Classification

Régression Logistique

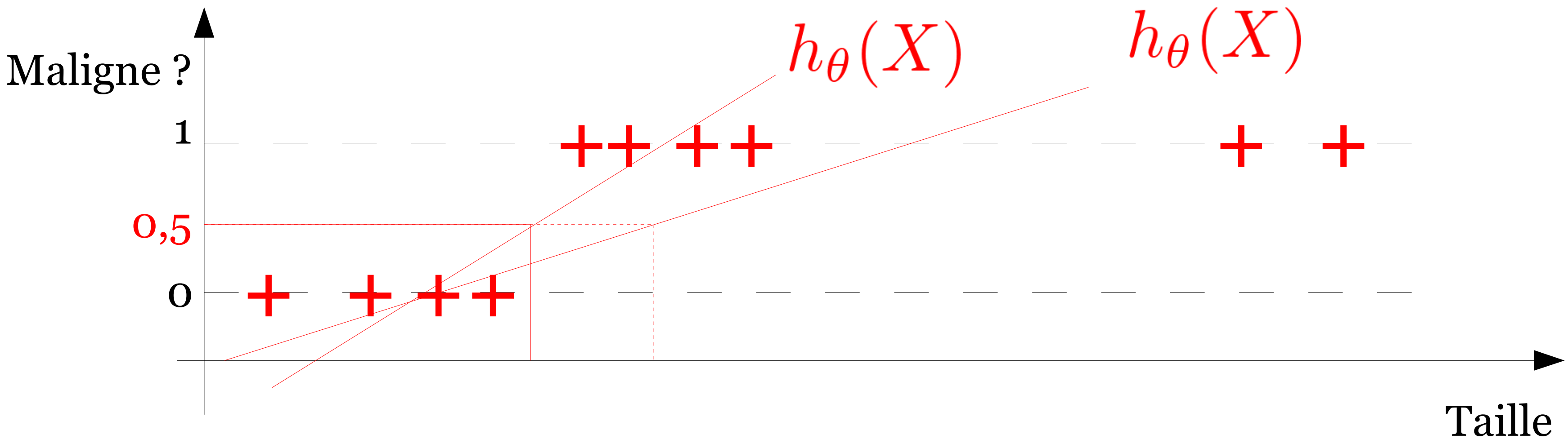
Définition du problème

- Iris : Setosa/Virginica/Versicolor,
- Tumeur : Bénigne/Maligne,
- Email : Normal/Spam.

$$y = \{0, 1, 2\}$$

$$y = \{0, 1\}$$

« Classe négative » « Classe positive»



Si :

- h(X)≥0,5 alors Y=1
- h(X)<0,5 alors Y=0

- Classification : $y = \{0, 1\}$
- h(X)>1 ou h(X)<0 ?
- Régression $\boxed{\text{logistique}}$: $0 \leq h(X) \leq 1$

