

Apprentissage Automatique

Motivation

Les modèles paramétriques

- Les estimations sont biaisées car ils dépendent des paramètres.
- Le rôle principal c'est l'estimation des paramètres
- Ces méthodes imposent certaines hypothèses.
- La forme de la fonction à estimer est connue.
- Facile à interpréter.
- Un ensemble fini de paramètres.

Exemple: Régression linéaire

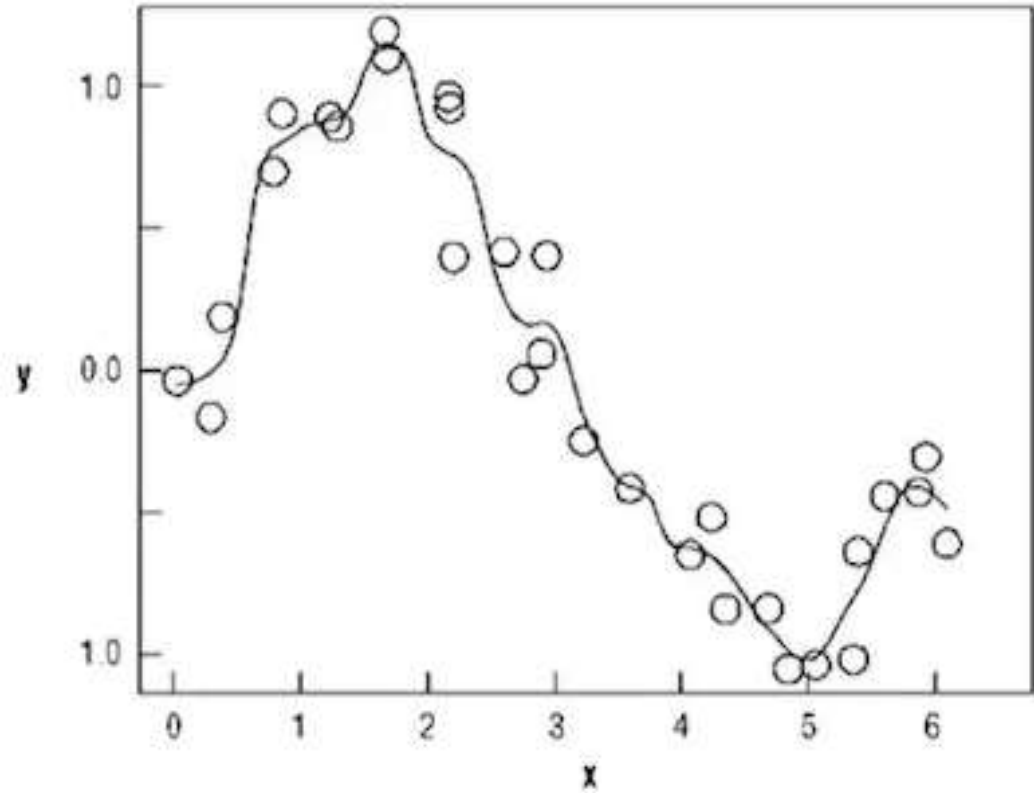
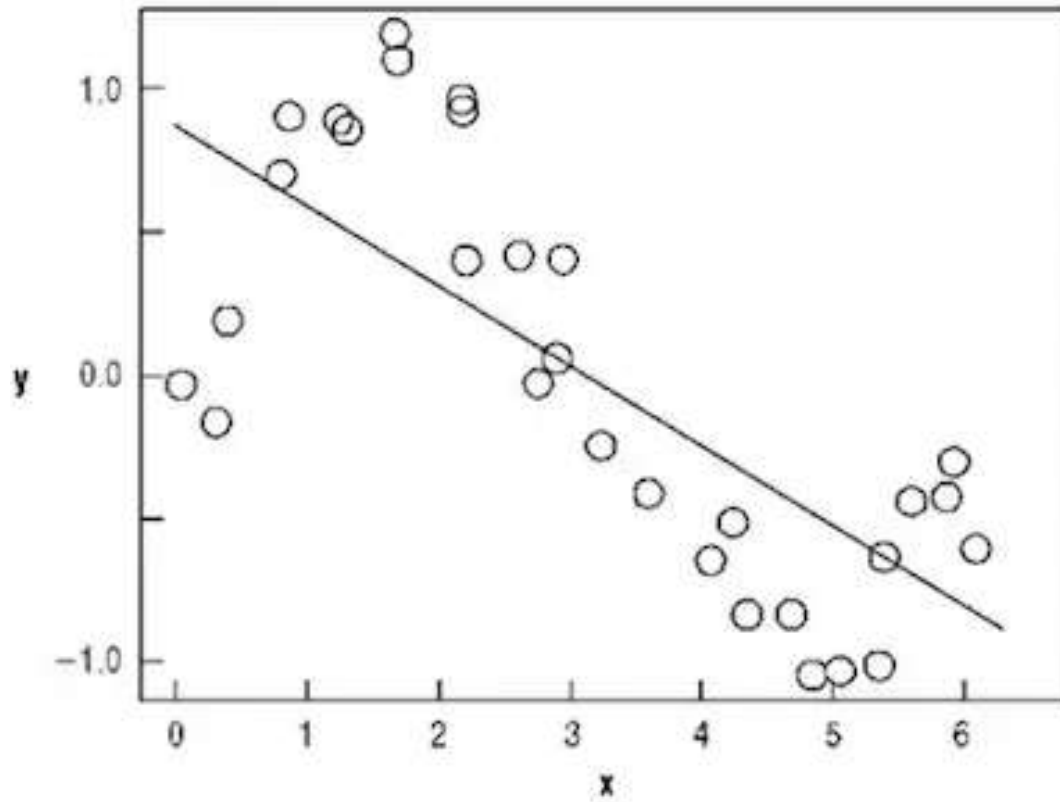
$$y_t = b_0 + a_0 x_t + \varepsilon_t$$

