Chapitre 3 : Modèle PAC rigide

24 février 2026

Plan du chapitre

Le Cadre d'étude : Hypothèses fondamentales

Le cadre est simplifié grâce à deux hypothèses fortes :

1. Cardinalité finie

L'ensemble des hypothèses (modèles) possibles $\mathscr{H}$ est de taille finie : $|\mathscr{H}| < \infty$.

2. Hypothèse de réalisabilité (Étiquetage dur)

Il existe une fonction inconnue $f \in \mathscr{H}$ telle que pour toute donnée $\mathbf{x} \in \mathscr{X}$, la vraie étiquette est strictement :

$$y = f(\mathbf{x})$$

Conséquences :
Une erreur du prédicteur h se traduit par $h(\mathbf{x}) \neq f(\mathbf{x})$.
Il est possible d'avoir une erreur empirique nulle.

Définition : Probablement Approximativement Correct

Une classe $\mathscr{H}$ est **PAC-apprenable** si minimiser le risque empirique sur l'échantillon $\mathscr{S}$ produit un prédicteur dont le risque de généralisation est faible, pour peu qu'on ait assez de données.

Définition formelle

Il existe une fonction de complexité $n_{\mathscr{H}}(\epsilon, \delta)$ telle que pour tous $\epsilon, \delta > 0$, si la taille de l'échantillon $n \geq n_{\mathscr{H}}(\epsilon, \delta)$:

$$\mathbb{P}[\mathscr{L}_{\mathbb{P},f}(h) < \epsilon] \geq 1 - \delta$$

- **Probablement** : avec une grande probabilité ($\geq 1 - \delta$).
- **Approximativement Correct** : marge d'erreur tolérée ($< \epsilon$).

Pourquoi ces deux limites (ϵ et δ) ?

Il est impossible d'exiger une certitude absolue ou une précision parfaite ($\epsilon = 0$ et $\delta = 0$). Ces limites sont inévitables pour deux raisons :

- **Risque sur la représentativité (δ) :** L'échantillon $\mathscr{S}$ peut, par malchance, être non représentatif. Puisque les tirages sont i.i.d., on pourrait théoriquement tirer n fois la même observation !
- **Risque sur la précision (ϵ) :** L'échantillon $\mathscr{S}$ peut être trop petit pour capturer tous les détails fins de la distribution des données et de la fonction d'étiquetage.

Preuve (1/4) : Notations et pire cas

Montrons qu'une classe finie $\mathscr{H}$ est PAC-apprenable.

- L'ensemble des prédicteurs mauvais (qui ne respectent pas la précision ϵ) est

$$\mathscr{H}_B = \{h \in \mathscr{H} : \mathscr{L}_{\mathbb{P},f}(h) > \epsilon\}$$

- L'ensemble des **échantillons trompeurs** est

$$\mathscr{M} = \{\mathscr{S} : \exists h \in \mathscr{H}_B, \mathscr{L}_{\mathscr{S}}(h) = 0\}$$

La MRE renvoie $h_{\mathscr{S}}$ tel que $\mathscr{L}_{\mathscr{S}}(h_{\mathscr{S}}) = 0$.
Si ce $h_{\mathscr{S}}$ se révèle être mauvais en réalité, c'est obligatoirement que l'échantillon $\mathscr{S}$ était trompeur :

$$\{\mathscr{S} : \mathscr{L}_{\mathbb{P},f}(h_{\mathscr{S}}) > \epsilon\} \subseteq \mathscr{M}$$

Preuve (2/4) : La borne de l'Union

Pour majorer la probabilité d'échec, il suffit de majorer la probabilité de tirer un échantillon trompeur $\mathscr{M}$.

Par définition, $\mathscr{M}$ est l'union des échantillons sur lesquels un des mauvais prédicteurs a une erreur nulle :

$$\mathbb{P}^n[\mathscr{M}] = \mathbb{P}^n\left[\bigcup_{h\in\mathscr{H}_B} \{\mathscr{S} : \mathscr{L}_{\mathscr{S}}(h) = 0\}\right]$$

En utilisant la **borne de l'union** $\left(\mathbb{P}(A\cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)\right)$:

$$\mathbb{P}^n[\mathscr{M}] \leq \sum_{h\in\mathscr{H}_B} \mathbb{P}^n\left[\{\mathscr{S} : \mathscr{L}_{\mathscr{S}}(h) = 0\}\right]$$

Preuve (3/4) : Probabilité d'erreur

Évaluons $\mathbb{P}^n[\mathscr{L}_{\mathscr{S}}(h) = 0]$ pour un mauvais modèle h fixe.
L'erreur empirique est nulle si h prédit juste sur **tous** les exemples :

$$\mathbb{P}^n[\mathscr{L}_{\mathscr{S}}(h) = 0] = \mathbb{P}^n[\forall i \in \{1..n\}, h(\mathbf{x}_i) = f(\mathbf{x}_i)]$$

Comme les variables sont indépendantes (i.i.d.) :

$$\mathbb{P}^n[\mathscr{L}_{\mathscr{S}}(h) = 0] = \prod_{i=1}^{n} \mathbb{P}[\mathbf{x}_i : h(\mathbf{x}_i) = f(\mathbf{x}_i)]$$

Or, la probabilité que h soit correct sur un exemple tiré au hasard est $1 - \mathscr{L}_{\mathbb{P},f}(h)$. Puisque $h \in \mathscr{H}_B$, on sait que $\mathscr{L}_{\mathbb{P},f}(h) > \epsilon$.

$$\prod_{i=1}^{n}(1 - \mathscr{L}_{\mathbb{P},f}(h)) < \prod_{i=1}^{n}(1 - \epsilon) = (1 - \epsilon)^n$$

Preuve (4/4) : Conclusion et Complexité d'échantillon

On réinjecte ce résultat dans notre somme initiale :

$$\delta = \mathbb{P}^n[\text{Échec de l'algorithme}] \leq \sum_{h \in \mathscr{H}_B} (1-\epsilon)^n$$

Puisque $\mathscr{H}_B \subseteq \mathscr{H}$, le nombre de mauvais modèles est inférieur à $|\mathscr{H}|$:

$$\delta \leq |\mathscr{H}|(1-\epsilon)^n$$

Rappel mathématique (Inégalité usuelle)

Pour tout $x \in \mathbb{R}$, on a $1 - x \leq e^{-x}$. Donc $(1-\epsilon)^n \leq e^{-\epsilon n}$.

On obtient donc la majoration : $\delta \leq |\mathscr{H}|e^{-\epsilon n}$ et finalement

Théorème PAC d'une classe finie

$$n \geq n_{\mathscr{H}}(\epsilon, \delta) = \frac{1}{\epsilon} \ln\left(\frac{|\mathscr{H}|}{\delta}\right)$$