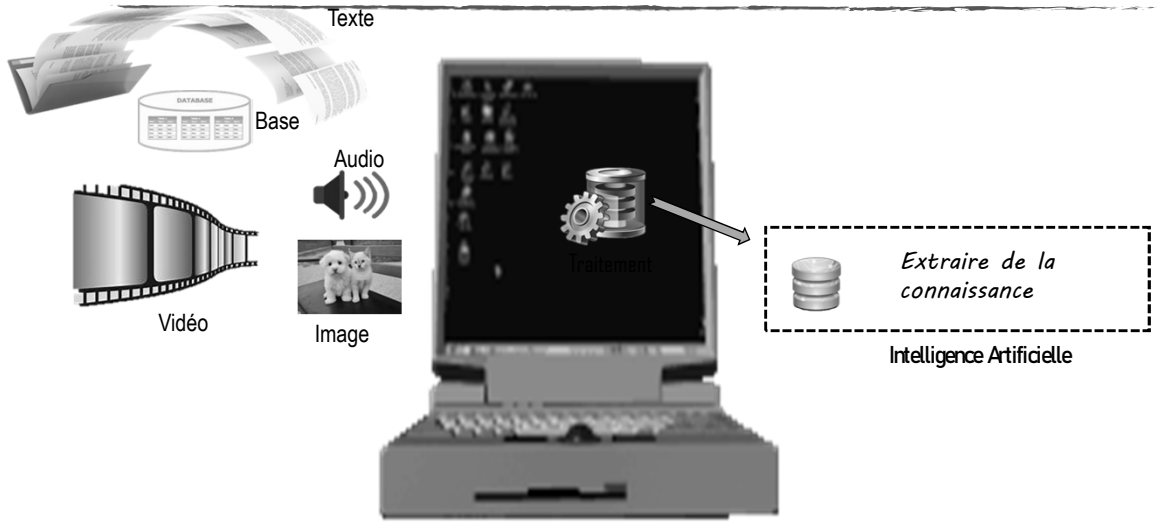
Chapitre 2

Apprentissage Artificiel

Phases de Modélisation

Introduction



Problématique

- On cherche à partir d'un ensemble de données X d'extraire une connaissances Y

Exemple:

En prenant en considération le nombre de contamination par le Covid-19 des journées précédentes; indiquer s'il y'aura une augmentation de cas de contamination dans la wilaya de Blida ?

Problématique

$X=\{x_i\}$

id	Journée	Augmentation
x1	Vendredi	Non
x2	Samedi	Non
x3	Dimanche	Non
x4	Lundi	Oui
x5	Mardi	Non
x6	Mercredi	Non
x7	Jeudi	Oui
x8	Vendredi	Oui
x9	Samedi	Non
x10	Dimanche	Oui
x11	Lundi	Non
x12	Mardi	Non
x13	Mercredi	Oui
x14	Jeudi	Non
x15	Vendredi	Non
x16	Samedi	Oui
x17	Dimanche	Non
x18	Lundi	Oui
x19	Mardi	Oui
x20	Mercredi	Non
x21	Jeudi	Non
x22	Vendredi	Oui
x23	Samedi	?? ⟸Y

