



IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



IA et apprentissage automatique

Lorenzo HERMEZ

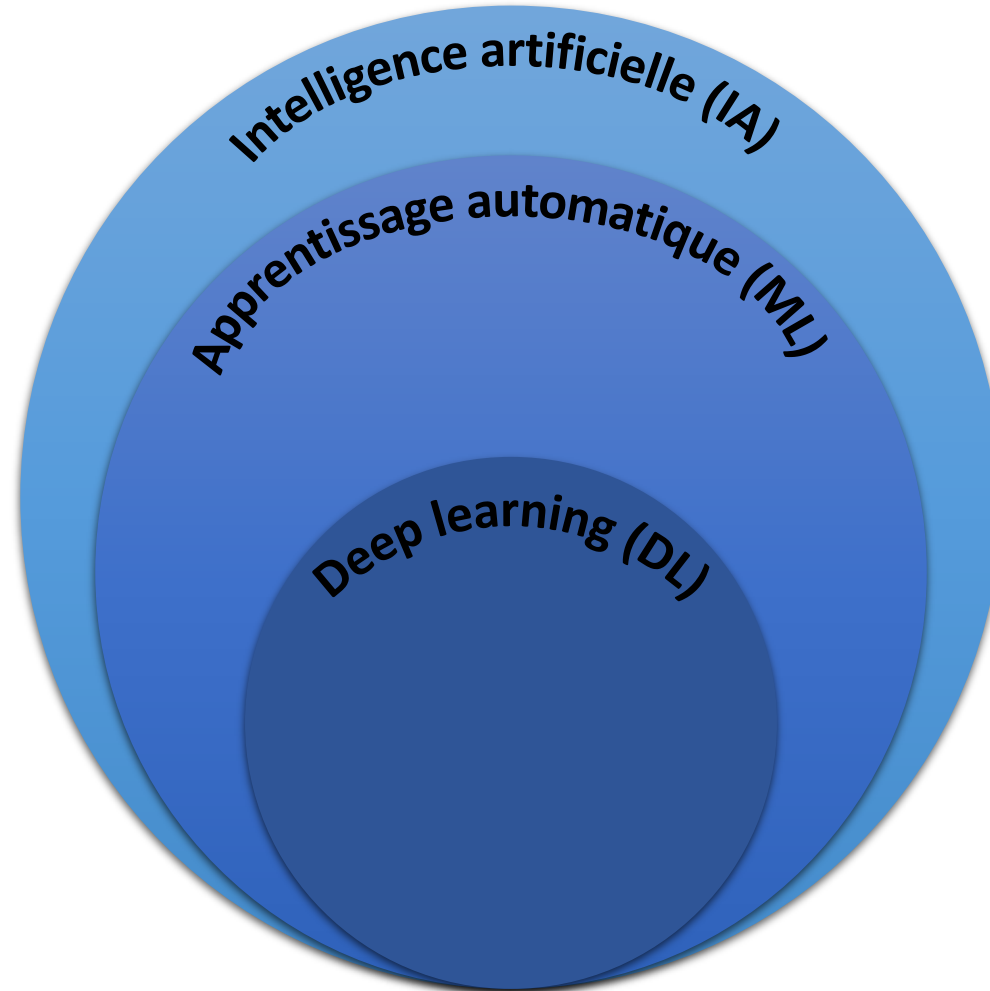
lorenzo.hermez@imt-nord-europe.fr

IMT Nord Europe, CERI SN, HIDE

#IMTomorrow
#IMTNordEurope

26 mars 2026

IA vs. apprentissage automatique

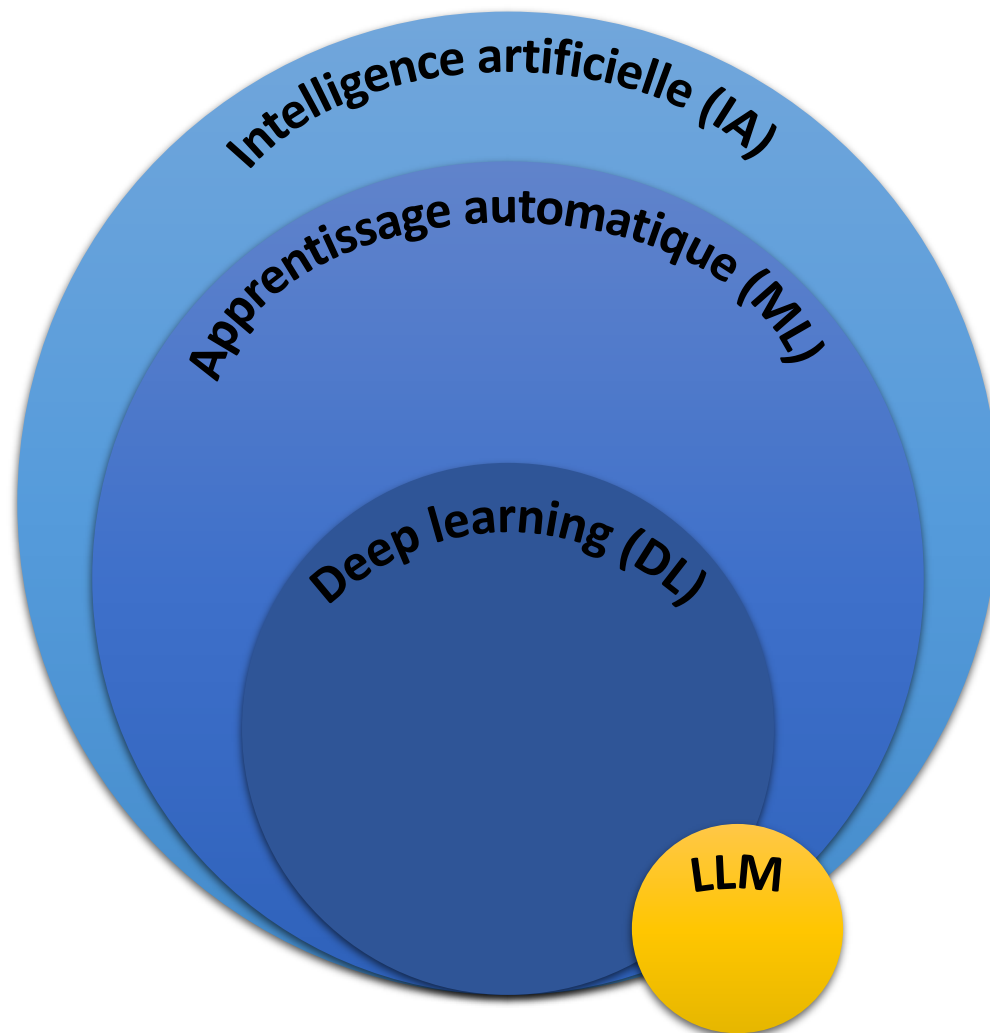




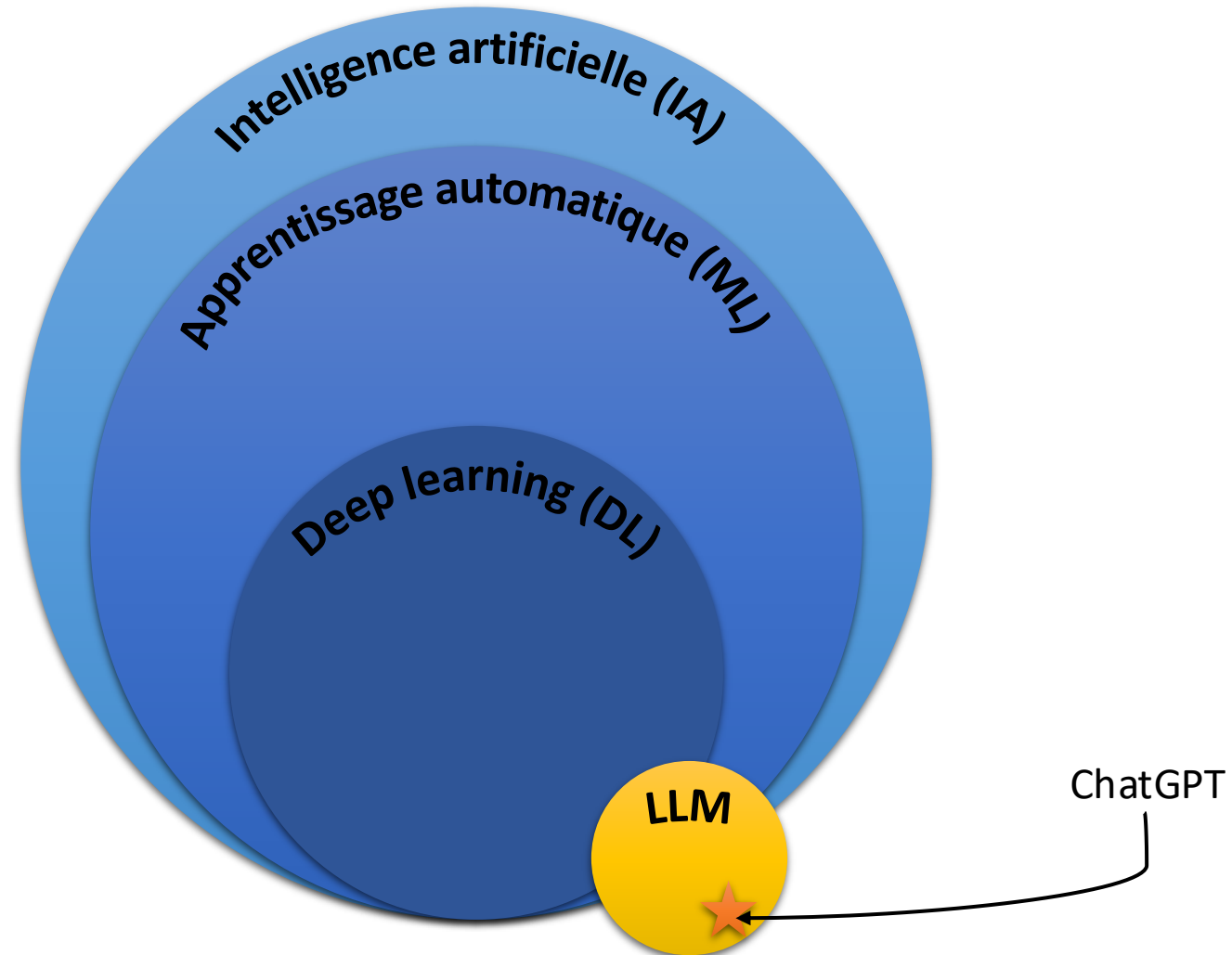
IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



IA vs. apprentissage automatique



IA vs. apprentissage automatique



Qu'est-ce qu'un modèle?



Un modèle est toujours entraîné à l'aide d'une **fonction de coût** permettant d'évaluer la qualité des prédictions.



IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille

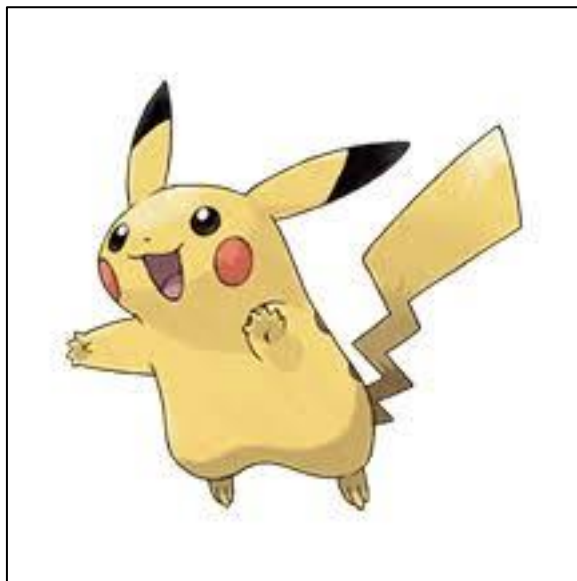


3 types d'apprentissage

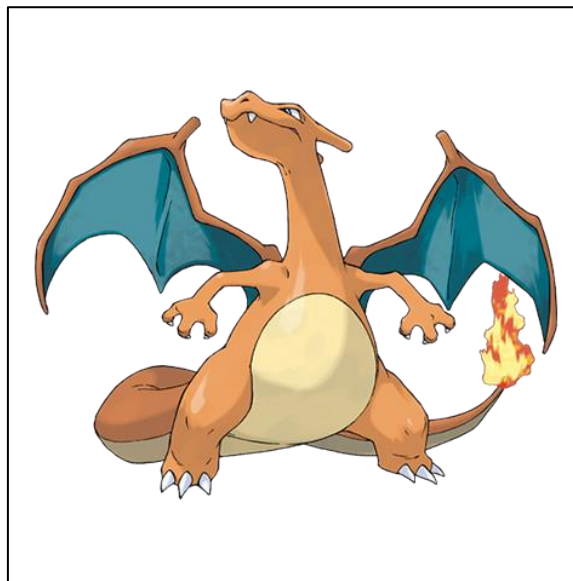
1) Apprentissage supervisé

Utilise des étiquettes d'entrée (X, y) pour entraîner les modèles et déterminer les sorties de futures données.

Apprentissage supervisé



Pikachu



Dracaufeu



Evoli

Apprentissage supervisé

Classification
prédiction discrète

Régression
prédiction continue

Classification
binaire

- Malade/Sain
- Positif/Négatif
- Crédit/Pas crédit
- Comestible/Non comestible

Classification
multiclasses

- Modèle de voitures
- Espèces de fruits
- Mots

- Température
- Salaire post-études
- Perte de poids post-chirurgie

3 types d'apprentissage

1) Apprentissage supervisé

Utilise des étiquettes d'entrée (X, y) pour entraîner les modèles et déterminer les sorties de futures données.

2) Apprentissage non-supervisé

Utilise des données non étiquetées X pour identifier des tendances dans les données.



IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



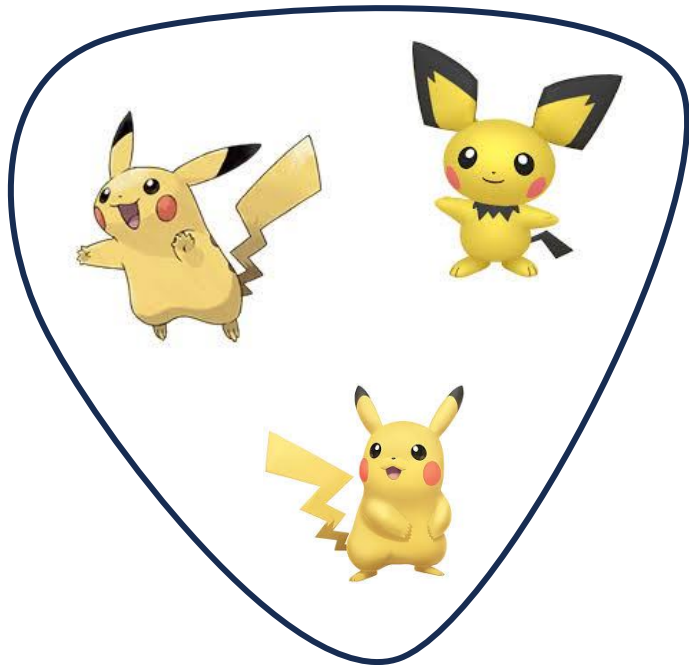
RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

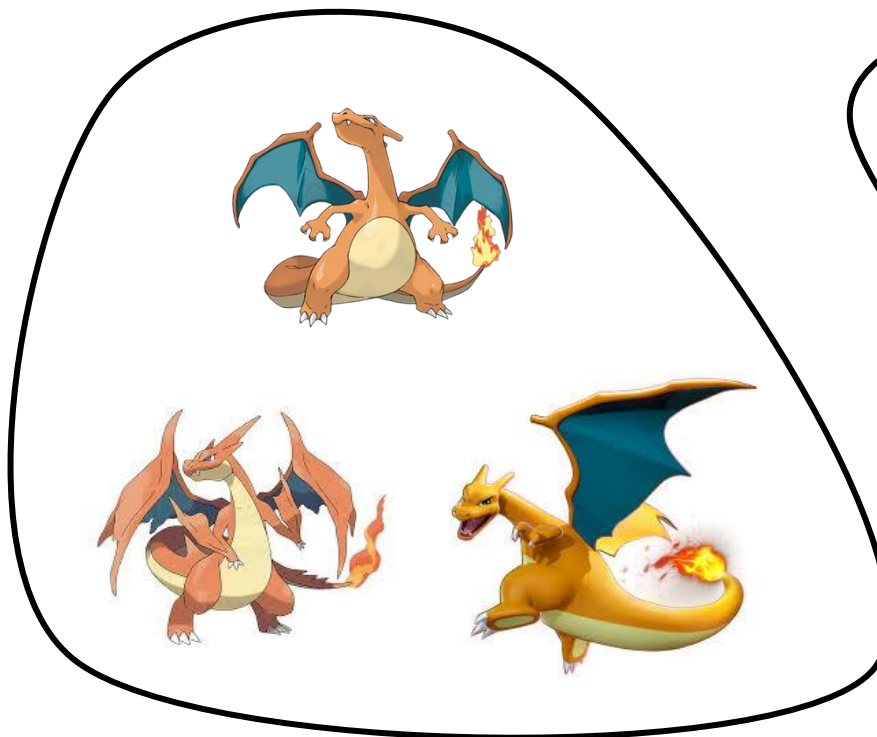
Apprentissage non-supervisé

Clustering

Cluster 1



Cluster 2



Cluster 3



Apprentissage non-supervisé

Clustering regroupement de données

- Identifier des groupes de patients similaires
- Identifier des protéines ayant des propriétés similaires

Détection de valeurs aberrantes¹

- Identifier d'éventuelles erreurs dans les données
- Identifier des patients atypiques

Réduction de la dimension

- Visualiser et comparer des données de grandes dimensions

¹ Samariya, D., Ma, J., Aryal, S., & Zhao, X. (2023). Detection and explanation of anomalies in healthcare data. *Health information science and systems*, 11(1), 20. <https://doi.org/10.1007/s13755-023-00221-2>

3 types d'apprentissage

1) Apprentissage supervisé

Utilise des étiquettes d'entrée (X, y) pour entraîner les modèles et déterminer les sorties de futures données.

2) Apprentissage non-supervisé

Utilise des données non étiquetées X pour identifier des tendances dans les données.

3) Apprentissage par renforcement

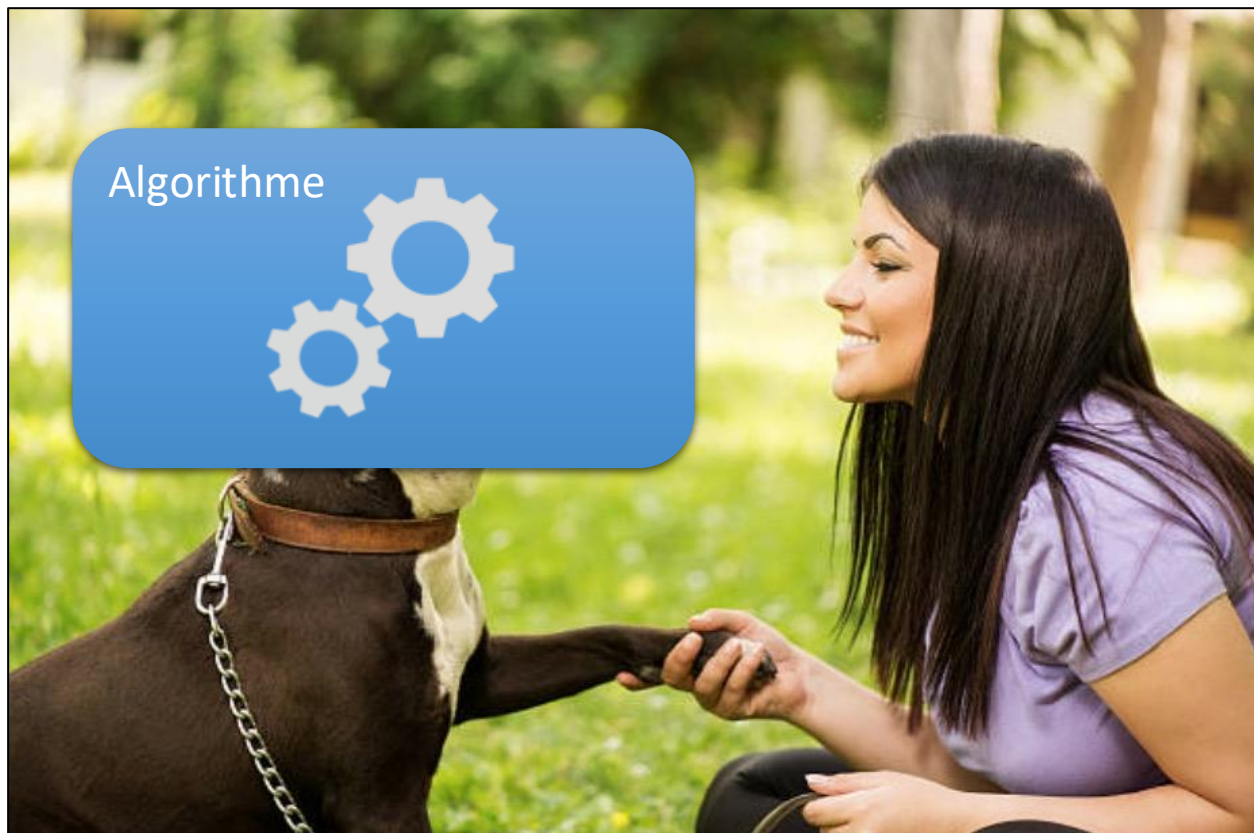
Apprentissage par agent dans un environnement interactif basé sur des récompenses et des pénalités.



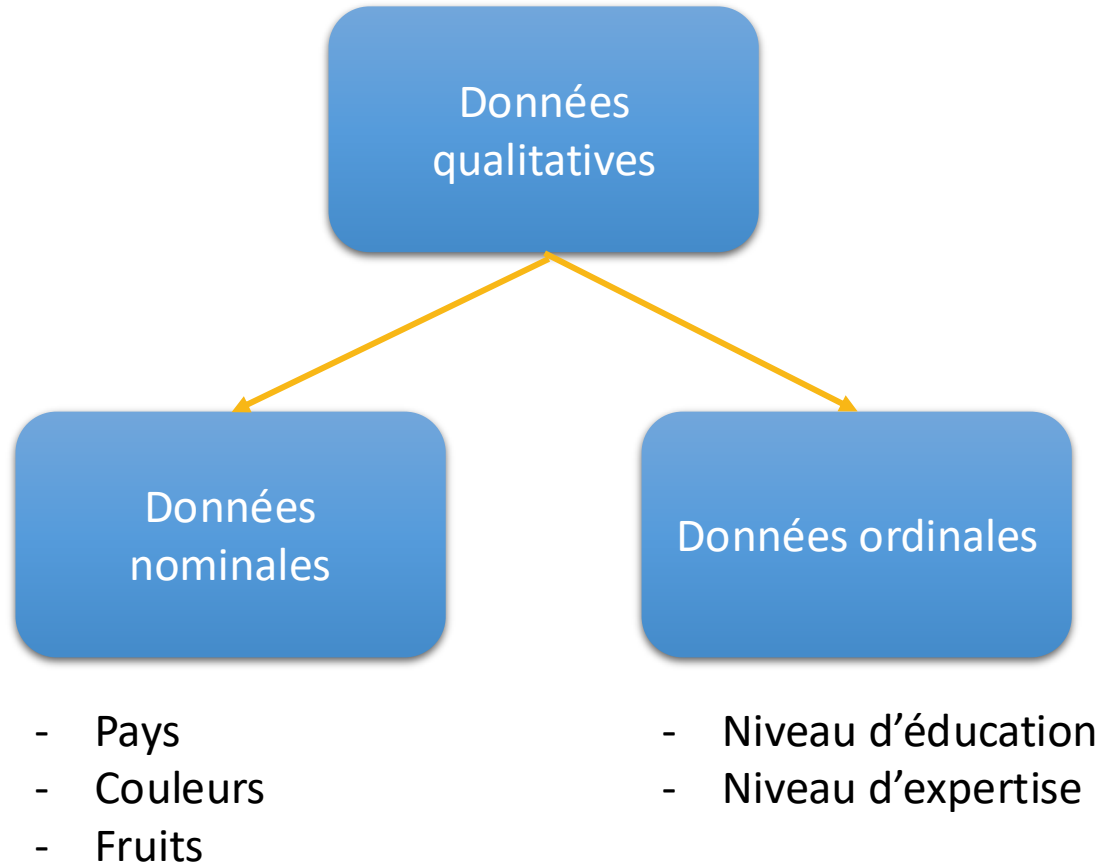
IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Apprentissage par renforcement



Les types de données



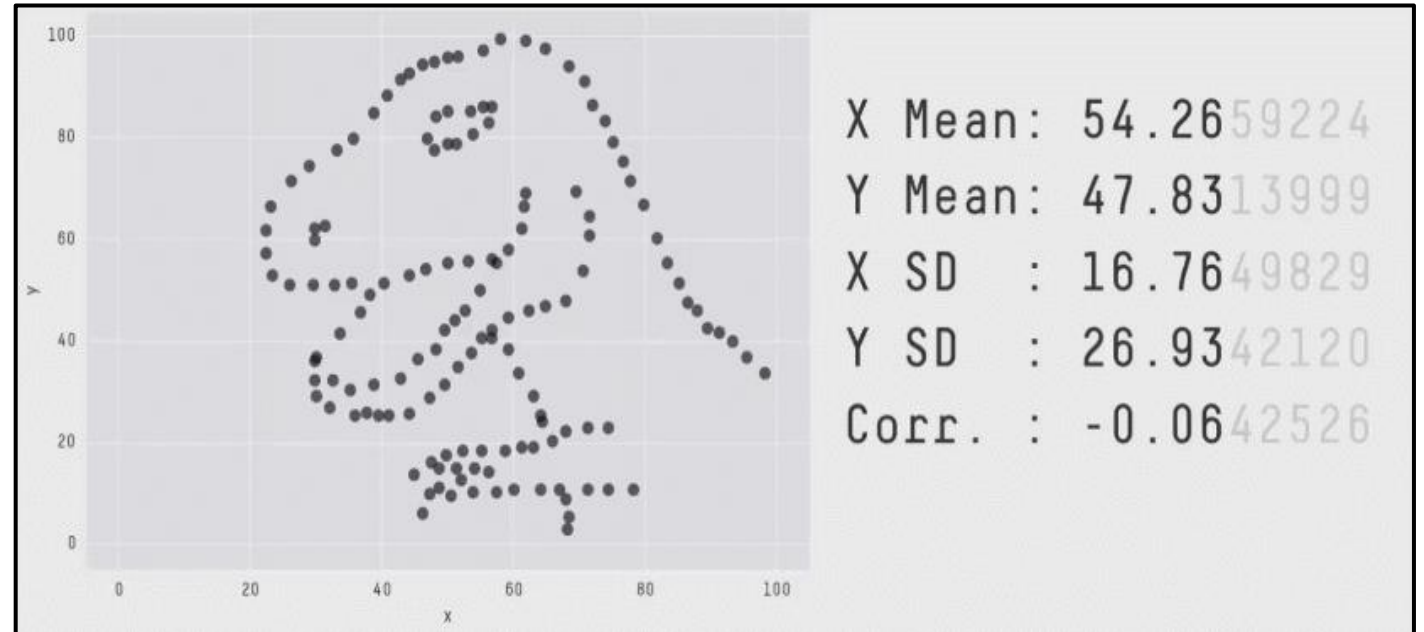
- Age
- Poids
- Température



Visualisation de données

Le plus important

- Les résumés statistiques ne suffisent pas à caractériser vos données
 - Même la corrélation ne suffit pas
- Vous devez **comprendre** les données que vous étudiez pour éviter les biais





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Apprentissage supervisé

#IMTtomorrow
#IMTNordEurope

Classification supervisée

- Chaque donnée est associée à une classe/étiquette.
- 1 exemple = une donnée et son étiquette (X, y) .
- Problème : étant donnée un ensemble d'exemples, comment prédire l'étiquette d'une donnée quelconque ?
- **Solution très simple (non-supervisée)**: on regarde l'étiquette des exemples les plus proches de la donnée à prédire et on tire une conclusion (vote majoritaire) => algorithme des plus proches voisins (k-NN)



Classification supervisée

k-NN

- Données

$X = [x_1, x_2]$ où x_1 est l'épargne du conjoint 1 et x_2 l'épargne du conjoint 2.

