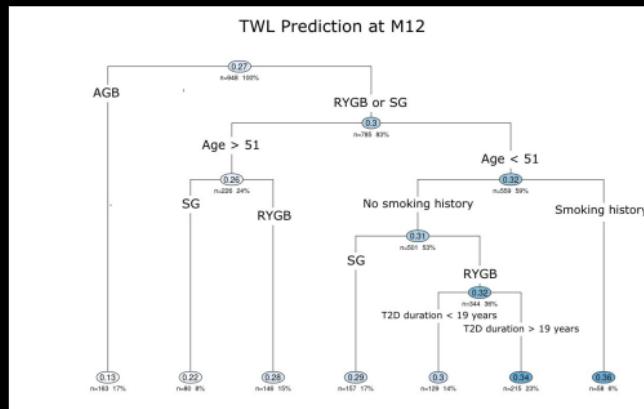
- ▶ 1 arbre de décision par instant de prédiction → 5 arbres.
- ▶ 7 attributs sélectionnés automatiquement, simples.  
Validé par une approche LASSO de sélection de variables qui a sélectionné les mêmes 7 attributs.



<https://bariatric-weight-trajectory-prediction.univ-lille.fr/>

# Prédiction de la courbe de poids post-chirurgie bariatrique

- ▶ 1 arbre de décision par instant de prédiction → 5 arbres.
- ▶ 7 attributs sélectionnés automatiquement, simples.  
Validé par une approche LASSO de sélection de variables qui a sélectionné les mêmes 7 attributs.



<https://bariatric-weight-trajectory-prediction.univ-lille.fr/>

# Prédiction de la courbe de poids post-chirurgie bariatrique

Ingrédients essentiels :

- ▶ disponibilité des données,
- ▶ ingénierie sur les données :
  - ▶ augmenter leur qualité,
  - ▶ identifier l'impact d'événements extérieurs (e.g. covid),
- ▶ travail multi-disciplinaire,
- ▶ réflexions sur l'utilisation du modèle,
- ▶ réflexions sur la manière de présenter le modèle,
- ▶ validation externe sur une dizaine de cohortes dans le monde.

Publié dans Saux *et al.*, *The Lancet Digital Health*, Sep. 2023.

# Segmentation, catégorisation, *clustering*, *embedding*

## Apprentissage non supervisé

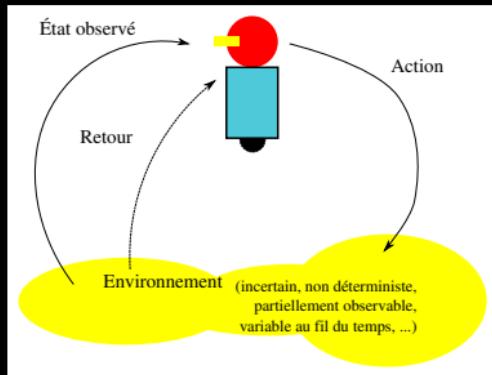
« non supervisé » = pas d'étiquette, juste des données.

- Segmentation de données, catégorisation de données, *Clustering* :
  - ▶ Objectif : regrouper les données qui se ressemblent.
  - ▶ Question : combien y a-t-il de groupes ?
- *Embedding* :
  - ▶ Objectif : trouver la variété où sont localisées les données.  
Cas particulier : ACP.
- Génération de données synthétiques qui ressemblent à des vraies (e.g. GANs).

Une distance est au cœur de ces méthodes.

# Prise de décision séquentielle dans l'incertain

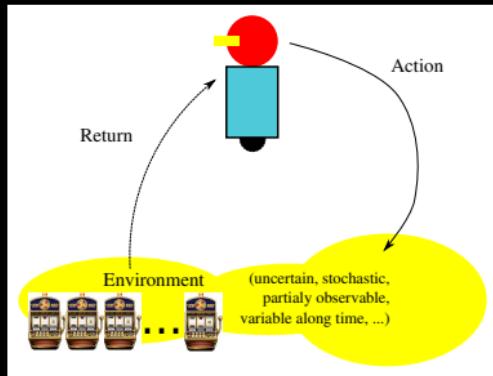
## Apprentissage par renforcement



- ▶ L'agent apprend à « faire quelque chose ».
- ▶ Apprentissage par essai-erreur.
- ▶ L'apprentissage supervisé est un outil pour l'AR.
- ▶ Utilisation de réseaux de neurones.
- ▶ Collaboration avec une équipe et une entreprise pour le contrôle *in vivo* de robots mous.

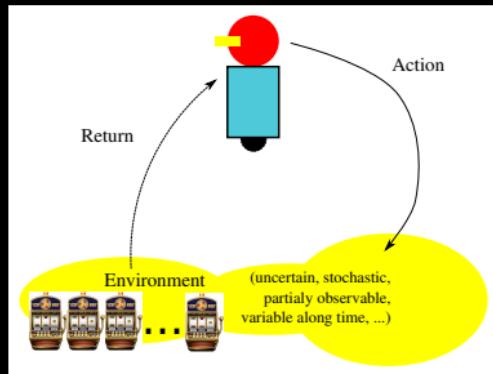
# Prise de décision séquentielle dans l'incertain

## Bandits



- ▶ Formalisme introduit en 1932 dans le cadre des essais cliniques.
- ▶ L'agent apprend quelle est la meilleure action/option parmi  $n$ .
- ▶ Nombreuses applications, notamment en médecine : détermination du meilleur traitement, dosage d'un médicament, traitement adaptatif, repositionnement de médicament, etc.

# Prise de décision séquentielle dans l'incertain



## ► ANR BIP-UP :

- 2023–2027, Université de Lille, CHU Lille, Inserm U. 1190, Inria/CRIStAL Scool.
- Bandits appliqués au suivi post-opératoire de patients.
- Prédiction de la prochaine visite utile.
- Difficultés :
  - faible quantité de données pour entraîner le modèle.
  - prise en compte du risque dans la prise de décision.

Autres travaux en cours : bandits pour les essais cliniques.

# Message pour la maison

- ▶ Beaucoup de modèles d'apprentissage supervisé : ils ont chacun leurs qualités et leurs défauts.
- ▶ Face à une étude de cas réelle :
  - ▶ visualiser les données,
  - ▶ ingénierie des données essentielles,
  - ▶ essayer plusieurs méthodes et comparer les résultats.
- ▶ La distance (explicite ou implicite) utilisée est capitale.  
Cette distance est fonction des attributs des données.
- ▶ Les bandits fonctionnent.
- ▶ Les meilleurs logiciels d'apprentissage automatique sont tous gratuits et accessibles librement pour les activités publiques.

R !