Régression Logistique

Hypothèse

Objectif

- 0≤h(X)≤1

$$h_\theta(X) = \theta^T X$$

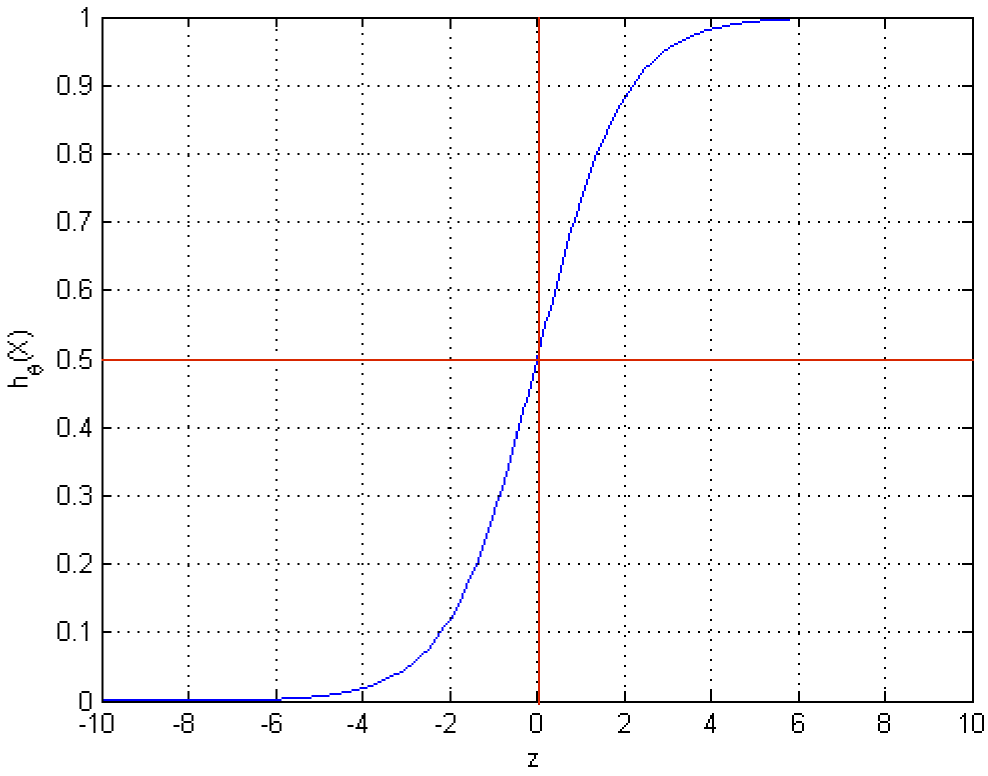
$$h_\theta(X) = g(\theta^T X)$$

$$h_\theta(X) = g(z)$$

$$z = \theta^T X$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

Fonction sigmoïde



Interprétation de l'hypothèse

- h(X) : probabilité que y=1 étant donné x,
- Exemple :

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{taille de la tumeur} \end{bmatrix}$$

$$h_\theta(X) = 0,7$$

- Il y a 70 % de chance que la tumeur soit maligne

Interprétation de l'hypothèse

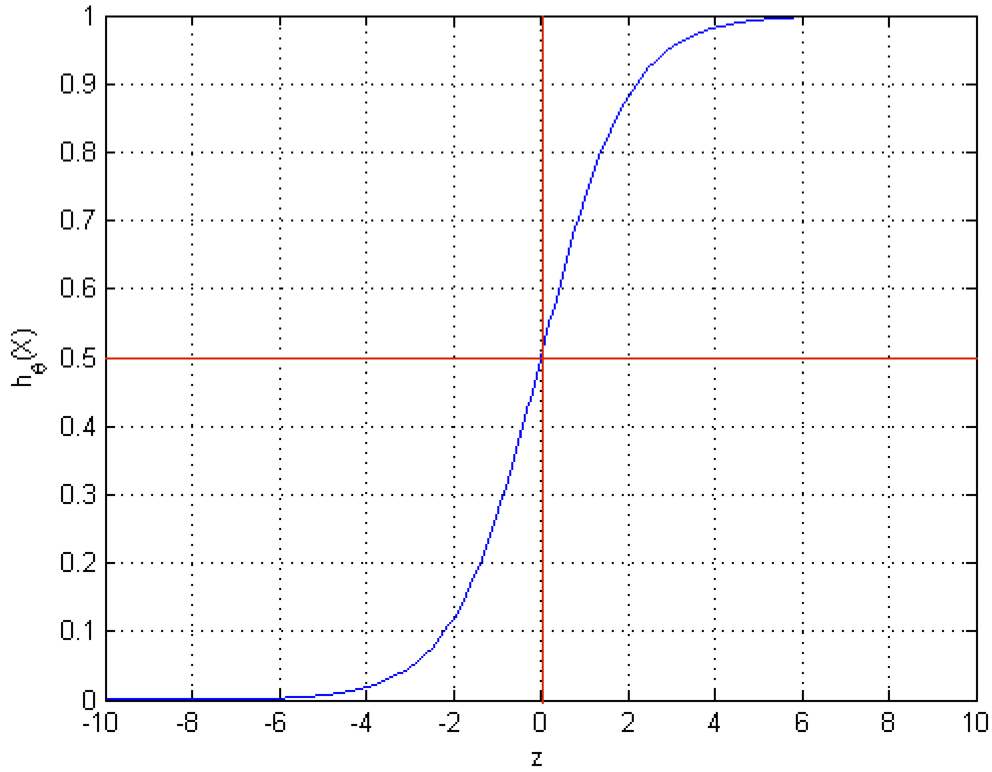
- $h(X) = p(Y=1|X;\theta)$
- Probabilité que Y=1, sachant X, paramétrisé par θ.
- $p(Y=1|X;\theta)+p(Y=0|X;\theta)=1$
- $p(Y=0|X;\theta)=1-p(Y=1|X;\theta)$