Le problème consiste à déterminer une sortie

à partir d'un ensemble de données d'entrée

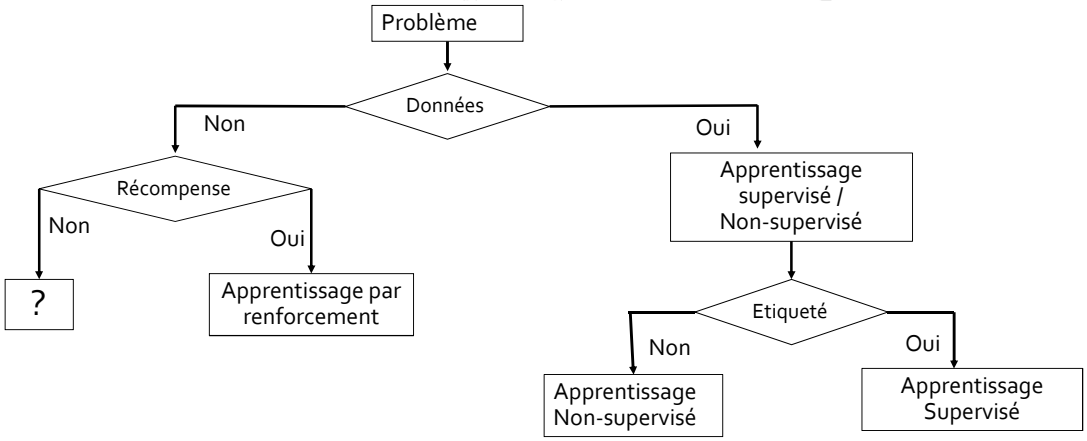
$Y = f(X)$ avec $X=\{x_i\}$

Si la fonction f est connue le problème est simple

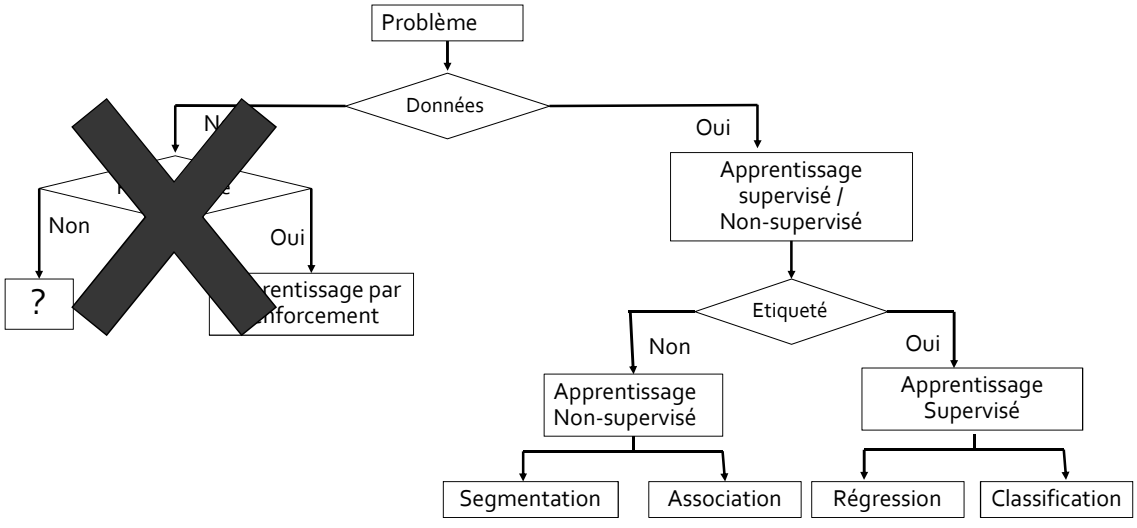
Sinon il faut déterminer la fonction f ou l'approximer

L'approximation de la fonction f en utilisant un modèle est un problème de modélisation.

Modélisation et Apprentissage Artificiel



Modélisation et Apprentissage Artificiel



Apprentissage supervise vs Apprentissage non supervise

Apprentissage supervisé

- Classification
  - Prédire la classe
    - Maligne/Begnine
- Régression
  - Prédire la valeur
    - Quantité du stock

Apprentissage non supervisé

- Association
  - Identifier des relations d'association entre les entités
    - Achat de pain ----> Achat de lait
- Segmentation
  - Réunir les entités similaires en groupe
    - Répartir les clients en groupes

Apprentissage supervise vs Apprentissage non supervise

Apprentissage supervisé

Point	Coordonnées		Classe
	X	Y	
P1	2	4	N
P2	3	3	P
P3	3	4	N
P4	4	2	N
P5	4	4	N
P6	5	1	N
P7	5	2	N
P8	5	3	N
P9	5	6	P
P10	5	8	P
P11	6	1	N
P12	6	5	P
P13	7	4	P
P14	7	6	P
P15	7	7	N
P16	8	3	P
P17	9	4	N
P18	9	6	P

Etiqueté

Apprentissage non supervisé

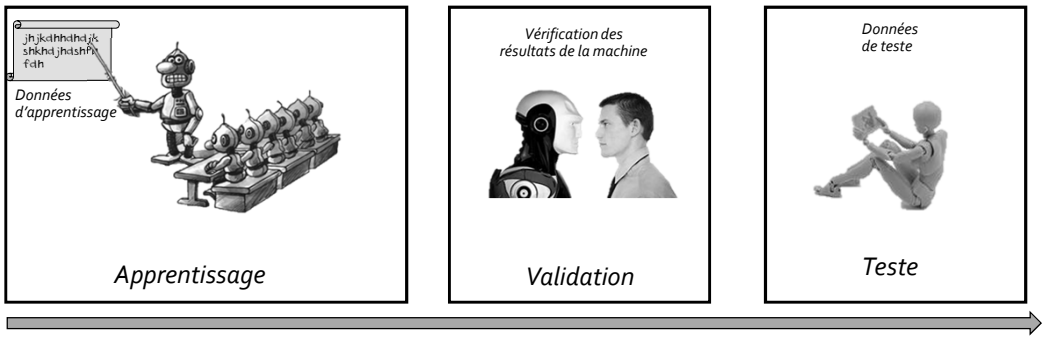
Point	Coordonnées		Classe
	X	Y	
P1	2	4	
P2	3	3	
P3	3	4	
P4	4	2	
P5	4	4	
P6	5	1	
P7	5	2	
P8	5	3	
P9	5	6	
P10	5	8	
P11	6	1	
P12	6	5	
P13	7	4	
P14	7	6	
P15	7	7	
P16	8	3	
P17	9	4	
P18	9	6	

Non étiqueté

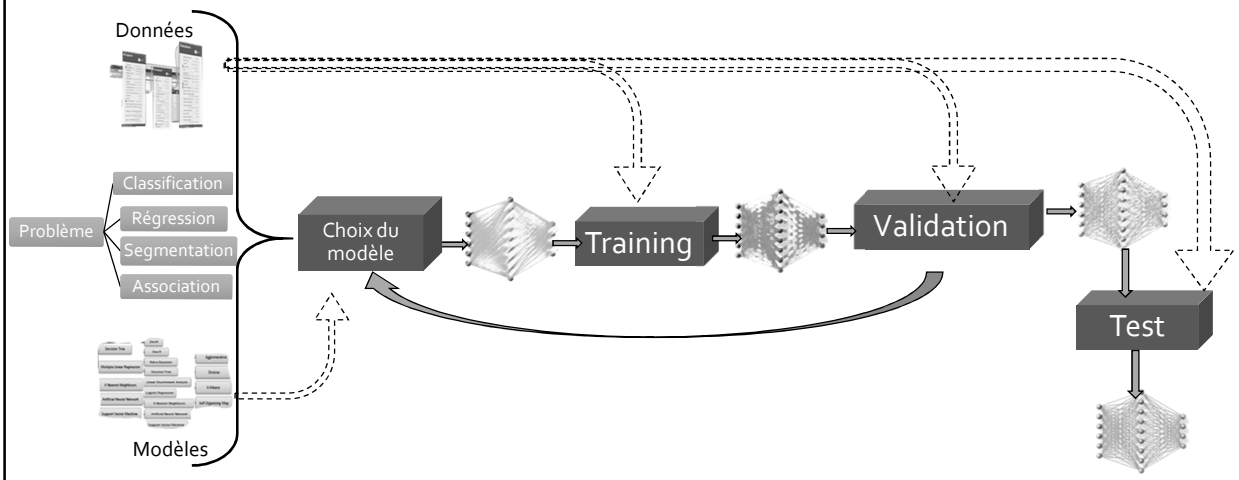
Techniques de modèlisation

- **Naïve Bayesian**
- **Decision Tree**
- **K Nearest Neighbours**
- **Logistic Regression**
- **Artificial Neural Network**
- **Support Vector Machine**
- **K-Means**
- **…**