Les modèles non-paramétriques

- Le rôle principal c'est l'estimation de la fonction cible.
- Pas d'hypothèses.
- La forme de la fonction à estimer est inconnue.
- Difficile à interpréter.
- Un ensemble infini de paramètres avec plus de degré de liberté et plus de flexibilité (une meilleure représentation des caractéristiques des données).

$$y_t = f(x_t) + \varepsilon_t$$

Modèle paramétrique /non-paramétrique



plan

- Chapitre 1** Concept Général d'Apprentissage Automatique
- Chapitre 2** Régression
- Chapitre 3** Overfitting – Régularisation – Cross Validation
- Chapitre 4** ANN
- Chapitre 5** SVR
- Chapitre 6** Ensemble méthode

Introduction

- Humain : apprendre des expériences
- Machine : suivre des instructions

Peut-on rendre une machine capable d'apprendre à partir des expériences?

Définition de l'Apprentissage Automatique

- L'apprentissage automatique concerne l'utilisation des bonnes **caractéristiques** pour construire le bon **modèle** qui permet de réaliser les bonnes **tâches**.
- C'est la capacité d'une machine à apprendre sans être explicitement programmée.

Conditions d'Utilisation de l'Apprentissage Automatique

Afin d'utiliser l'apprentissage automatique dans la résolution d'un problème, il faut vérifier la validité de trois critères:

- 1- L'existence d'un modèle à apprendre:** il existe une corrélation entre les variables exogènes et endogènes. Donc, on sait que le modèle existe même si on le connaît pas exactement
- 2- La modélisation mathématique est impossible:** on ne peut pas résoudre le modèle mathématiquement (pas de solution analytique)
- 3- L'existence des données:** il existe des données qui représentent le modèle.

Conditions d'Utilisation de l'Apprentissage Automatique

Dans le cas de ces trois critères, il existe qu'une seule condition qui est suffisante:

- Supposons que le modèle n'existe pas, dans ce cas on peut utiliser l'apprentissage automatique pour vérifier l'existence ou non du modèle.
- Supposons qu'on peut résoudre le modèle mathématiquement, dans ce cas la ML n'est pas la solution recommandée parce qu'elle n'est pas la technique optimale, mais elle peut être utilisée.
- L'existence du **3^{ème} critère** est primordiale car la ML concerne l'apprentissage des données, s'il n'y a pas de données, on peut rien apprendre.