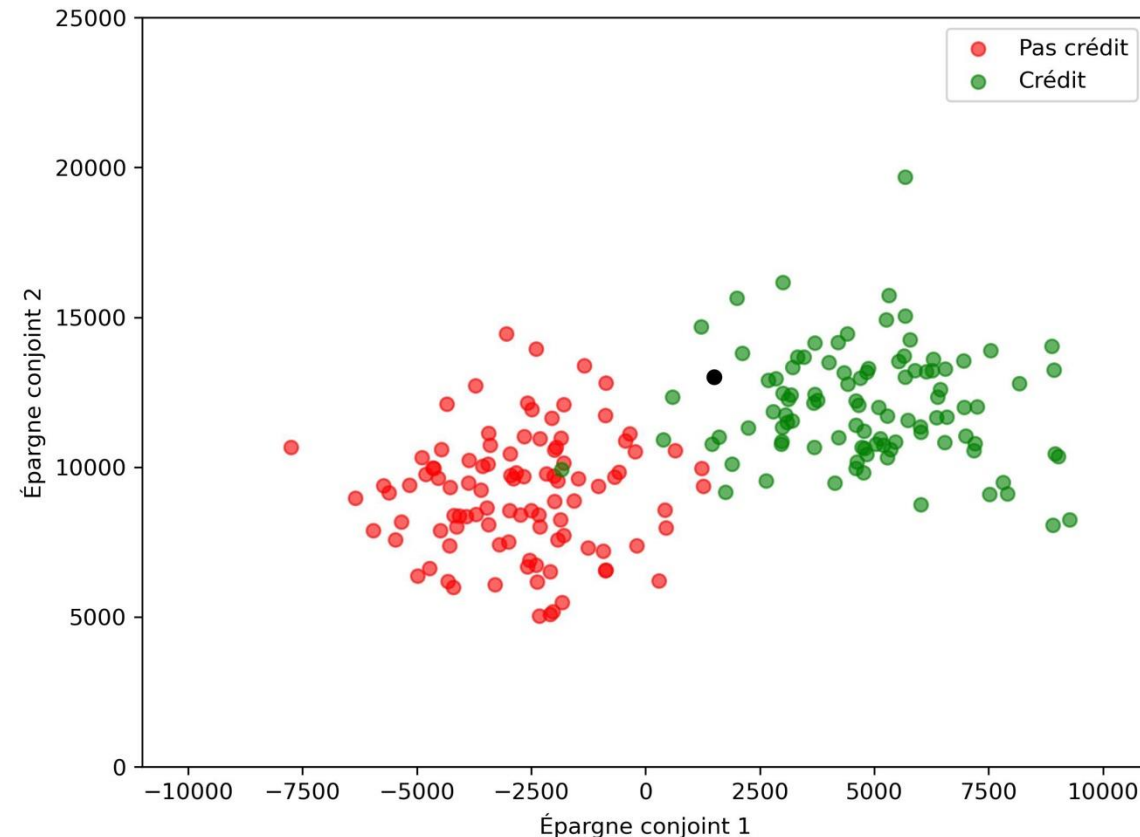
- Objectif

prédire si un couple aura un crédit ou pas

- Inférence

prédit à l'aide des **k voisins les plus proches**

→ Notion de distance pour mesurer la **similarité** des données



Notions de distances

La distance classique est la **distance Euclidienne**: $d_2(x, \tilde{x}) = \sqrt{\sum_{i=1}^d (x_i - \tilde{x}_i)^2}$

→ Ce n'est pas toujours la meilleure distance à utiliser !

Autres distances:

- **distance de Manhattan**: $d_1(x, \tilde{x}) = \sum_{i=1}^d |x_i - \tilde{x}_i|$

ex: analyse d'expression génique, analyse d'IRM ou comparaison d'histogrammes de textures de tissus, ...

- **distance de Minkowski**: $d_p(x, \tilde{x}) = \sqrt[p]{\sum_{i=1}^d (x_i - \tilde{x}_i)^p}$

ex: algorithme k-NN pour diagnostic clinique, extraction de caractéristiques en radiomique, clustering de patients, ...



IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Classification supervisée

Autres exemples

Il n'y a pas d'algorithme magique !

➤ dépend du type de données en entrée.

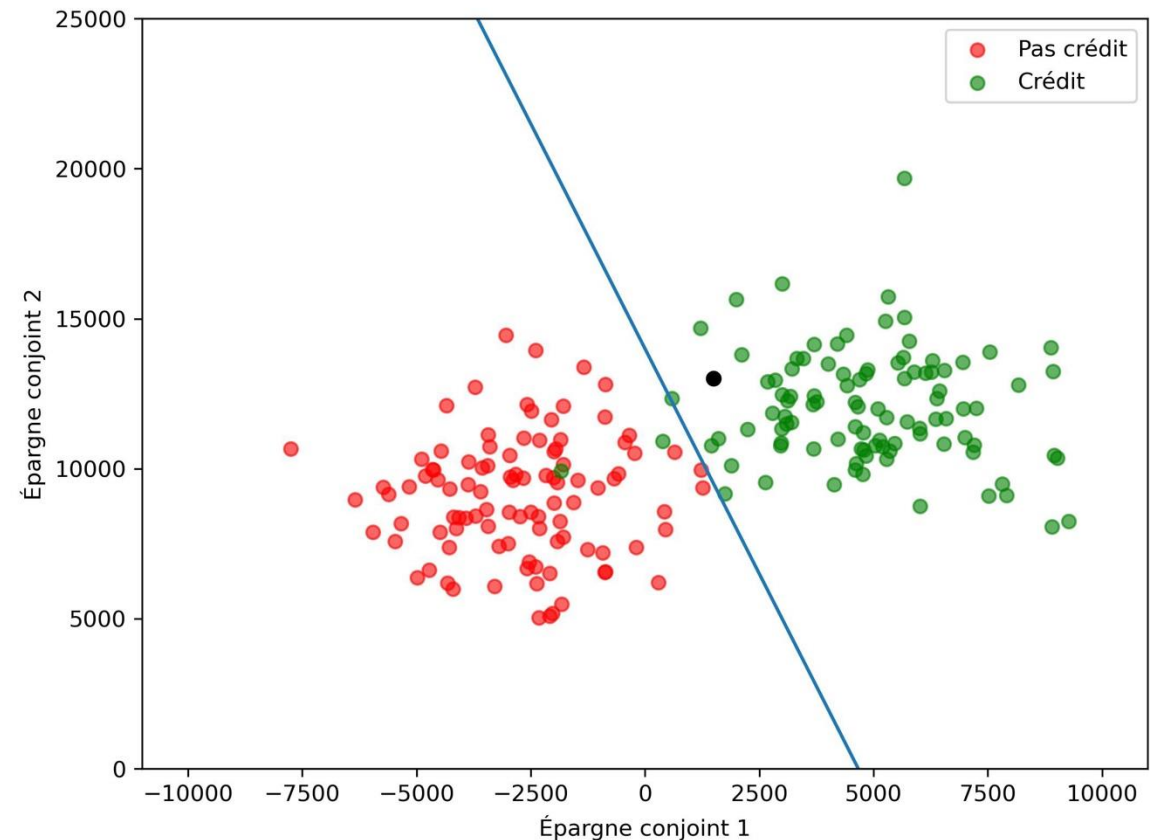


Classification supervisée

Autres exemples

Il n'y a pas d'algorithme magique !

- dépend du type de données en entrée.
- Si on suppose que l'espace des données est séparable par un hyperplan:
 - SVM linéaire



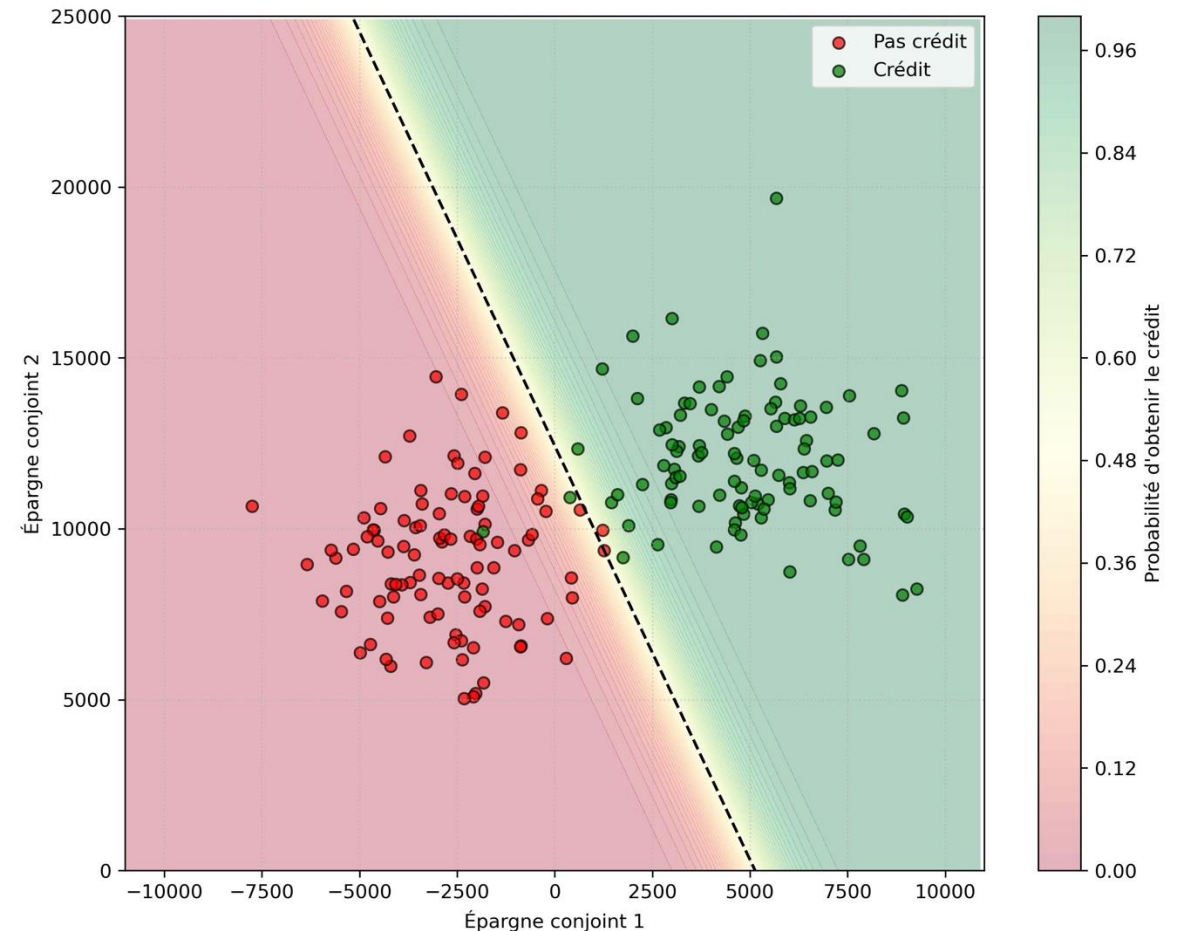
Classification supervisée

Autres exemples

Il n'y a pas d'algorithme magique !

- dépend du type de données en entrée.
- Si on suppose que l'espace des données est séparable par un hyperplan:
 - SVM linéaire
 - Régression logistique

$$\mathbb{P}(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

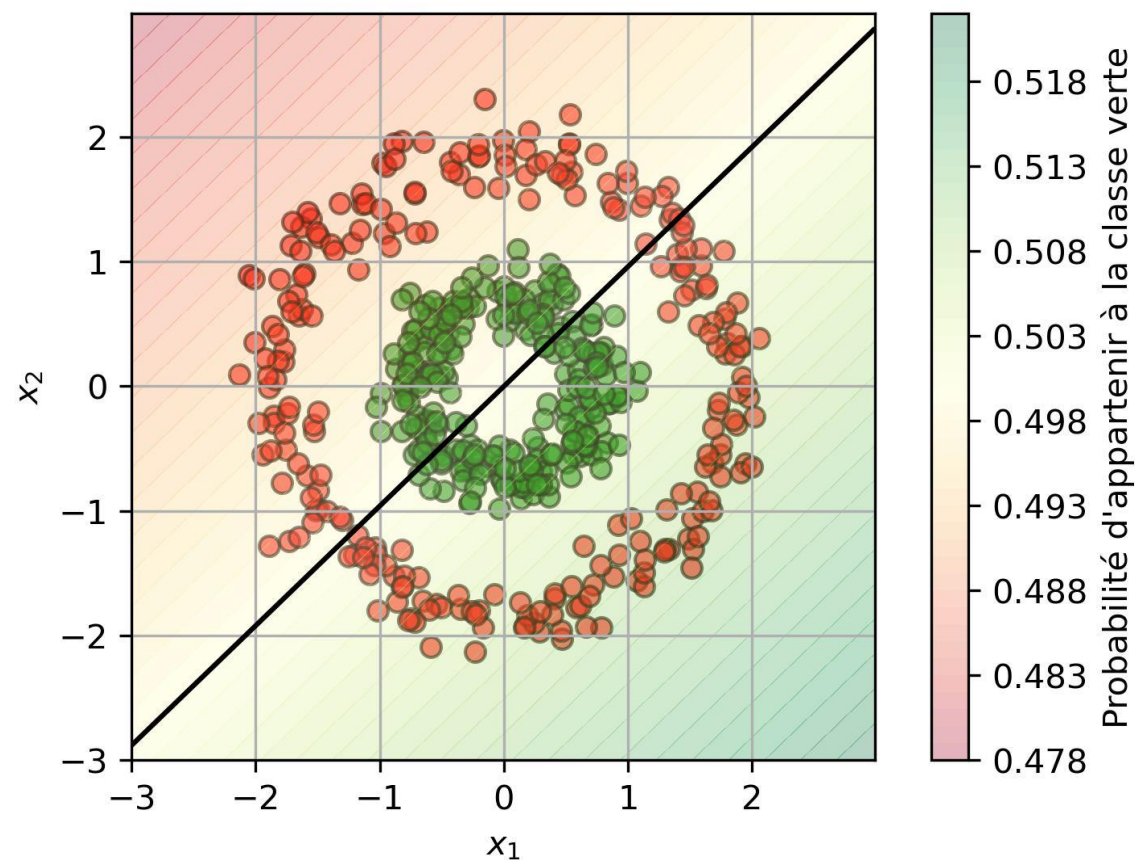


Classification supervisée

Autres exemples

Il n'y a pas d'algorithme magique !