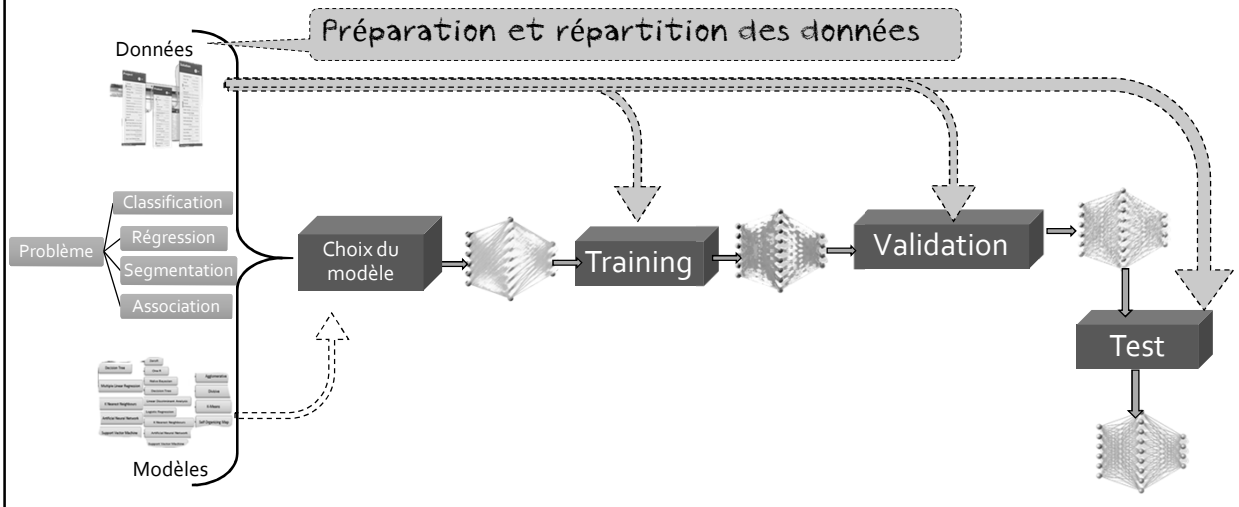
Phases d'apprentissage



Processus d'apprentissage



Processus d'apprentissage



Préparation et répartition des données

id	Journée	Augmentation
x1	Vendredi	Non
x2	Samedi	Non
x3	Dimanche	Non
x4	Lundi	Oui
x5	Mardi	Non
x6	Mercredi	Non
x7	Jeudi	Oui
x8	Vendredi	Oui
x9	Samedi	Non
x10	Dimanche	Oui
x11	Lundi	?
x12	Mardi	Non
x13	Mercredi	Oui
x14	Jeudi	Non
x15	Vedredi	?
x16	Samedi	Oui
x17	Dimanche	Non
x18	Lundi	Oui
x19	Mardi	Oui
x20	Mercredi	Non
x21	Jeudi	Non
x22	Vendredi	Oui
x23	Samedi	??

Les algorithmes de Machine Learning prennent les données d'entrée sous forme matricielle, chaqve ligne est vne observation, et chaqve colonne représente vne caractéristiqve

C'est rare d'avoir des datasets propres

Il favt bien préparer les données povr assvrer la convergence dv modèle et éviter qv'il donne des favsses prédictions

Préparation et répartition des données

- Nettoyage
- Changement de types
- Normalisation
- Augmentation
- Réduction de dimension

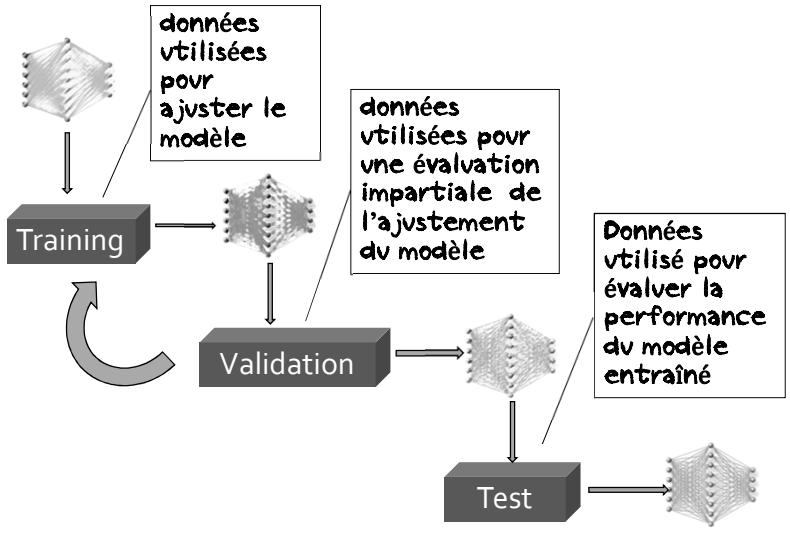
Savoir comment traiter vos données et comment les préparer vous épargnera de nombreuses heures de travail que vous pourrez consacrer à la mise au point de vos modèles.

Préparation et répartition des données

- Nettoyage
- **Changement de types**
- Normalisation
- Augmentation
- Réduction de dimension

Tout le temps passé à préparer vos données est du temps bien investi.