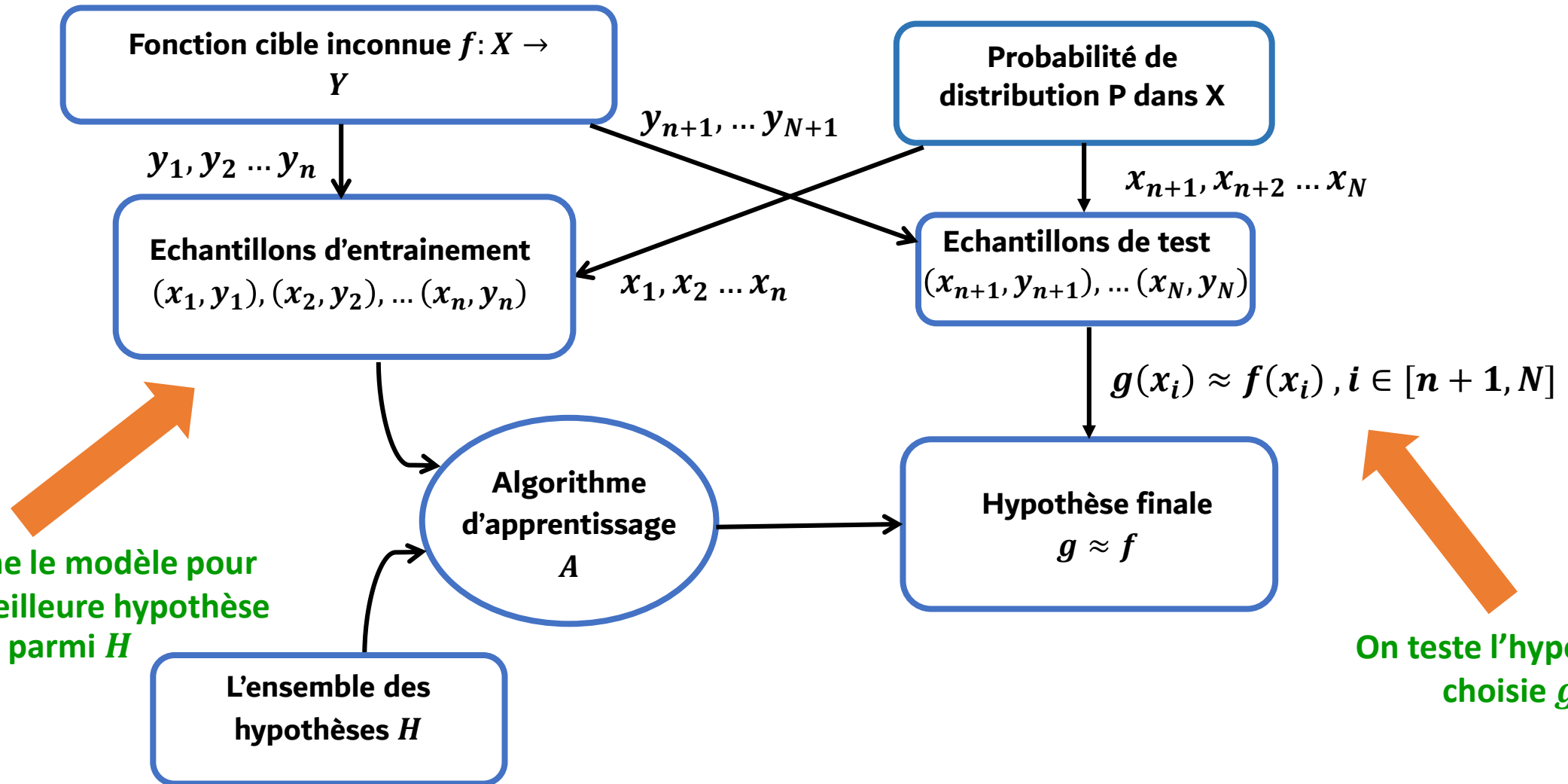
Tâche de l'Apprentissage Automatique

- Régression.
 - Régression statique
 - Régression dynamique (Série temporelle).
- Classification.
- Clustering.

Entités de Base de l'Apprentissage Automatique

- \mathbf{x}_t est une entrée de d coordonnées $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(d)})$
- \mathbf{y}_t est une sortie de e coordonnées $\mathbf{y}_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(e)})$
- \mathbf{X} l'espace des entrées de dimension d
- \mathbf{Y} l'espace des sorties de dimension e
- \mathbf{D} l'ensemble des données $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)$ de dimension N
- f la fonction cible inconnue $f: \mathbf{X} \rightarrow \mathbf{Y}$
- \mathbf{H} l'ensemble des hypothèses $\mathbf{H} = \{h\}$
- g l'hypothèse finale $g: \mathbf{X} \rightarrow \mathbf{Y}, (g \in \mathbf{H})$
- \mathbf{A} l'algorithme d'apprentissage

Processus de l'Apprentissage Automatique



On entraine le modèle pour choisir la meilleure hypothèse g parmi H

On teste l'hypothèse choisie g

Ensemble des données D - Types de présentations des données

- Il faut noter que l'ensemble des données peut être présenté au processus d'apprentissage sous plusieurs formes.
- Les données sont créées et présentées au début du processus d'apprentissage.
- Il existe plusieurs variantes de ce type de protocoles:
 - Apprentissage actif,
 - Apprentissage en ligne (online),
 - Apprentissage hors ligne (off line)...