- dépend du type de données en entrée.
- Si on suppose que l'espace des données est **séparable par un hyperplan**:
 - SVM linéaire
 - Régression logistique
- Si on suppose que l'espace des données est **non-linéairement séparable**:

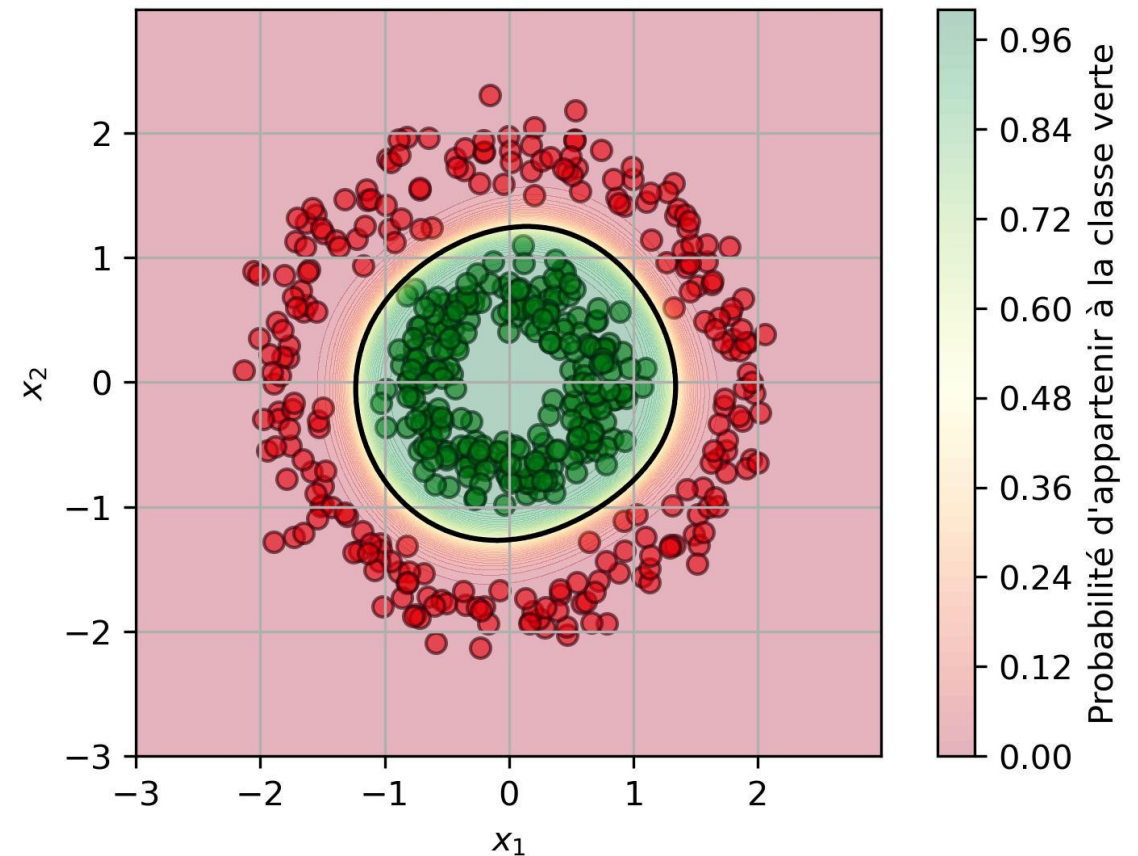


Classification supervisée

Autres exemples

Il n'y a pas d'algorithme **magique** !

- dépend du type de données en entrée.
- Si on suppose que l'espace des données est **séparable par un hyperplan**:
 - SVM linéaire
 - Régression logistique
- Si on suppose que l'espace des données est **non-linéairement séparable**:
 - Perceptron multi-couches
 - SVM non-linéaire





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Classification supervisée

Autres exemples

Il n'y a pas d'algorithme magique !

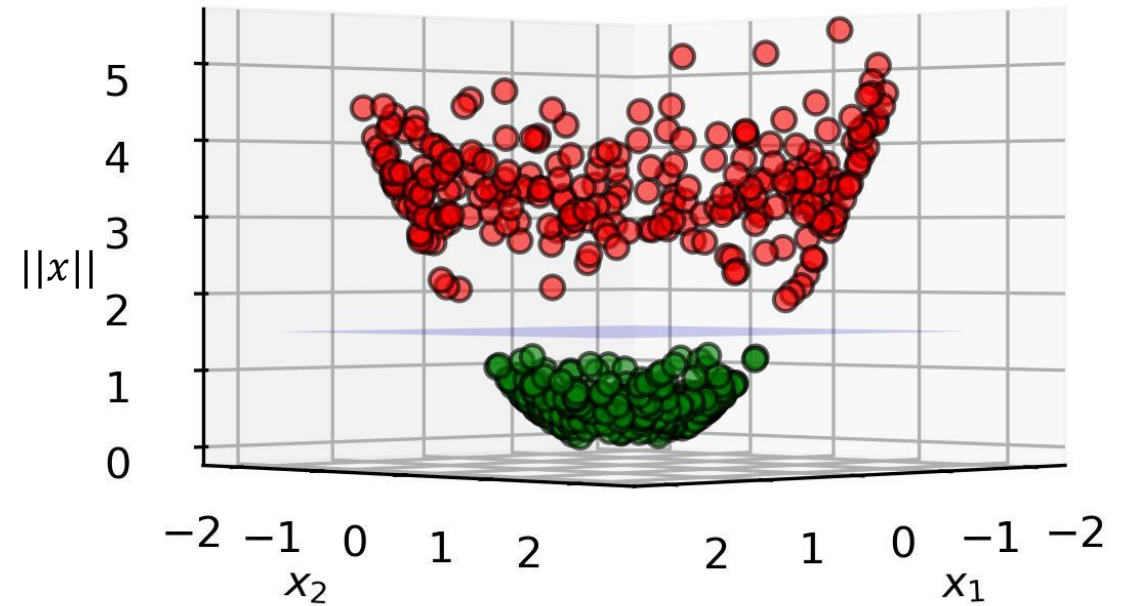
- dépend du type de données en entrée.
- Si on suppose que l'espace des données est séparable par un hyperplan:
 - SVM linéaire
 - Régression logistique
- Si on suppose que l'espace des données est non-linéairement séparable:
 - Perceptron multi-couches
 - SVM non-linéaire
- fonctionnent avec des données quantitatives.

Classification supervisée

Autres exemples

Il n'y a pas d'algorithme **magique** !

- dépend du type de données en entrée.
- Si on suppose que l'espace des données est **séparable par un hyperplan**:
 - SVM linéaire
 - Régression logistique
- Si on suppose que l'espace des données est **non-linéairement séparable**:
 - Perceptron multi-couches
 - SVM non-linéaire
- fonctionnent avec des **données quantitatives**.



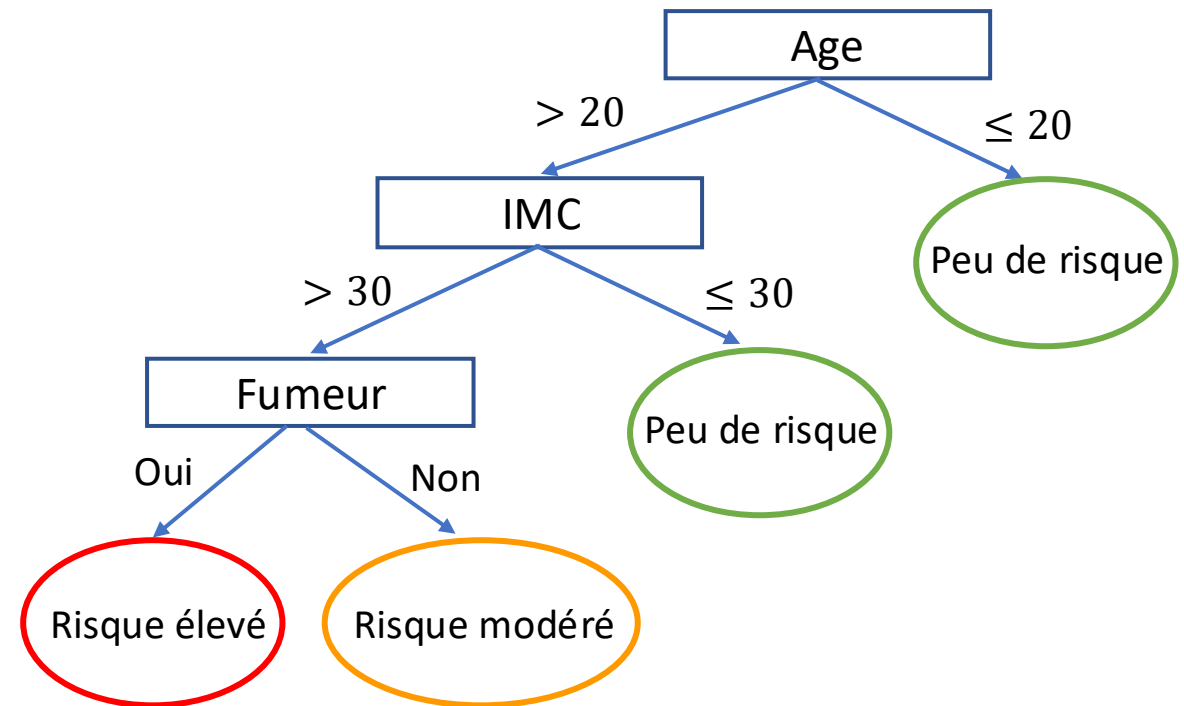
Feature engineering

Classification supervisée

Arbres de décision

- Introduit récemment: années 2000
- Modèles faciles à comprendre et facilement **interprétables**.
- **Ensemble de règles de classification** basant leur décision sur les tests des attributs, organisés de manière arborescente

Prédiction du risque de pathologie cardiaque



Classification supervisée

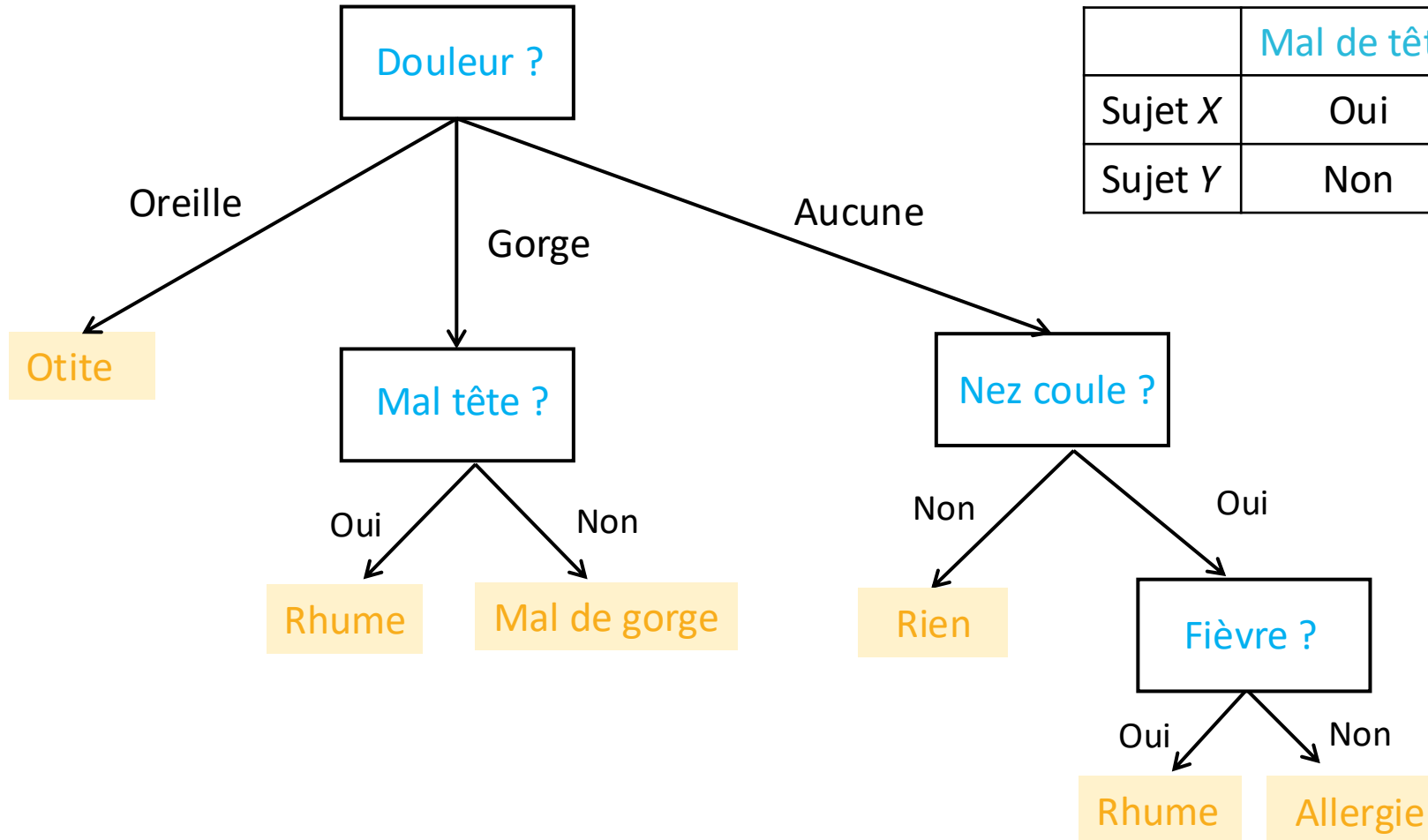
Arbres de décision

- Chaque observation est décrite par un vecteur d'attributs.
- Traitement de **données qualitatives** : non numériques.

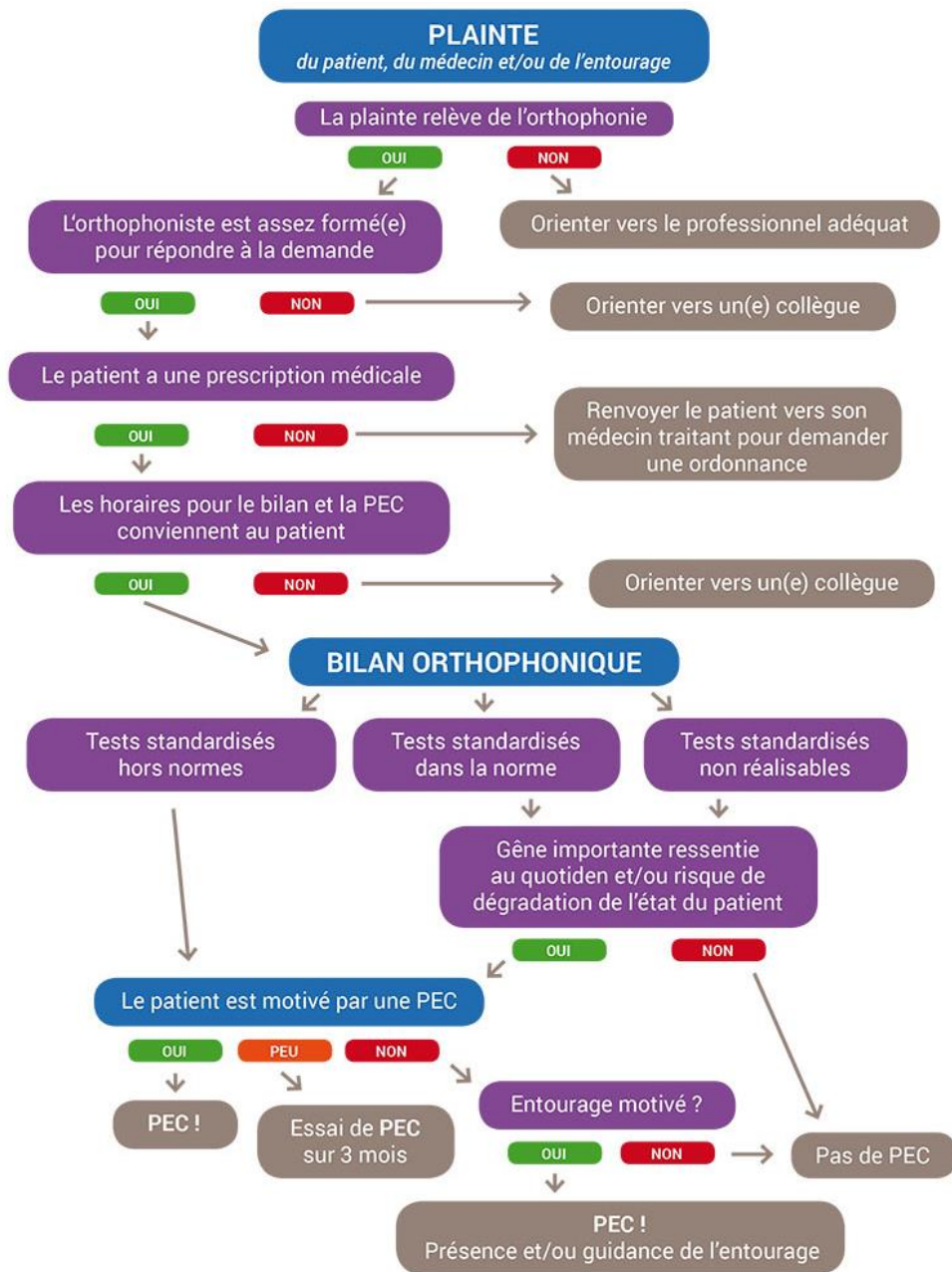
	Mal de tête	Nez coule	Age	Douleur	Diagnostic
Sujet X	Oui	Oui	35	Gorge	Rhume
Sujet Y	Non	Oui	21	Oreille	Otite