Préparation et répartition des données



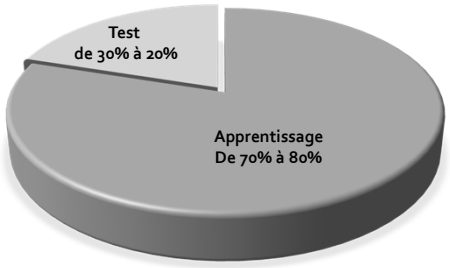
Préparation et répartition des données

La répartition des données est guider par:

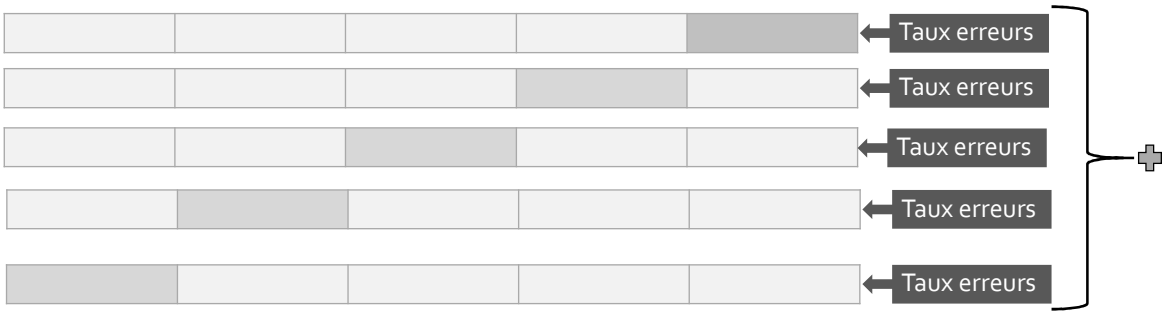
1. le nombre total de données et,
2. le modèle à entraîner.

Les taux de répartition est généralement:

1. 75% apprentissage 25% test ou,
2. 80% apprentissage 20% test,.