Ensemble des données D - Types de présentations des données

■ Apprentissage Actif

L'algorithme d'apprentissage est capable d'interagir avec l'utilisateur en lui demandant de choisir le point \mathbf{x}_t à partir de l'ensemble des inputs, puis l'algorithme produit la valeur cible de \mathbf{x}_t .

■ Apprentissage Hors ligne

Les données sont présentées à l'algorithme sous la forme d'un lot (toutes à la fois).

Ce type d'apprentissage est utilisé quand les données sont non-continues.

Ensemble des données D - Types de présentations des données

▪ Apprentissage En ligne

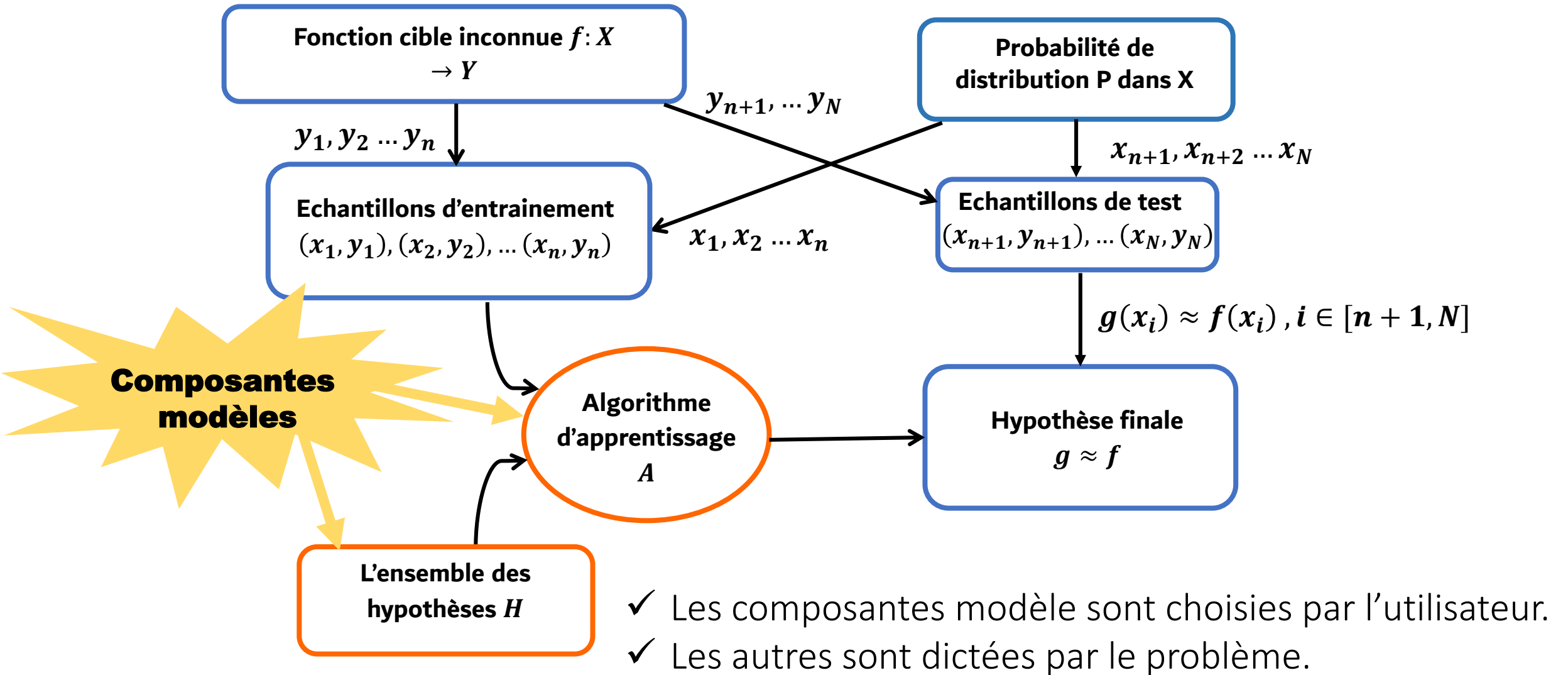
Les données sont présentées à l'algorithme d'une manière incrémentale (une par une).