Arbres de décision

Classification par série de tests



	Mal de tête	Nez coule	Age	Douleur	Diagnostic
Sujet X	Oui	Oui	35	Gorge	Rhume
Sujet Y	Non	Oui	21	Oreille	Otite



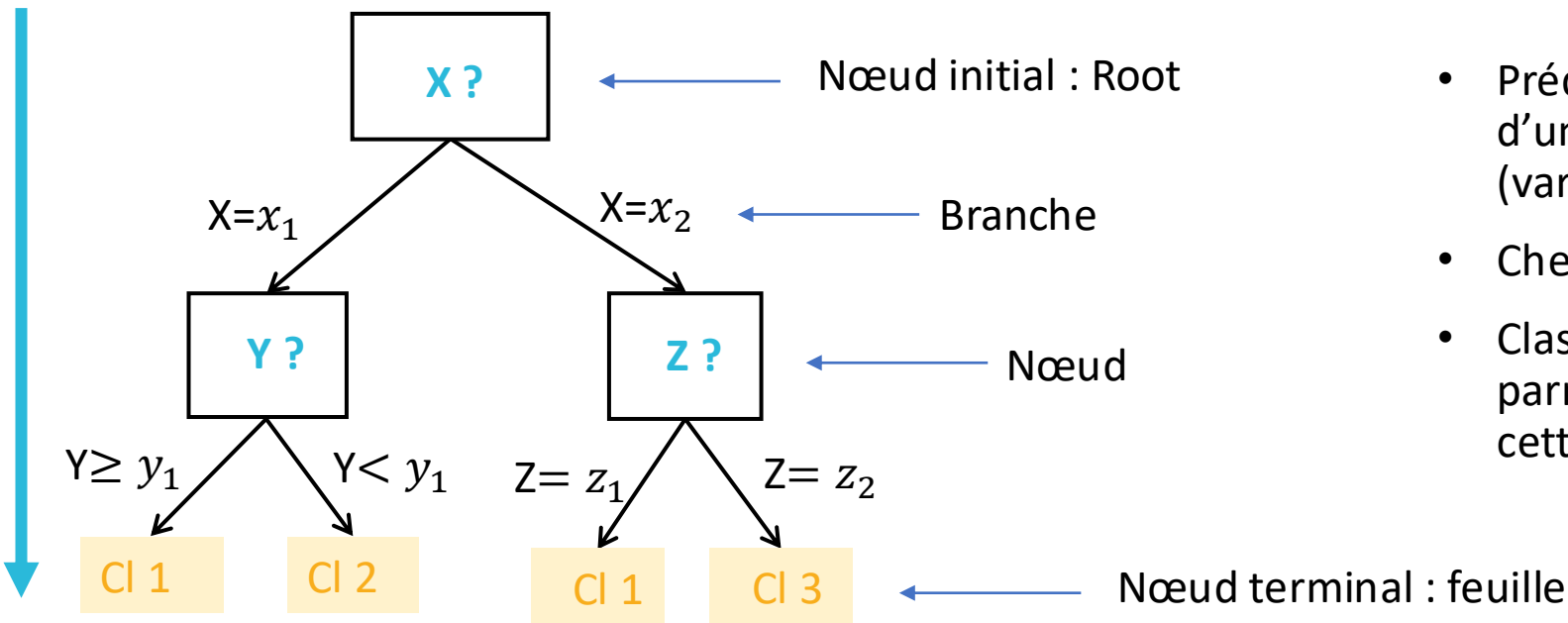
Arbres de décision

En orthophonie

Arbres de décision

Résumé

- Une méthode de **classification supervisée**, sous forme symbolique d'arbre :
 - **Nœuds** : testent les attributs
 - **Branches** : une branche pour chaque valeur possible de l'attribut testé
 - **Feuilles** : catégories



- Prédire la valeur de la cible à partir d'un ensemble de valeurs d'attributs (variables prédictives)
- Chemin de prédiction
- Classe d'une feuille: classe majoritaire parmi les exemples appartenant à cette feuille

Classification supervisée: arriver autant que possible à des **feuilles pures** avec des **branches les plus courtes** et **les moins nombreuses possibles**

Arbres de décision

Un peu de théorie: l'algorithme CART

- **Données:** Soient $X = \{X_1, \dots, X_N\}$ avec $X_i \in \mathbb{R}^p, \forall i \in \llbracket 1, N \rrbracket$ et $Y = [y_1, \dots, y_N]^T \in \mathbb{R}^N$. Ainsi, chaque individu i est représenté par le couple (X_i, y_i) .
- **Objectif:** Créer récursivement un arbre de décision en partitionnant les individus de telle sorte que les individus avec le même *label* soient regroupés ensemble.
- **Méthode:** Supposons que les données au nœud m soient représentées par Q_m avec $|Q_m| = n_m$
- Pour chaque test candidat $\theta = (k, t_m)$ consistant en la variable $k \in \llbracket 1, p \rrbracket$ et le seuil t_m . On partitionne l'ensemble Q_m en deux sous-ensembles $Q_m^G(\theta)$ et $Q_m^D(\theta)$ tels que:

$$Q_m^G(\theta) = \{(X_i, y_i) \in Q_m \mid x_{i,j} \leq t_m\}$$

$$Q_m^D(\theta) = Q_m \setminus Q_m^G(\theta)$$

- La qualité des tests θ à chaque nœud m est donnée à l'aide d'une fonction de coût (régression) ou d'impureté (classification), notée H .

Arbres de décision

Un peu de théorie: l'algorithme CART

- L'impureté G est définie par:

$$G(Q_m, \theta) = \frac{n_m^G}{n_m} H(Q_m^G(\theta)) + \frac{n_m^D}{n_m} H(Q_m^D(\theta)) \text{ où } n_m^G = |Q_m^G(\theta)| \text{ et } n_m^D = |Q_m^D(\theta)|$$

- On sélectionne le test θ^* qui minimise cette impureté:

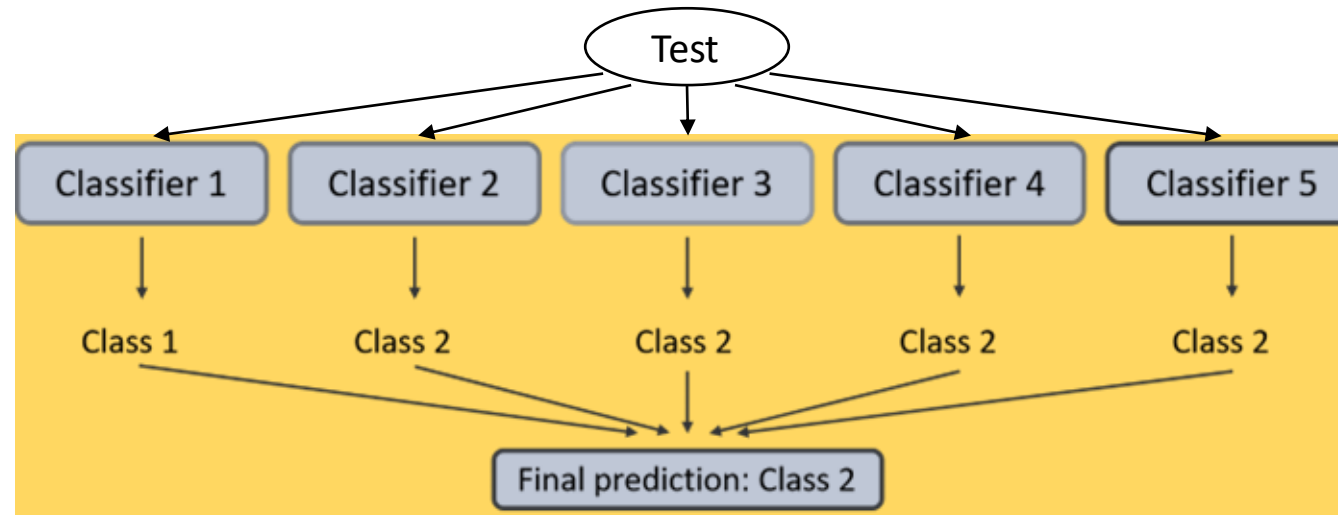
$$\theta^* = \underset{\theta}{\operatorname{argmin}} G(Q_m, \theta)$$

- On recommence alors cette procédure avec les deux sous-ensembles $Q_m^G(\theta^*)$ et $Q_m^D(\theta^*)$ jusqu'à:
 - $n_m = 1$
 - ou $n_m < \min_{\text{samples}}$
 - ou $m = \max_{\text{depth}}$
 - ou que tous les individus en ce nœud soient de la même classe : **feuille pure**

Ensemble learning

L'union fait la force

- Différents classifieurs avec un taux de performance moyen
- Agréger les prédictions de chacun des classifieurs permet une meilleure classification
- Même si chacun des classifieurs est mauvais, l'ensemble peut être meilleur





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Ensemble learning

Sagesse de la foule

- La foule d'individus a plus souvent raison qu'un expert seul, à condition que la foule soit :

Ensemble learning

Sagesse de la foule

- La foule d'individus a plus souvent raison qu'un expert seul, à condition que la foule soit :

Grande



Compétente



Diversifiée



Ensemble learning

Sagesse de la foule

- La foule d'individus a plus souvent raison qu'un expert seul, à condition que la foule soit :

Grande



Compétente



Diversifiée



Trois stratégies

- Bagging (e.g., Random Forest)
- Boosting (e.g., AdaBoost)
- Stacking

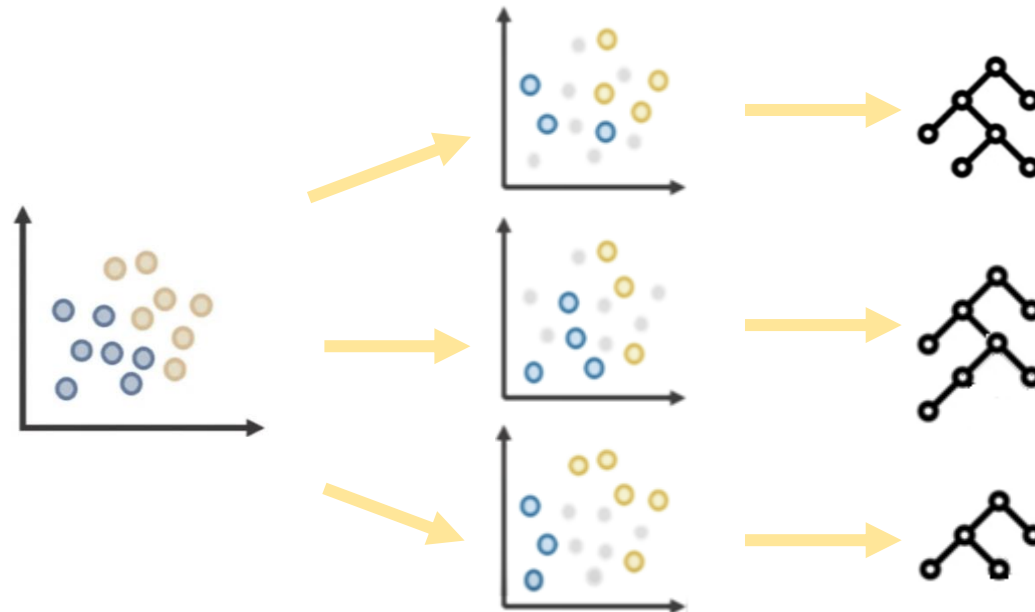
Ensemble learning

Bagging

- Y variable cible et $\forall i \in \llbracket 1, N \rrbracket$, $x_i \in \mathbb{R}^p$, N observations
- $f(x)$ modèle appris sur un échantillon $S = ((x_1, y_1), \dots, (x_N, y_N))$

Bagging

Créer **plusieurs répliques** de $f(x)$, en entraînant chaque copie sur une partie **aléatoire** du dataset



L. Breiman

- Bagging predictors, Machine Learning, 24(2), 1996
- Random forests, Machine Learning, 45, 2001

Ensemble learning

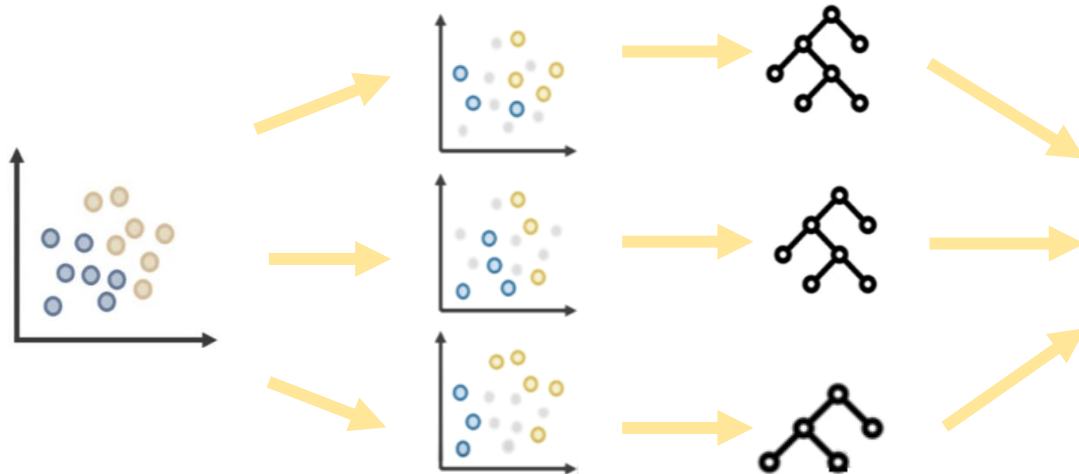
Bagging

- M échantillons bootstrap S_1, \dots, S_M issus de S par tirage aléatoire avec remise
- Sur chaque échantillon k , apprendre un modèle $f_k(x)$
- Prévission de X_{test} par agrégation de modèles