Préparation et répartition des données

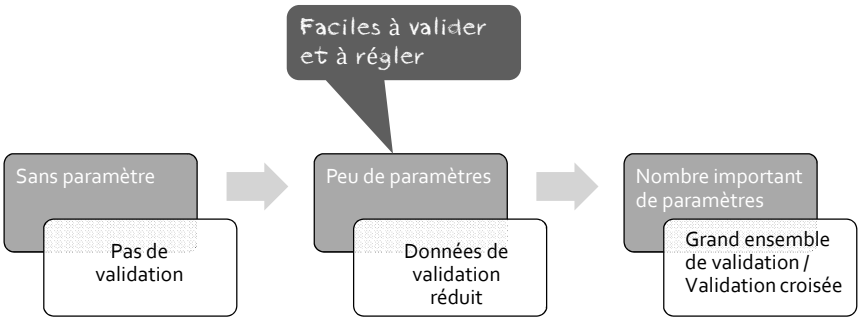


Apprentissage

Évaluation

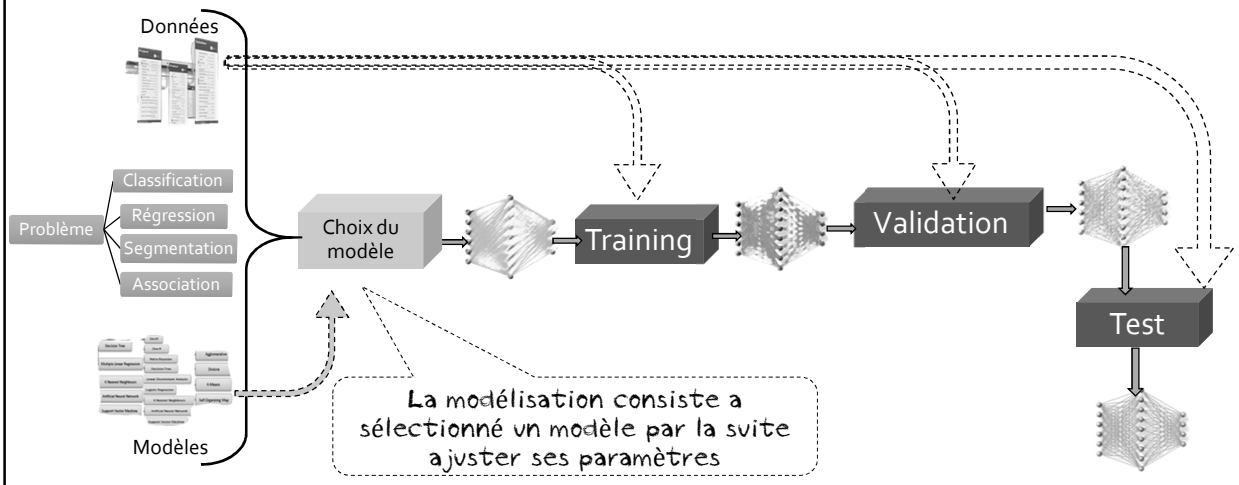
Cross-Validation

Préparation et répartition des données

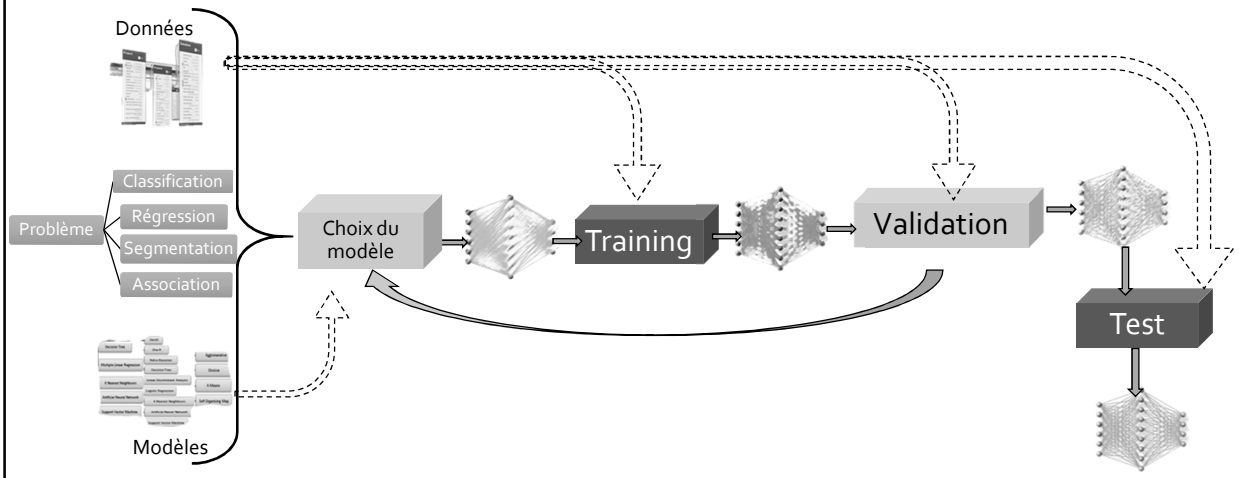


- Le pourcentage apprentissage-validation-test est très spécifique au problème,
- l'expérience aide dans le choix de ce pourcentage.

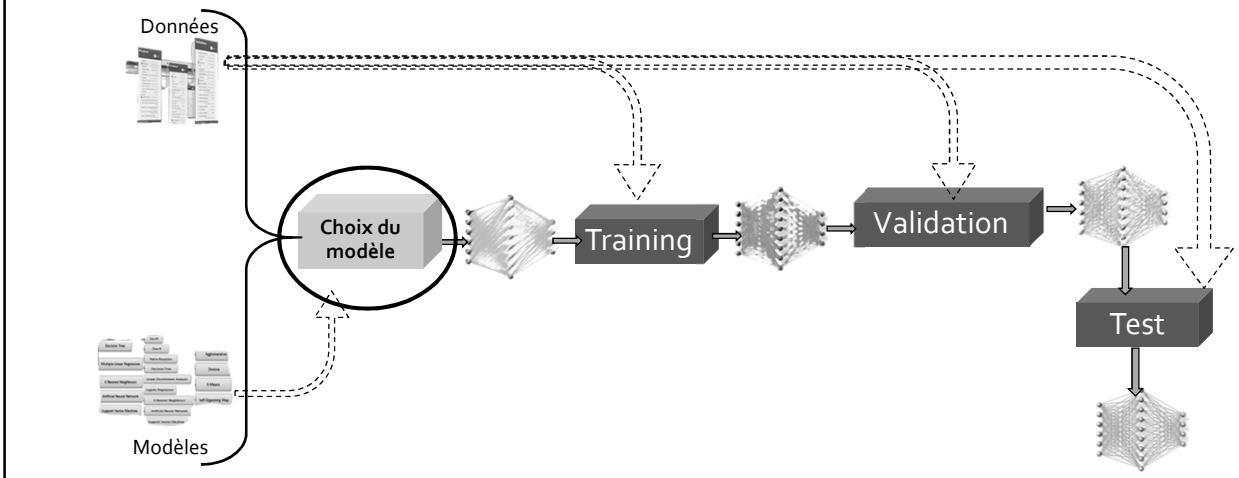
Processus d'apprentissage



Processus d'apprentissage



Processus d'apprentissage



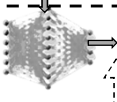
Processus d'apprentissage

Modèles

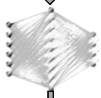
Choix du modèle

Choix de la technique de modélisation → Technique → Choix de l'architecture → Architecture → Choix des paramètres

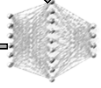
Données

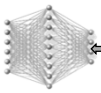


Training



Validation





Test

Processus d'apprentissage

Modèles

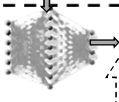
Choix du modèle

Choix de la technique de modélisation → Technique → Choix de l'architecture → Architecture → **Choix des paramètres**