Ce type d'apprentissage est utilisé quand les données sont continues et que l'algorithme doit procéder au fur et à mesure du temps.

Il est utile dans le cas où il y a des contraintes de calcul ou de stockage qui nous empêchent de traiter tous les données en lot.

On l'appelle aussi un apprentissage stochastique parce que les observations sont présentées d'une manière aléatoire.

Les Composantes Modèles



Les Composantes Modèles - Exemples

- L'ensemble des hypothèses et l'algorithme d'apprentissage sont appelés le **modèle de l'apprentissage automatique**

Exemple:

L'ensemble des hypothèses	L'algorithme d'apprentissage
Les hypothèse des Réseaux de neurones	Algorithme de ANN (Backpropagation)
Hypothèse de la programmation génétique	Algorithme génétique GA
Les hypothèses des machines à vecteur support SVM	Algorithme de SVM

Les Types d'Apprentissage

- La condition fondamentale de l'apprentissage de données est l'utilisation d'un ensemble d'observation pour résoudre un processus.
 - Cette condition reste large et difficile dans certains cas.
- Par conséquent, il existe différents types d'apprentissage permettant de s'adapter avec différentes situations et conditions.
- Les trois principaux types d'apprentissage sont:
 - **Apprentissage Supervisé**
 - Apprentissage Non-supervisé
 - Apprentissage par Renforcement

Exemple: Prédiction de la mortalité mensuelle due à la pollution

Caractéristiques
Température
Humidité
% Monoxide de carbone
% Dioxide de soufre
% Dioxide de nytrogène
% Hydrocarbones
% ozone
...

Données à prévoir
Mortalité totale
Mortalité respiratoire
Mortalité cardiovasculaire

Exemple: Prédiction de la mortalité mensuelle due à la pollution

- C'est un problème d'apprentissage de régression dynamique.
- Les données historiques $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$,
- y_t c'est la mortalité.
- L'entreprise veut utiliser l'apprentissage pour trouver une hypothèse g qui représente comment la mortalité est liée à ces paramètres.

Ensemble des Hypothèses

- Soit le vecteur d'entrées $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(d)})$: les données d'une ville.
- Différents poids sont attribués aux différentes coordonnées de \mathbf{x}_t reflétant leur importance relative dans la mortalité.
- Les coordonnées pondérées sont ensuite combinées pour former « le taux de mortalité ».

Ensemble des Hypothèses

- Cette formule peut s'écrire d'une manière plus compacte:

$$\mathbf{h}(x_t) = \left(\sum_{i=1}^d \mathbf{w}_i x_t^{(i)} - \mathbf{b} \right)$$

- Tel que \mathbf{b} est le seuil et \mathbf{h} est une hypothèse ($\mathbf{h} \in H$)
- \mathbf{h} dépend du poids et du seuil, donc pour choisir les hypothèses, l'utilisateur doit modifier les paramètres \mathbf{w}_i et \mathbf{b}
- d le nombre des caractéristiques.

Ensemble des Hypothèses

$$h(\mathbf{x}_t) = \sum_{i=1}^d \mathbf{w}_i \mathbf{x}_t^{(i)} - \mathbf{b}$$

$$h(\mathbf{x}_t) = \sum_{i=1}^d \mathbf{w}_i \mathbf{x}_t^{(i)} + \mathbf{w}_0$$

Où $\mathbf{w}_0 = -\mathbf{b}$

Introduction d'une nouvelle coordonnée $\mathbf{x}_0 = 1$

$$h(\mathbf{x}_t) = \sum_{i=0}^d \mathbf{w}_i \mathbf{x}_t^{(i)}$$

Ensemble des Hypothèses

$$\mathbf{h}(\mathbf{x}_t) = \mathbf{w}^T \mathbf{x}_t$$

$$\mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix} \quad \text{et} \quad \mathbf{x}_t = \begin{pmatrix} x_t^{(0)} \\ x_t^{(1)} \\ \vdots \\ x_t^{(d)} \end{pmatrix}$$

- L'espace des entrées est le suivant:

$$\mathbf{X} = \{\mathbf{1}\} \times \mathbb{R}^d = \left\{ \left[x_t^{(0)}, x_t^{(1)}, \dots, x_t^{(d)} \right]^T \mid x_t^{(0)} = \mathbf{1}, (x_t^{(1)}, \dots, x_t^{(d)}) \in \mathbb{R} \right\}$$