➤ Régression: $\widehat{f}_M(X_{test}) = \frac{1}{M} \sum_{k=1}^M \widehat{f}_{S_k}(X_{test})$

➤ Classification: $\widehat{f}_M(X_{test}) = \underset{j}{\operatorname{argmax}} \operatorname{Card} \{k | \widehat{f}_{S_k}(X_{test}) = j\}$

➔ Vote majoritaire ($f_1(x), \dots, f_M(x)$)



Prévission par agrégation des modèles

Moyenner les prévisions de plusieurs modèles permet de réduire l'erreur de prédiction

Forêts aléatoires

- Tirer au hasard dans la base d'entraînement M échantillons avec remise (échantillons de taille n)
 - Construire un arbre pour chaque échantillon k avec tirage aléatoire des attributs
 - Pour la construction de chaque nœud de chaque arbre, tirer au hasard q attributs parmi p pour émettre la décision associée au nœud
 - Classification par vote majoritaire
- Maximiser la diversité des arbres
 - Appris sur différents attributs
 - Construits sur des échantillons différents

Forêts aléatoires

- Tirer au hasard dans la base d'entraînement M échantillons avec remise (échantillons de taille n)
 - Construire un arbre pour chaque échantillon k avec tirage aléatoire des attributs
 - Classification par vote majoritaire
- Maximiser la diversité des arbres
 - Appris sur différents attributs
 - Construits sur des échantillons différents
- Critiques ?
 - Peu de sur-apprentissage
 - Meilleure performance que les arbres de décision
 - Parallélisation possible

 - Difficilement interprétable
 - Entraînement plus lent



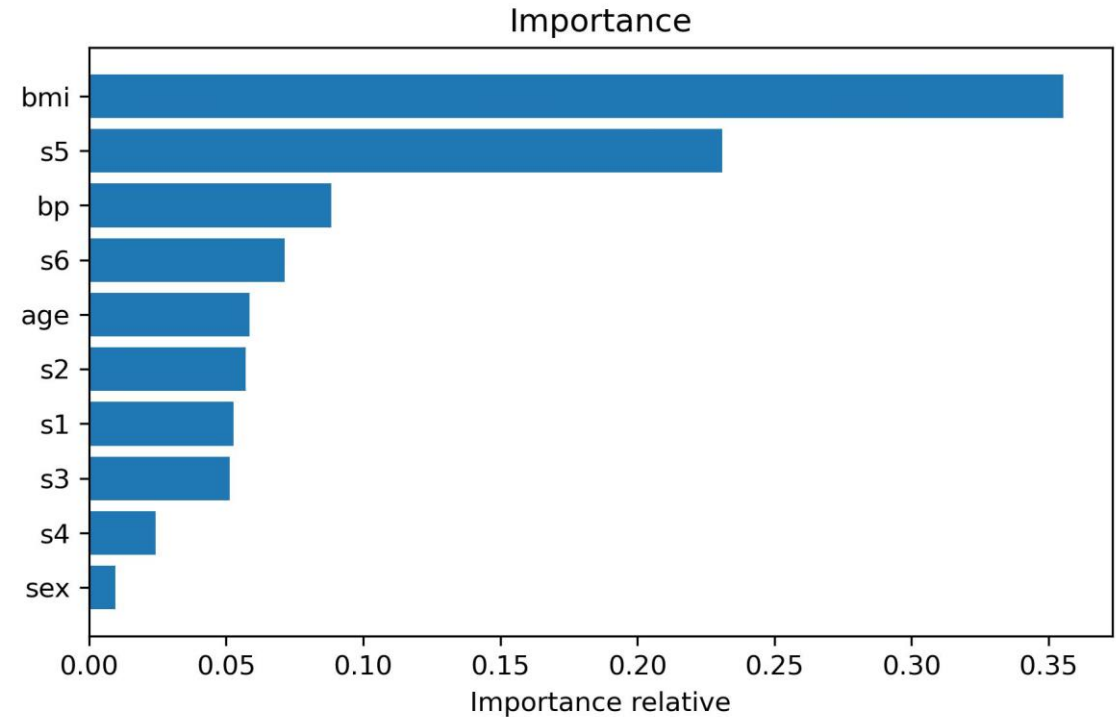
IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Forêts aléatoires

- Permet également de savoir quels sont les attributs les plus importants pour prédire la variable cible = **feature importance**
- Pour prédire le diabète:
 - **bmi**: IMC
 - **s5**: niveau de triglycérides
 - **bp**: pression sanguine



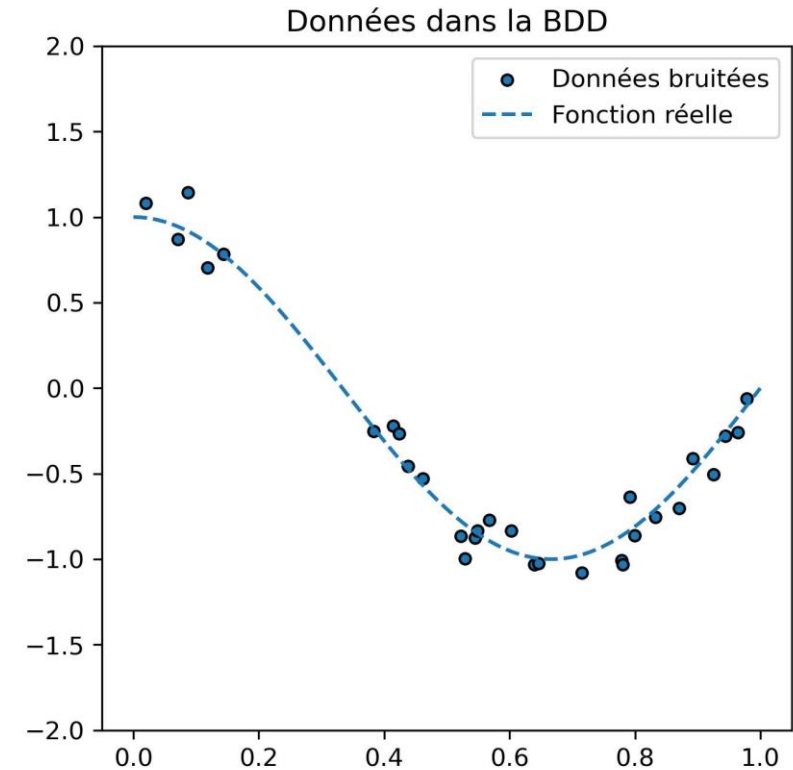


IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Apprentissage supervisé

- Tout modèle est une fonction qui dépend de paramètres.
- Apprendre = **ajuster** les paramètres du modèle
- Le choix d'un modèle (donc du nombre de paramètres) dépend du nombre de données, et surtout, de la **complexité de la distribution** des données.



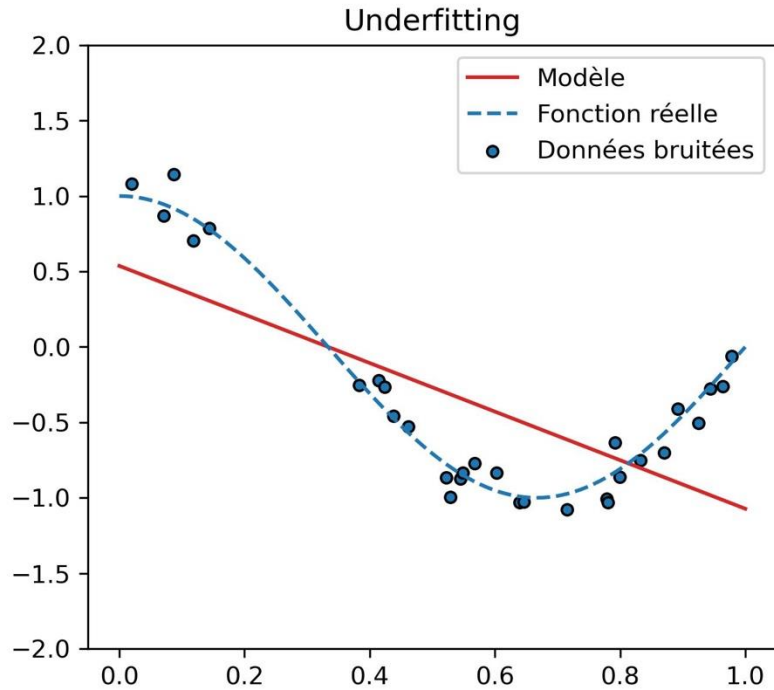


IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Apprentissage supervisé

Ex: régression polynomiale





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



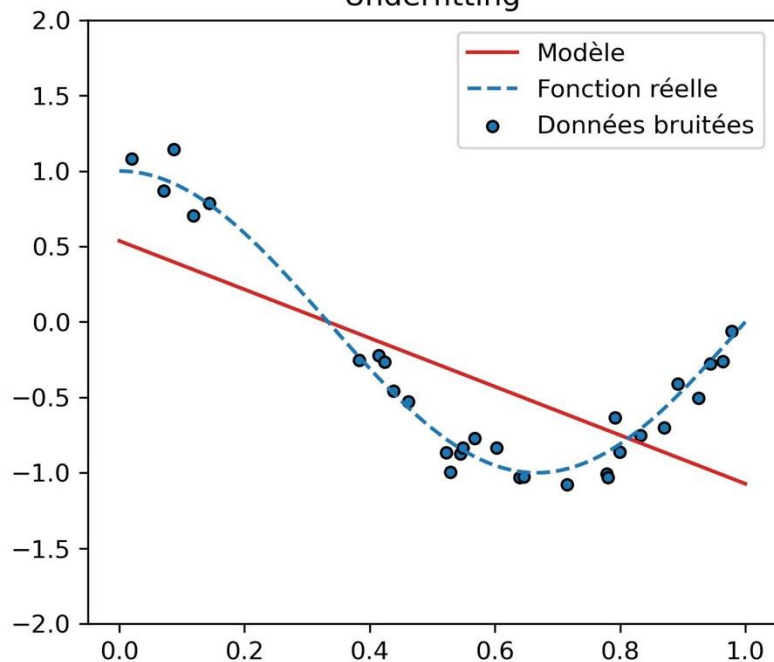
RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

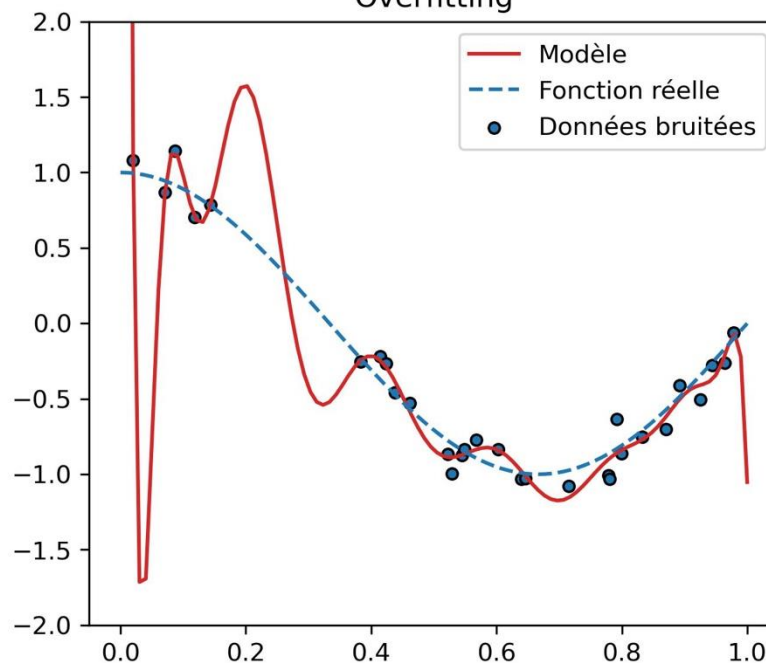
Apprentissage supervisé

Ex: régression polynomiale

Underfitting



Overfitting





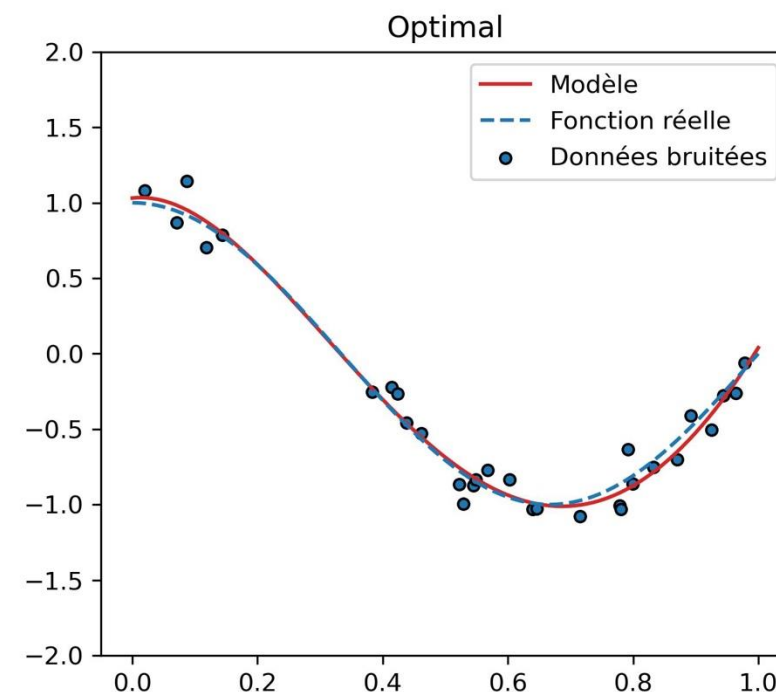
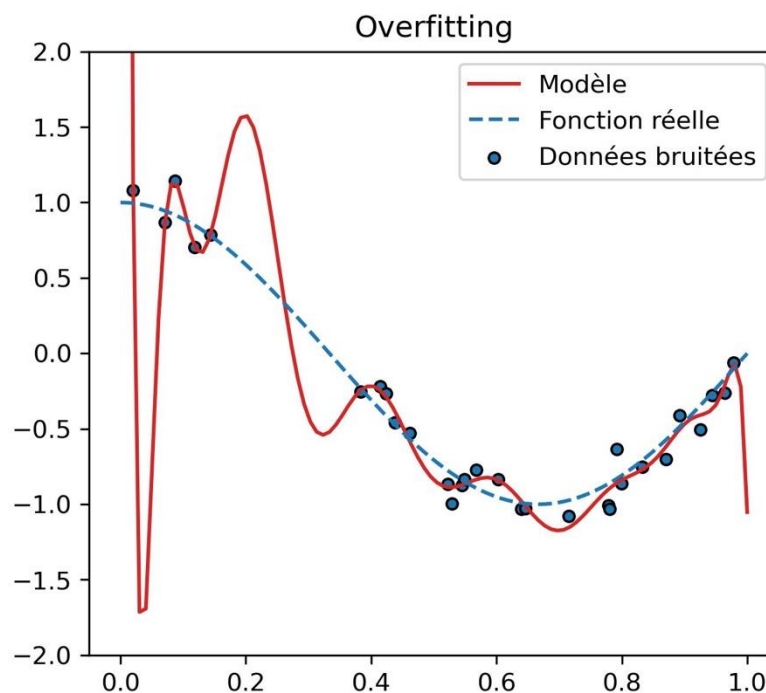
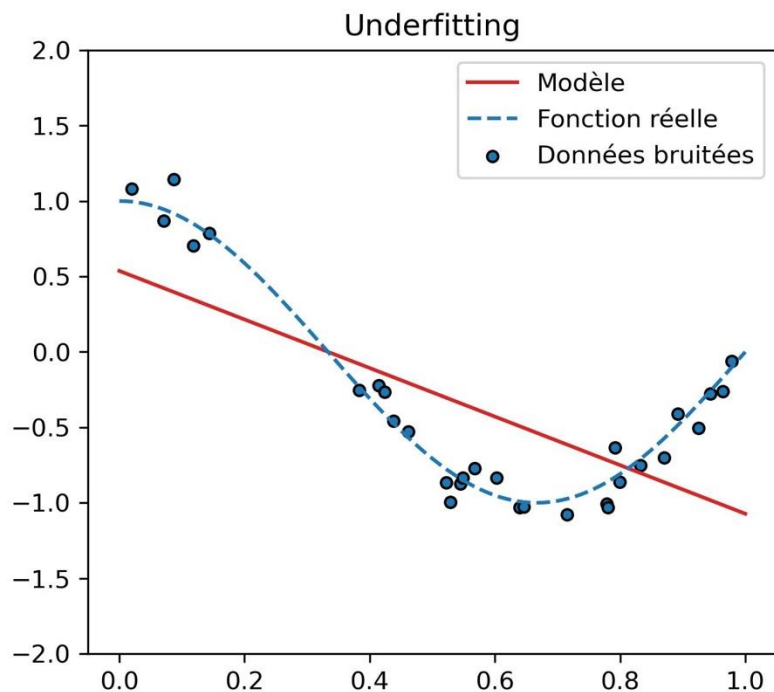
IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