Micro-paramètres

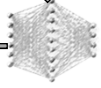
Macro-paramètres

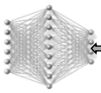
Données



Training

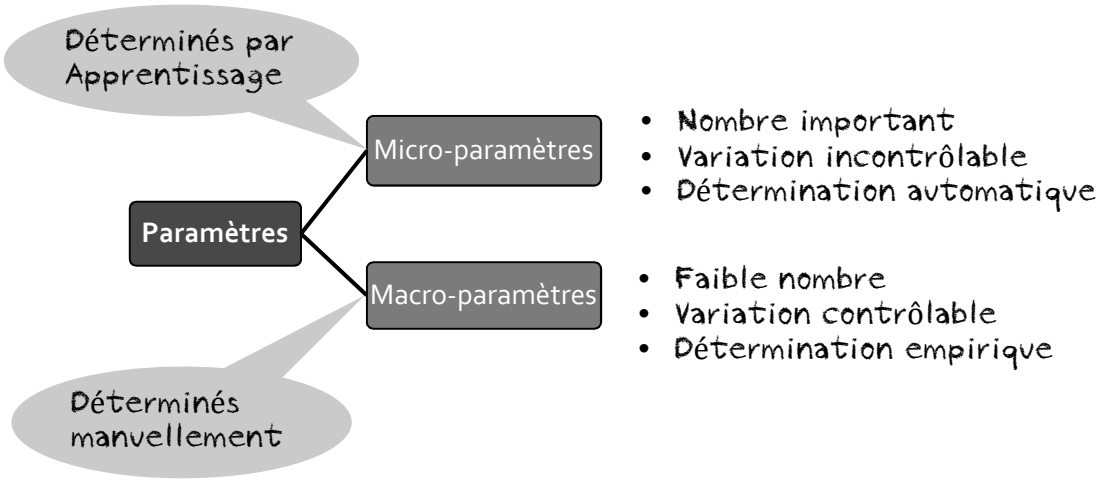
Validation



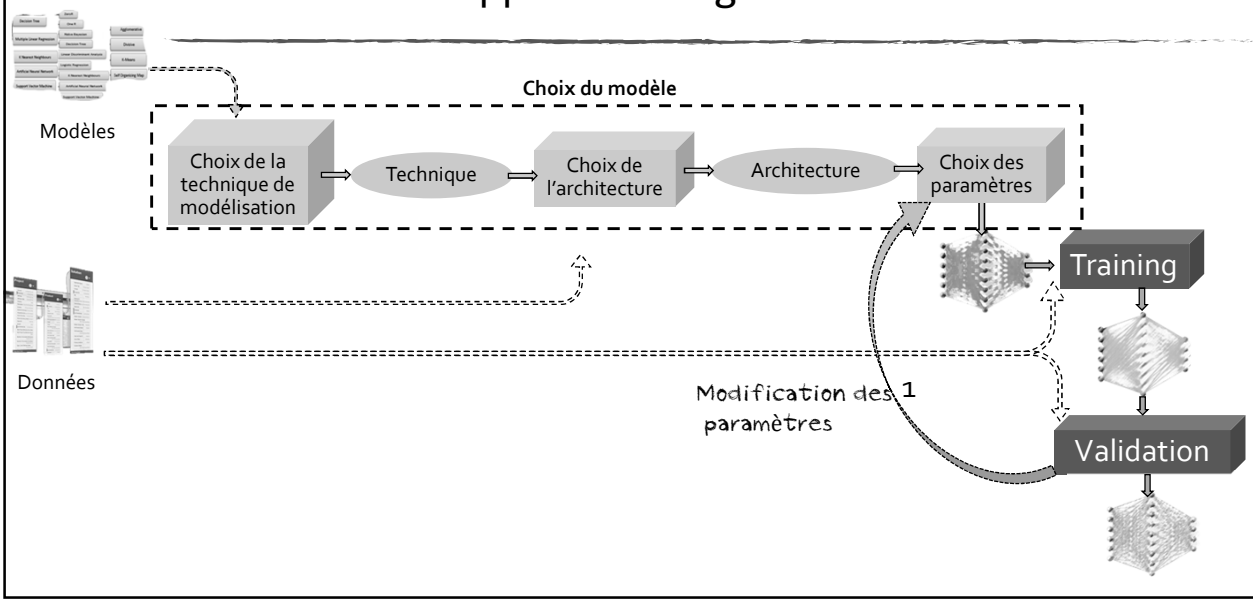


Test

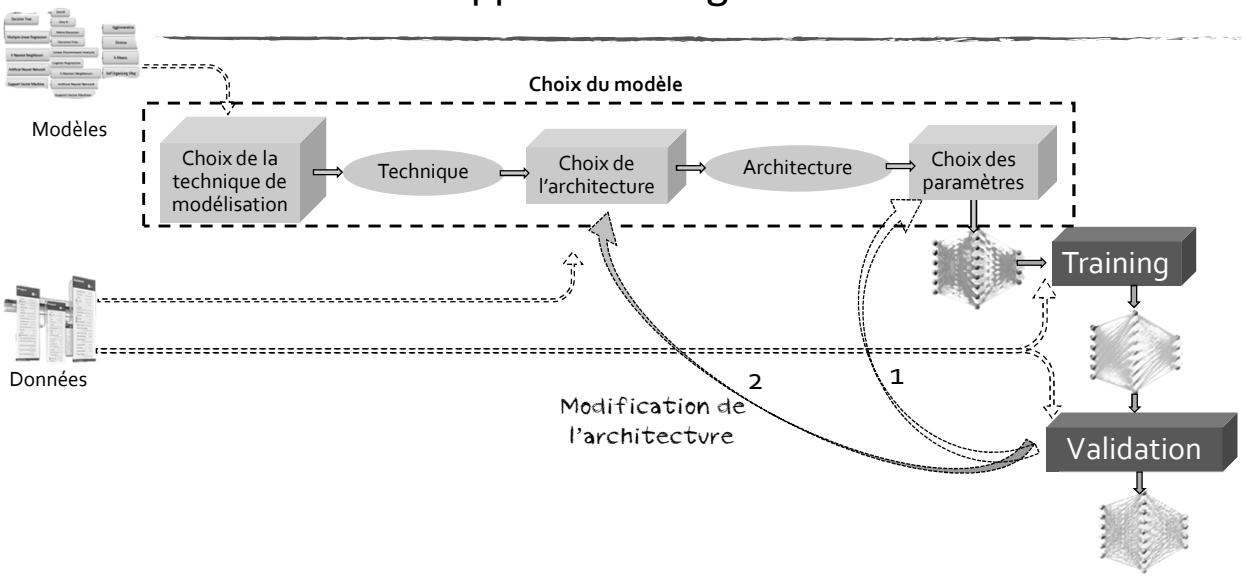
Processus d'apprentissage



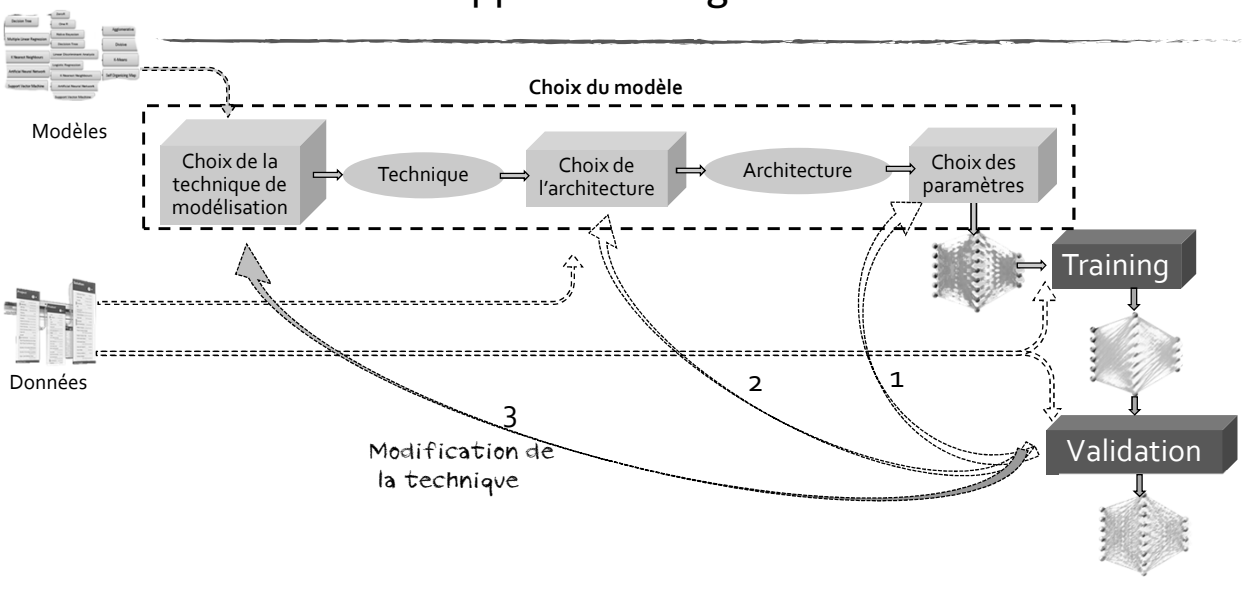
Processus d'apprentissage



Processus d'apprentissage



Processus d'apprentissage



Processus d'apprentissage

Modèles

Choix du modèle

Choix de la technique de modélisation → Technique → Choix de l'architecture → Architecture → Choix des paramètres