- L'ensemble des hypothèses: indique les ressources à utiliser.

$$\mathbf{H} = \{ \mathbf{w} \in \{-\mathbf{b}\} \times \mathbb{R}^d / \mathbf{b} \in \mathbb{R} \}$$

- Pour notre cas on va prendre H finie:

$$\mathbf{H} = \{ \mathbf{h}_i / i \in [1, M] \}$$

Ensemble des Hypothèses – Les Poids

- Les poids w_1, \dots, w_d peuvent être négatifs ou positifs tout dépend de la relation qui existe entre l'entrée et la sortie
 - Si la relation est proportionnellement positive \rightarrow le poids est positif
 - Si la relation est proportionnellement négative \rightarrow le poids est négatif
- Le choix optimal des poids et du seuil permet de définir l'hypothèse finale $g \in H$ que l'algorithme produit.
- Ceci revient à dire que l'ensemble des hypothèses H contient une infinité de vecteurs de poids et de seuils.
- D'où il faut choisir les valeurs les plus performantes pour une meilleure approximation de la fonction cible f .

Mesure de l'Erreur - Définition

La mesure d'erreur répond à la question suivante:

Que signifie $h \approx f$?

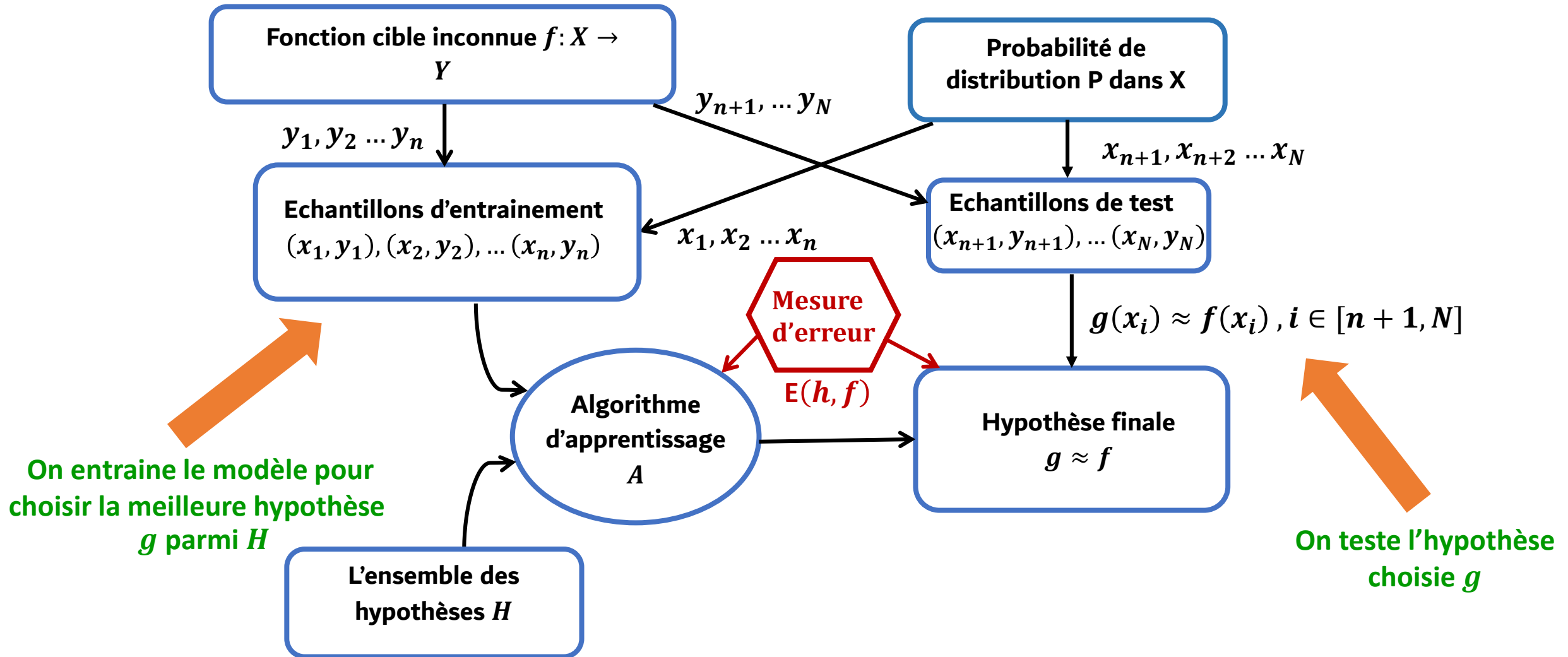
La mesure d'erreur permet de:

- Quantifier comment peut l'hypothèse h approximer correctement f pour guider l'algorithme à passer d'une hypothèse à l'autre.
- Quantifier la distance qui existe entre f et h .

$$\textbf{Erreur} = E(h, f)$$

- Si $E(h, f) = 0 \rightarrow h$ reflète parfaitement f
- Si $E(h, f) \gg 0 \rightarrow$ chercher une autre hypothèse h

Mesure de l'Erreur - Mise à Jour du Diagramme d'Apprentissage



Mesure de l'Erreur - Types

Dans le concept d'apprentissage, il existe deux types d'échantillons:

- Echantillons d'entraînement (x_1, x_2, \dots, x_n)
- Echantillons de test $(x_{n+1}, x_{n+2}, \dots, x_N)$

On distingue deux types d'erreurs:

- Erreur à l'intérieur de l'échantillon :

Erreur Empirique E_{emp}

- Erreur à l'extérieur de l'échantillon :

Erreur Généralisée E_{gen}

Mesure de l'Erreur - Types

- L'erreur empirique:

$$E_{emp}(h) = E_{train}(h) = \frac{1}{n} \sum_{i=1}^n (h(\mathbf{x}_t^{(i)}) - f(\mathbf{x}_t^{(i)}))^2$$

- L'erreur généralisée:

$$E_{gen}(h) = E_{test}(h) = \frac{1}{n} \sum_{i=n+1}^N (h(\mathbf{x}_t^{(i)}) - f(\mathbf{x}_t^{(i)}))^2$$

On ne peut pas connaître E_{gen} , mais on peut l'estimer à travers E_{test} .

Faisabilité d'Apprentissage

- Donc la faisabilité d'apprentissage ($E_{gen}(g) \approx 0$) est répartie en deux questions :

$$1. E_{gen}(g) \approx E_{emp}(g) ?$$

$$2. E_{emp}(g) \approx 0 ?$$

- Ceci permet de mieux comprendre le rôle que jouent les différentes composantes du problème d'apprentissage.
- Et donc traiter la « complexité » de ces composantes.

Faisabilité d'Apprentissage - Complexité de H

Si M augmente $\rightarrow E_{emp}(g)$ est un mauvais estimateur de $E_{gen}(g)$.

- M est une mesure de la « complexité » de l'ensemble d'hypothèses H .
- Si nous voulons répondre par « oui » à la première question:
 - la complexité de H devra être faible.
- Si nous voulons répondre par « oui » à la deuxième question:
 - la complexité de H devra être forte, puisque g provient de H .

Faisabilité d'Apprentissage - Complexité de f

- Si nous voulons répondre par « oui » à la première question:
 - la complexité de f devra être forte.
- Si nous voulons répondre par « oui » à la deuxième question:
 - la complexité de f devra être faible.

Il est impossible d'apprendre une fonction très complexe f .

La plupart des fonctions objectives dans la vie réelle ne sont pas trop complexes. Nous pouvons les apprendre à partir d'un ensemble raisonnable de données D et d'un ensemble raisonnable d'hypothèses H .

Complexité de f : Cibles Bruitées

- La fonction cible n'est pas toujours une fonction
- Une fonction est une notion mathématique qui retourne une valeur unique pour n'importe quel point dans un domaine

Exemples:

Taux de mortalité

- Deux villes différentes peuvent avoir les mêmes caractéristiques mais un taux de mortalité différent.