Apprentissage supervisé

Ex: régression polynomiale





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Exemple de projets made in Lille

#IMTomorrow
#IMTNordEurope

Prédiction de la courbe de poids post-chirurgie bariatrique



- Collaboration entre Inserm U1190 (CHU de Lille) et Scool (Inria/CRISTAL/UdL)
- Comment prédire la courbe de poids à partir des données de la cohorte ABOS:
 - 1500 patients
 - Suivi longitudinal pendant +5 ans
 - Données recueillies lors de visites à t0 puis 1, 3, 12, 24, 60 mois
- Perte de poids représentée en % (TWL)
- Beaucoup d'attributs pour chaque patient (434) dont des données anthropométriques et biologiques.

Prédiction de la courbe de poids post-chirurgie bariatrique

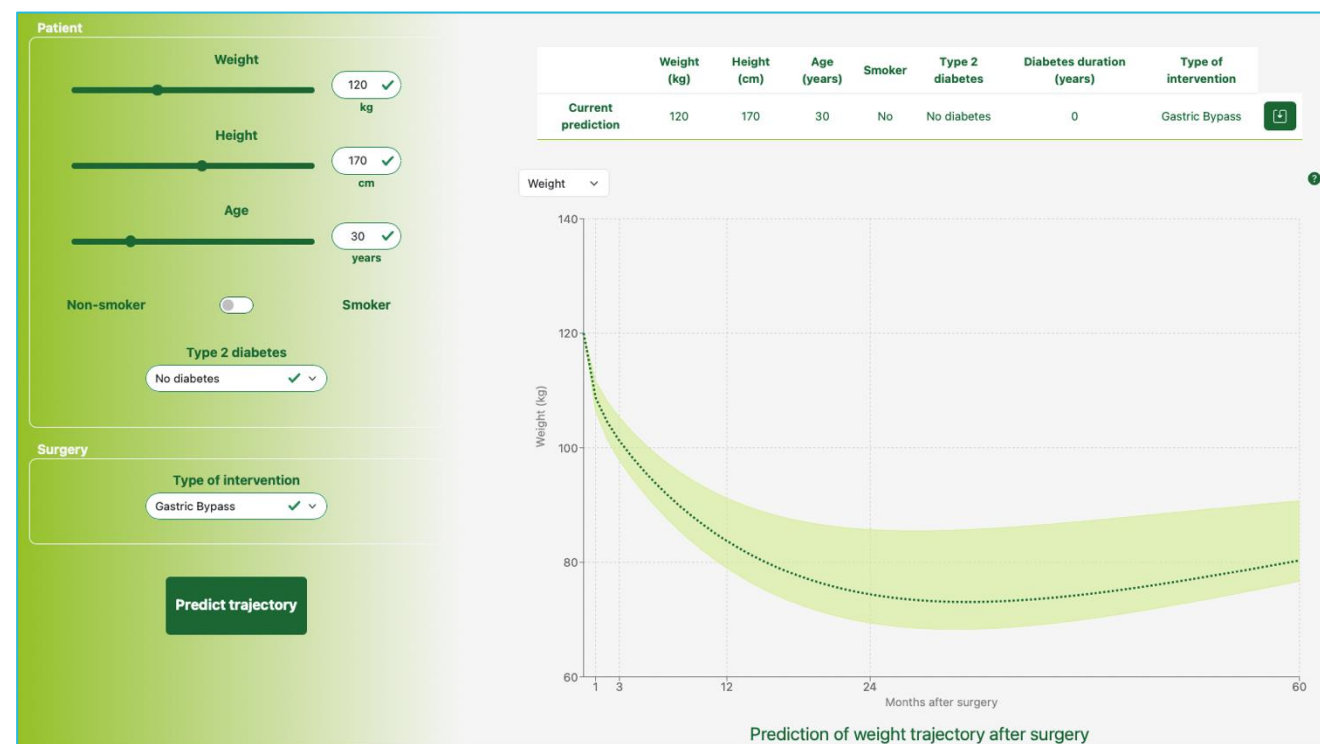
- Modèle utilisé: **arbre de décision** car fonctionne bien dans le **contexte médical** (peu de données avec des attributs de plusieurs types) et permet l'**interprétation** des résultats.
- Un arbre de décision par visite puis **interpolation** dans le but de prédire une courbe
- Pourquoi ?: Permettre aux patients de se projeter dans leur perte de poids et aux cliniciens de **décider** de l'intervention la plus adéquate.
- Validation: Il faut choisir des attributs **simples** à recueillir pour tester le modèle sur des bases de données externes.

Prédiction de la courbe de poids post-chirurgie bariatrique

- Sélection de variables: 7 attributs sélectionnés automatiquement à l'aide des arbres de décision. Une sélection par une approche type LASSO a choisi les mêmes attributs (moins un) utilisés par la suite.

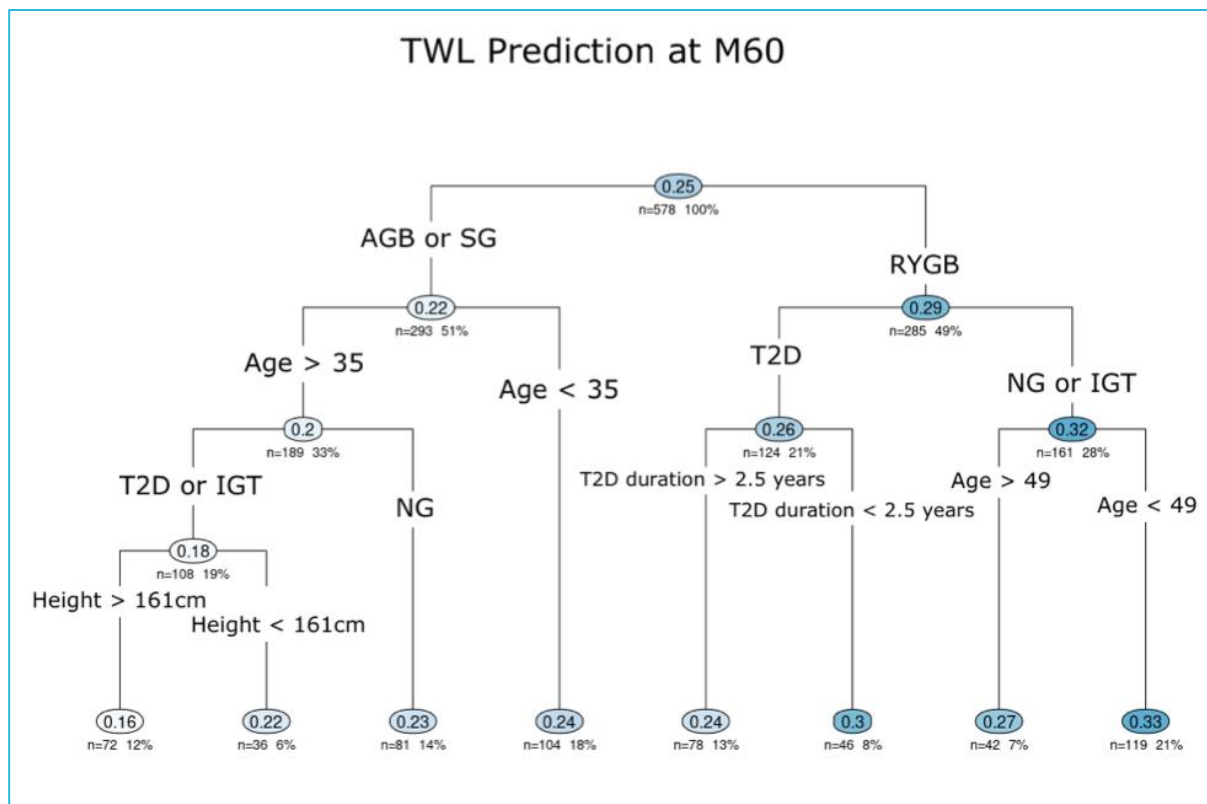
- ❖ Age
- ❖ Type d'intervention
- ❖ Poids initial
- ❖ Fumeur
- ❖ Type de diabète
- ❖ Durée du diabète

- Interface utilisée plus d'1M de fois !





Prédiction de la courbe de poids post-chirurgie bariatrique





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Apprentissage non-supervisé

#IMTomorrow
#IMTNordEurope

L'apprentissage non supervisé vise à **découvrir des structures cachées** dans des données **non étiquetées**.

Objectif: Identifier des motifs ou des groupes au sein des données

L'agent apprend, sans professeur, à partir **des relations** entre les observations

Cas d'usage:

- Exploration de données : outliers, prétraitement
- Segmentation de clients ou d'images
- Analyse de séquence génomique
- Profilage clinique
- Réduire la dimensionnalité des données

Pourquoi apprendre à classer des données non étiquetées ?



Pourquoi ?

- Coût / impossibilité d'étiqueter
- Mauvaise connaissance des données

Comment ?

- Etude de la structure inhérente aux données
- Pour palier le manque de connaissances sur les classes, on fait **des hypothèses fortes** sur l'espace de décision (métrique dans l'espace, forme des classes,...)

Hard clustering

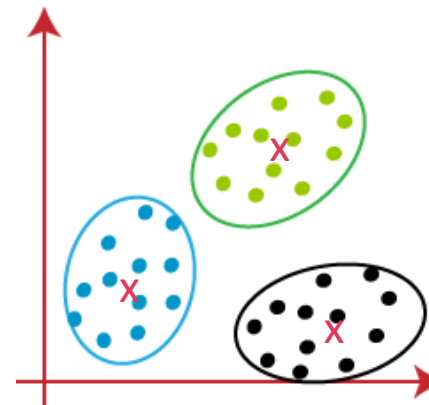
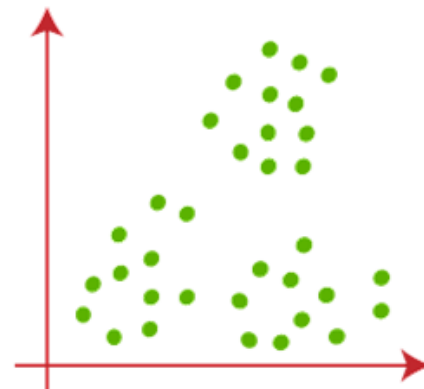
K-Moyennes

➤ Fonctionnement

- Partitionne les données en **K clusters exclusifs**
- Point dans un groupe = déterminé par la distance entre ce point et le **centroïde** du cluster

➤ Utilisation

- Nombre de clusters connu
- Regroupement rapide de grands ensembles de données



$K = 3$



Hard clustering

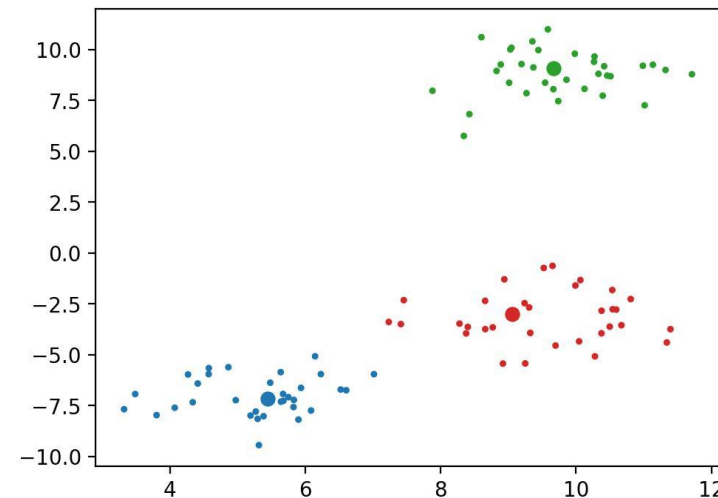
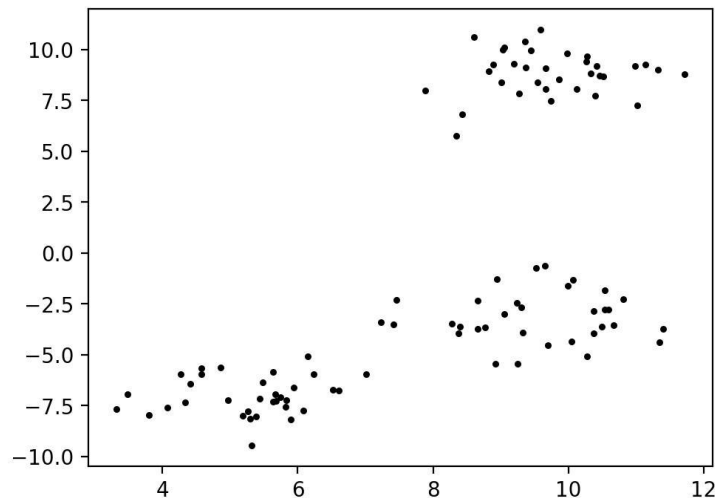
K-Medoides