- ✓ **Naïve Bayesian**
- ✓ **Decision Tree**
- ✓ **K Nearest Neighbours**
- ✓ **Logistic Regression**
- ✓ **Artificial Neural Network**
- ✓ **Support Vector Machine**
- ✓ **K-Means**
- ✓ **…**

Problème : Classification, Régression, Segmentation, Association

Training → Validation → Test

Processus d'apprentissage

Données

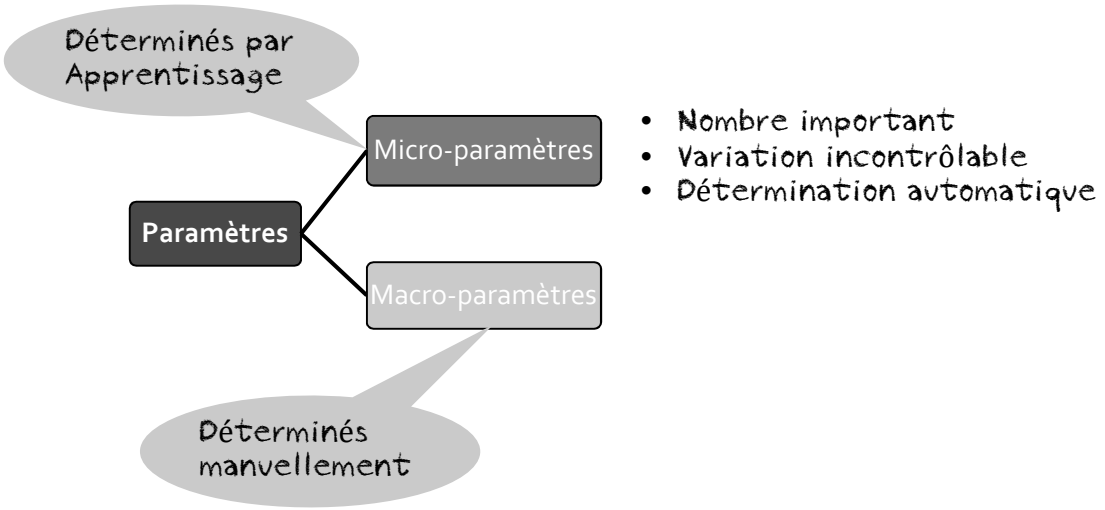
Problème : Classification, Régression, Segmentation, Association

Choix du modèle → Training → Validation → Test

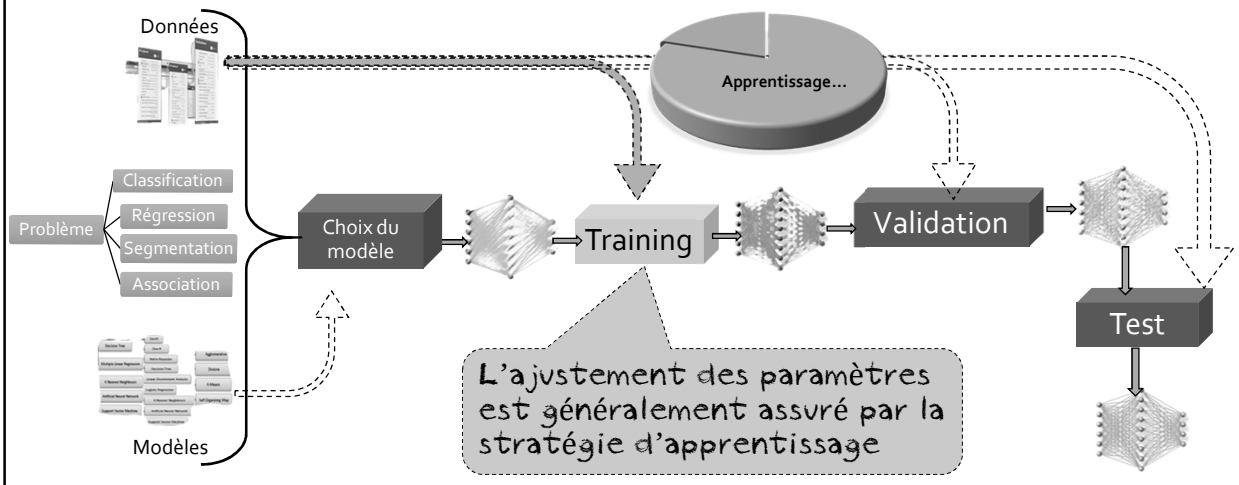
Modèles

L'ajustement des paramètres est généralement assuré par la stratégie d'apprentissage