Complexité de f : Cibles Bruitées

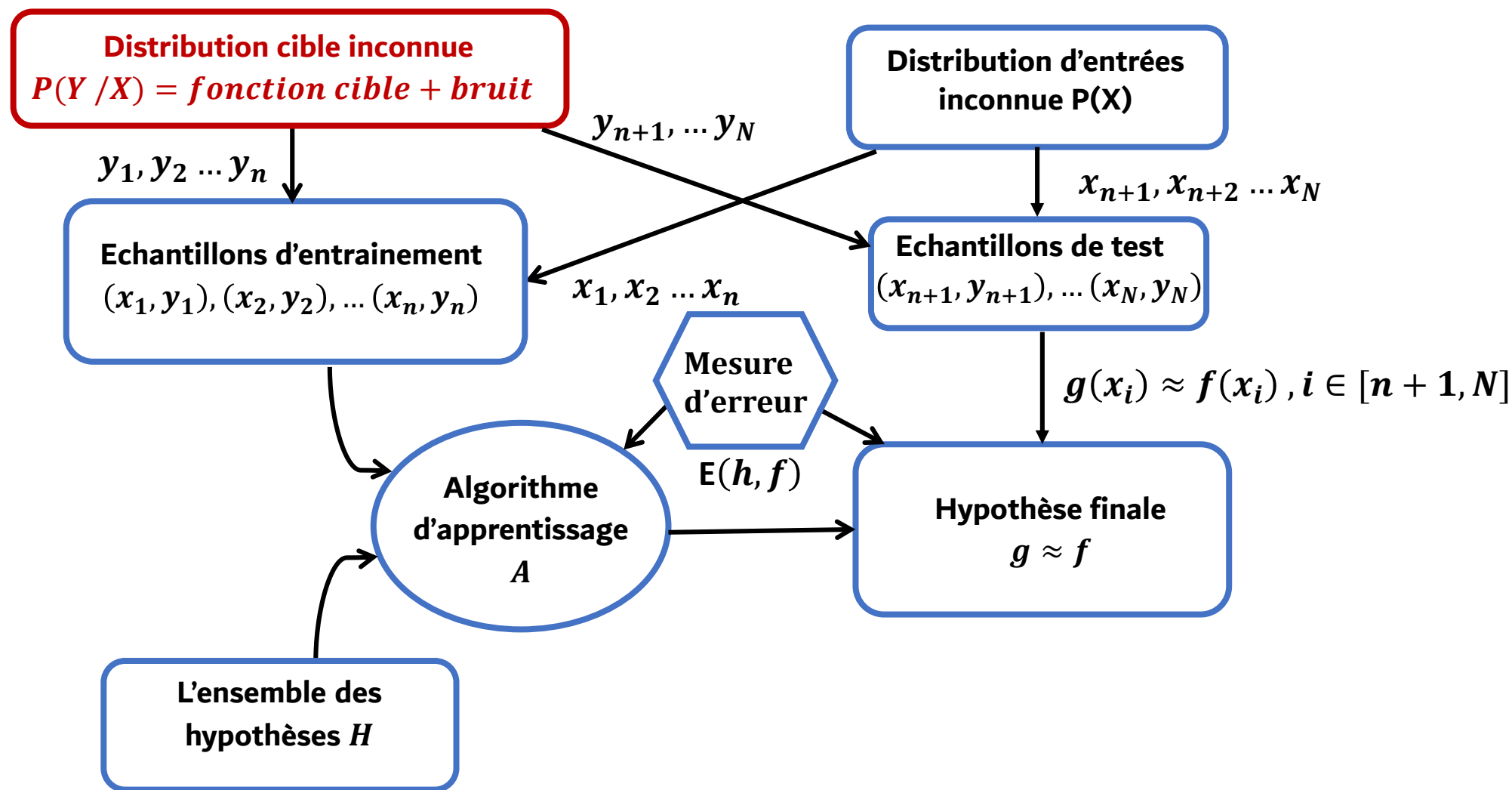
- Dans de nombreuses applications, les données que nous utilisons pour apprendre ne sont pas générées par une fonction cible déterministe.
- Elles sont générées de manière bruitées de telle sorte que la sortie n'est pas uniquement déterminée par l'entrée. **Modèle Stochastique**
- On dit que \mathbf{y}_t est influée par \mathbf{x}_t plutôt que déterminée par \mathbf{x}_t :

Fonction cible $f(\mathbf{x}_t)$ \rightarrow distribution cible $P(\mathbf{y}_t/\mathbf{x}_t)$

\mathbf{y}_t déterministe $\rightarrow \mathbf{y}_t$ probabiliste

$$f(\mathbf{x}_t) = E(\mathbf{y}_t/\mathbf{x}_t) + \textit{bruit}$$

Cibles Bruitées - Mise à Jour du Diagramme d'Apprentissage



Cibles Bruitées - Limites

- L'apprentissage d'une cible bruitée est difficile que l'apprentissage d'une cible déterministe.
- Dans le cas bruité:
 - $E_{gen}(g)$ peut être proche de $E_{emp}(g)$
 - $E_{emp}(g)$ sera mauvaise, car il est difficile de s'adapter avec le bruit.
- Il faut noter que l'erreur et le bruit sont deux notions qui lient le phénomène d'apprentissage aux situations pratiques.

FIN