➤ Fonctionnement

- Similaire au k -moyennes
- **Exigence:** centre de clusters coïncident avec les points dans les données → **Medoïds**

➤ Utilisation

- Nombre de clusters connu
- Regroupement rapide de grands ensembles de données



Hard clustering

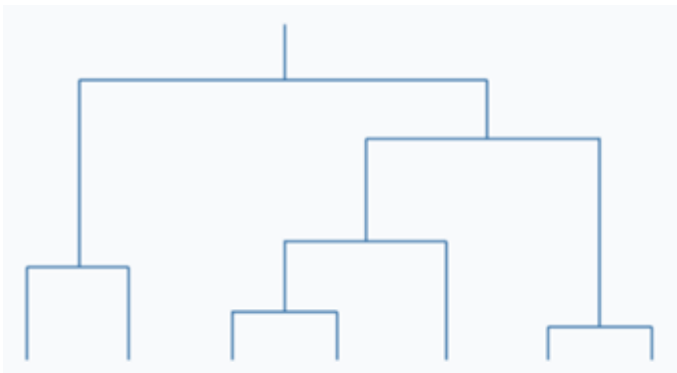
Clustering hiérarchique

➤ Fonctionnement

- Analyse les similitudes entre des paires de points
- Regroupe des objets en un arbre hiérarchique → grappes imbriquées

➤ Utilisation

- Nombre de clusters inconnu
- La visualisation guide la sélection





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille

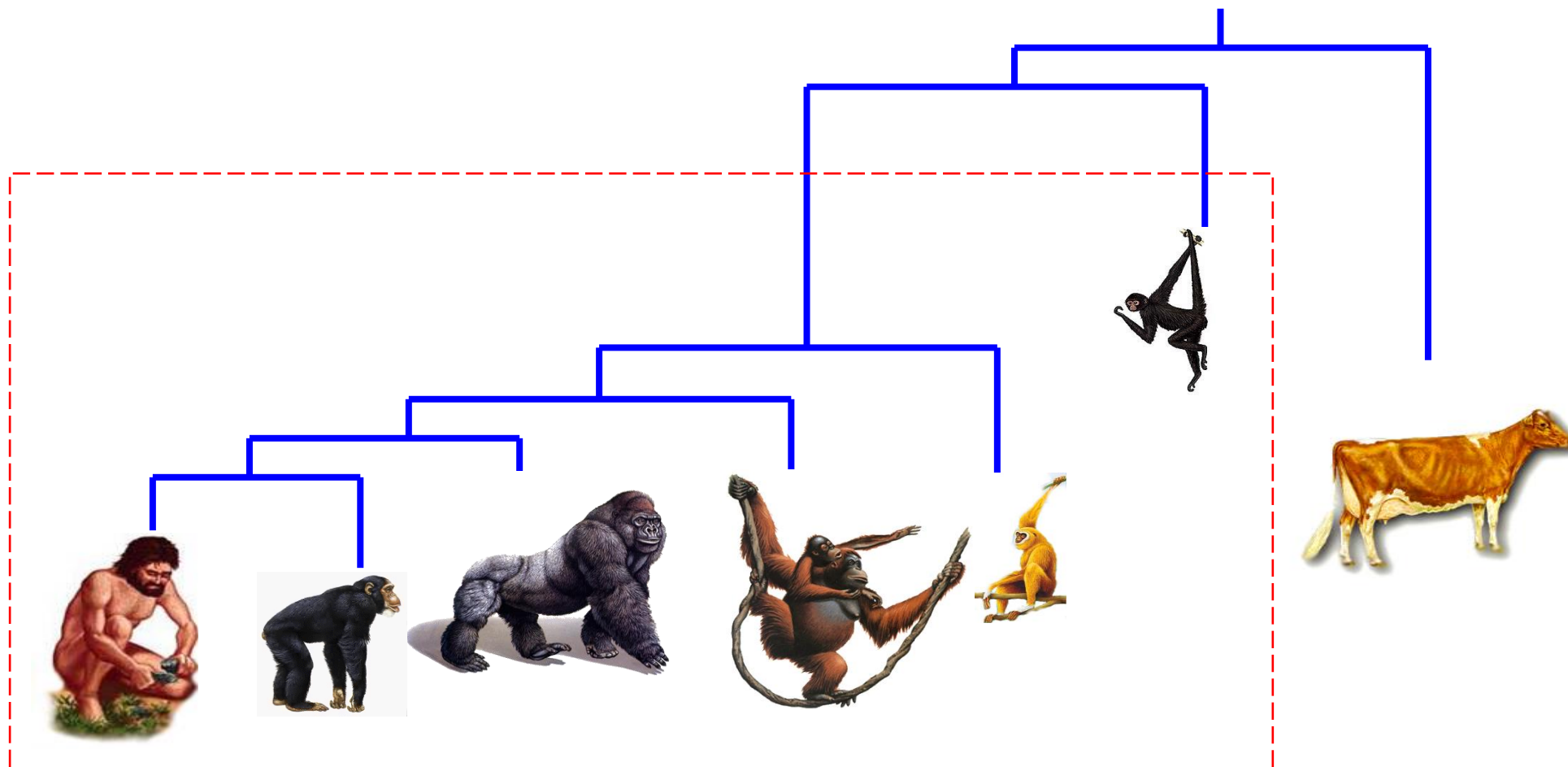


RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

Hard clustering

Clustering hiérarchique



Hard clustering

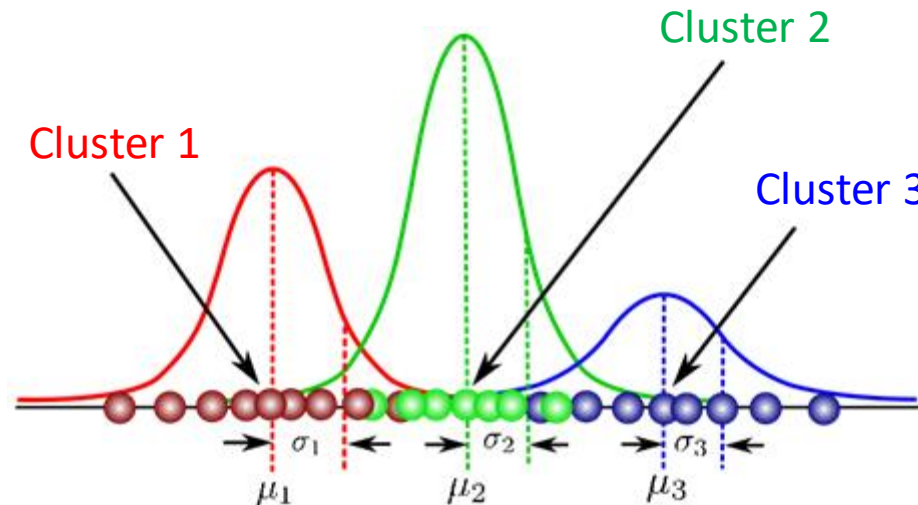
Gaussian Mixture Model

➤ Fonctionnement

- Les données des clusters proviennent de différentes distributions normales multivariées avec certaines probabilités

➤ Utilisation

- Une donnée appartient à plusieurs clusters



Clustering

Regroupement des données

- Similaires au sein d'un même cluster
- Dissimilaires entre différents clusters

Dispersion intra-classe à minimiser

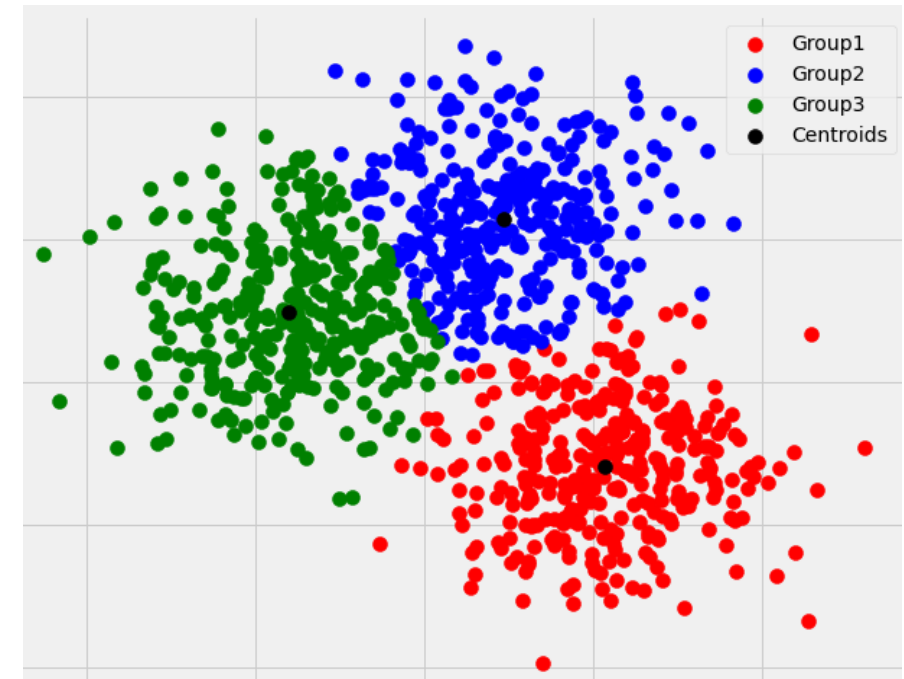
$$WSS = \sum_{i=1}^k \sum_{x \in C_i} d(x, m_i)^2$$

Indicateur de compacité

Dispersion inter-clusters à maximiser

$$BSS = \sum_{i=1}^k n_i * d(m_i, m)^2$$

Indicateur de séparabilité des classes





IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Compression d'images



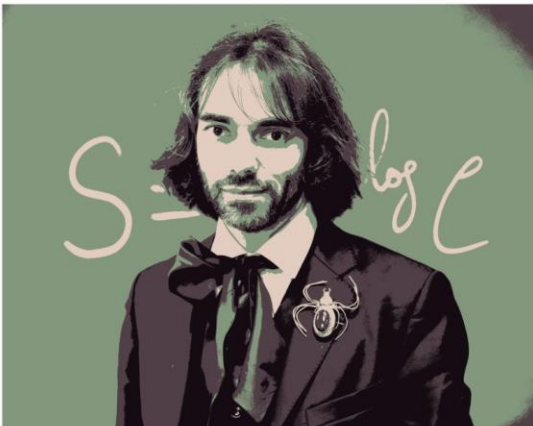
Original



K=2



K=4



K=16

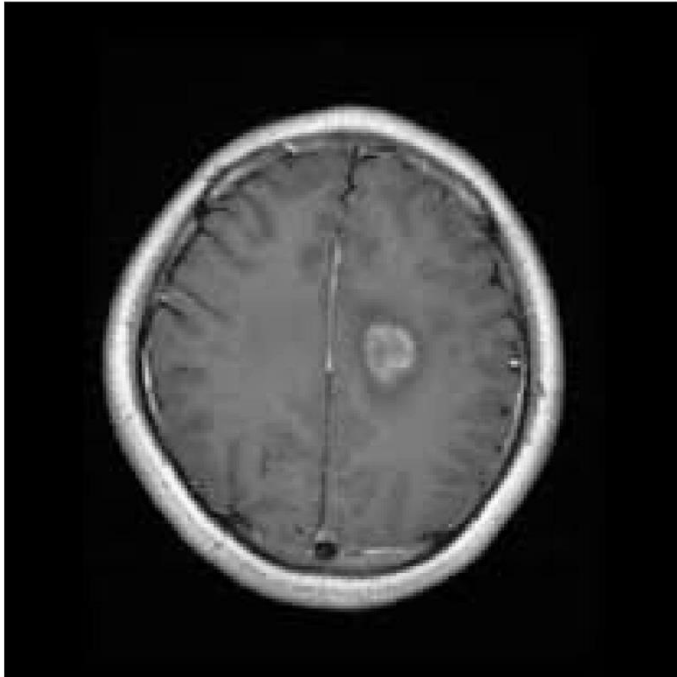


- Chaque pixel caractérisé par un **vecteur de paramètres**
- Ici **p=3**, correspondant à la couleur RGB du pixel

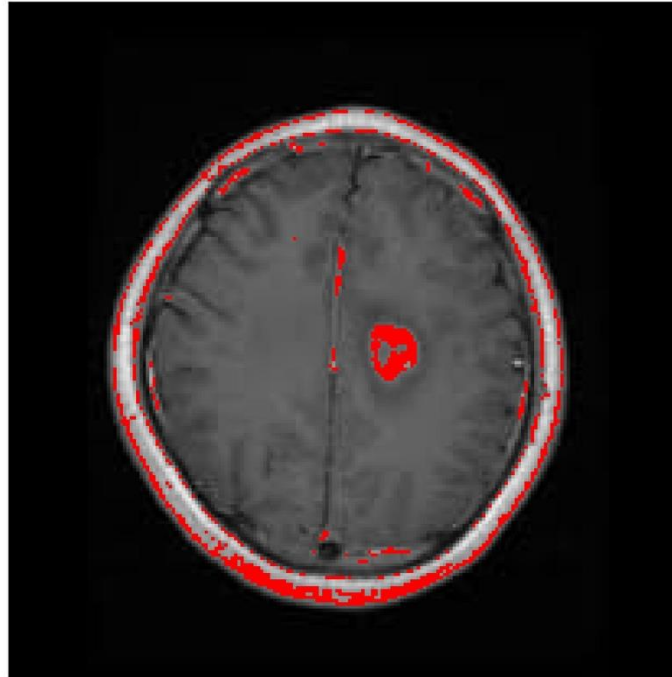
$$X = \begin{bmatrix} & R & G & B \\ x_{11} & x_{12} & x_{13} \\ \vdots & & \\ x_{i1} & x_{i2} & x_{i3} \\ \vdots & & \\ x_{N1} & x_{N2} & x_{N3} \end{bmatrix}$$

Segmentation d'images

Original Image



Segmentation



$$X = \begin{bmatrix} & R & G & B \\ x_{11} & x_{12} & x_{13} \\ \vdots & & \\ x_{i1} & x_{i2} & x_{i3} \\ \vdots & & \\ x_{N1} & x_{N2} & x_{N3} \end{bmatrix}$$

- Tumeur regroupée au sein d'un cluster
- Améliore la visibilité de la tumeur en considérant le cluster comme un masque
- Nécessite une analyse plus fine pour isoler complètement la tumeur

Take-home message

- Beaucoup de modèles d'apprentissage supervisé.
- Sur une étude de cas médicale:
 - visualisation **obligatoire** (!),
 - **feature engineering** souvent nécessaire,
 - essayer les différents modèles et les **confronter**.
- La notion de distance est importante car mesure la **similarité** entre les données.
- Les logiciels d'apprentissage automatique sont tous gratuits, open-sources et collaboratifs = **tout le monde** y a accès !



IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille



Merci pour votre attention

Lorenzo HERMEZ

lorenzo.hermez@imt-nord-europe.fr

IMT Nord Europe, CERI SN, HIDE

#IMTomorrow
#IMTNordEurope

26 mars 2026