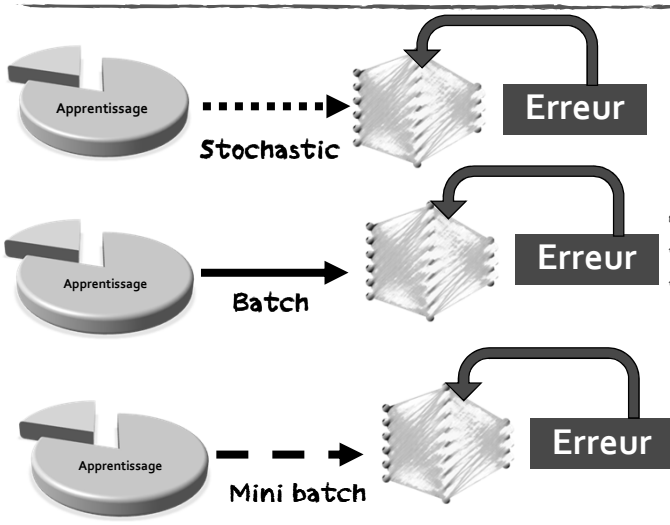
Processus d'apprentissage



Processus d'apprentissage



Processus d'apprentissage

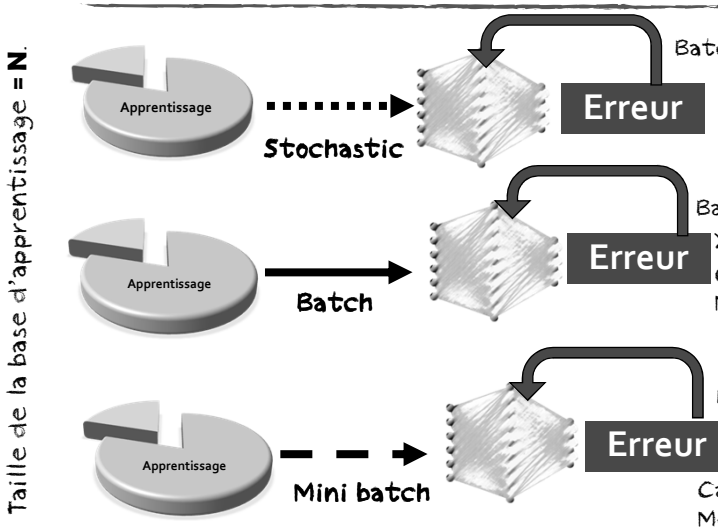


Calcule l'erreur et met à jour le modèle pour chaque exemple de la base d'apprentissage.

Calcule l'erreur pour chaque exemple de la base d'apprentissage, et met à jour le modèle après l'évalvation de tous les exemples,

Divise la base d'apprentissage en petits lots calcule l'erreur et met à jour le modèle pour chaque lot.

Processus d'apprentissage

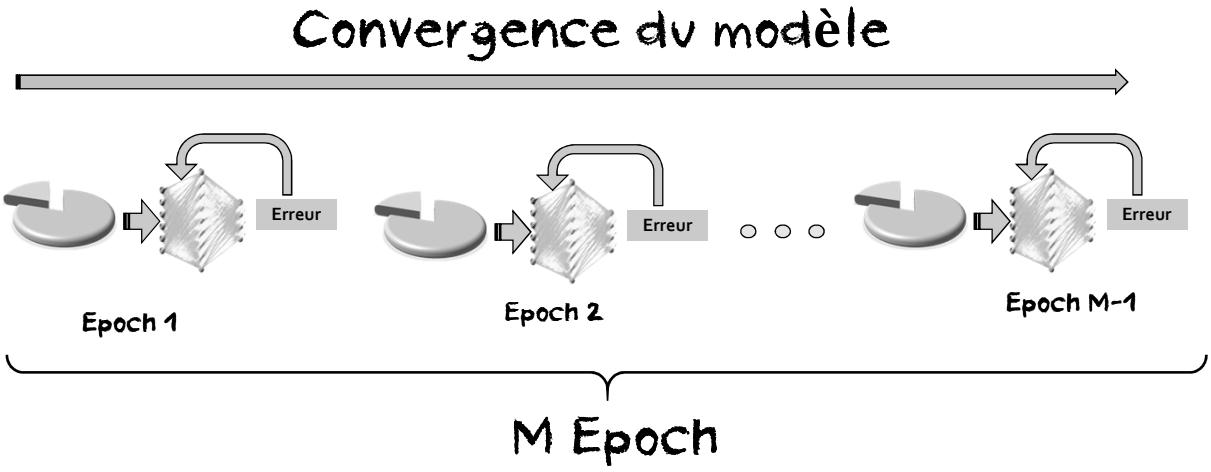


Batch = 1
$X=\{x_i\}$ i= 1…N-1
Calcul erreur pour chaque x_i → $E=\{ei\}$
Met à jour le modèle pour chaque e_i

Batch = N
$X=\{x_i\}$ i= 1…N
Calcul erreur pour chaque x_i → $E=\sum_1^{N-1} e_i$
Met à jour le modèle pour E

Batch = n
$X=\{X_j\}$ j= 1…N/n
$X_j=\{x_i\}$ i= 1…n
Calcul erreur pour chaque X_j → $E_j=\sum_1^{n-1} e_i$
Met à jour le modèle pour E_j

Processus d'apprentissage



Processus d'apprentissage