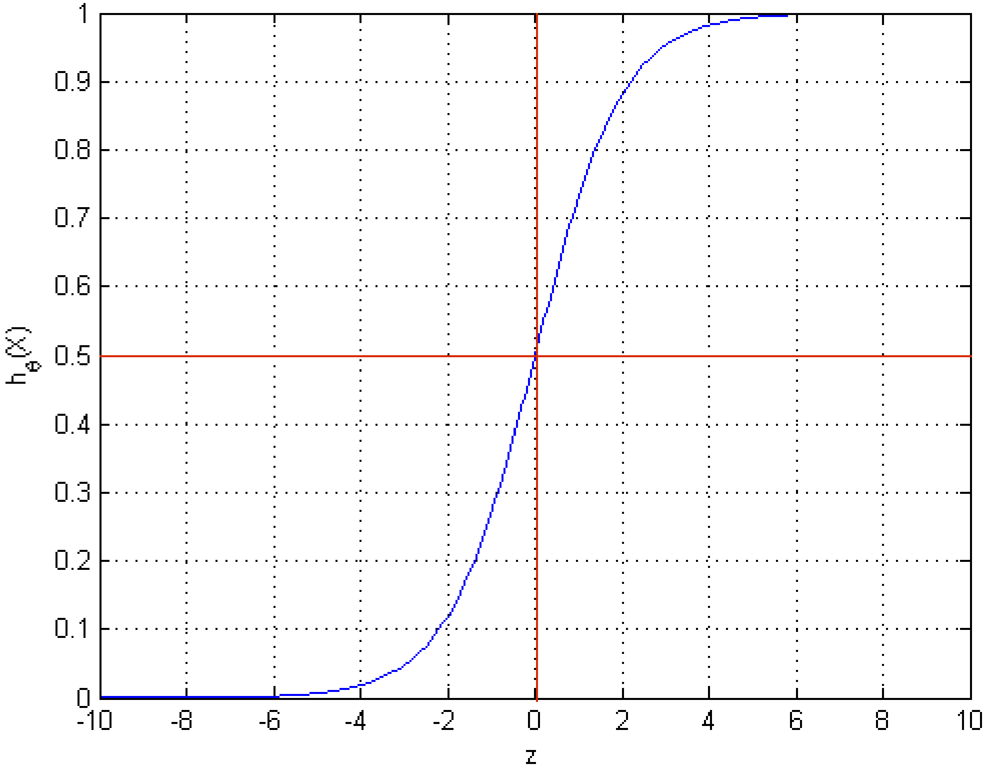
Régression Logistique

Frontière de Décision

$$h_\theta(X) = \theta^T X$$

$$g(z) = \frac{1}{1+e^{-z}} \qquad z = \theta^T X$$



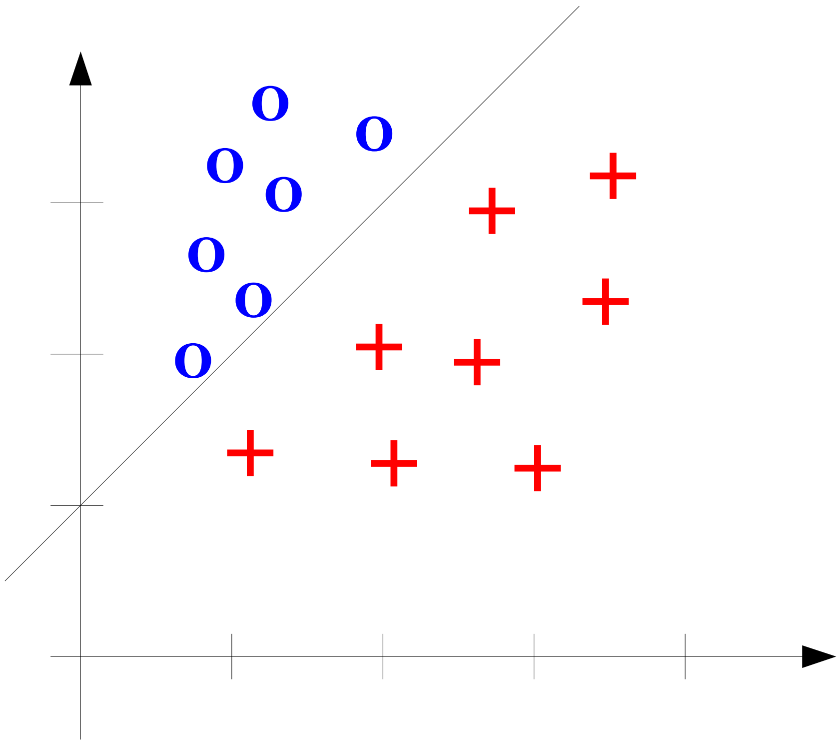


$$z = \theta^T X$$

- Y=1 si h(X)≥0,5
- Y=0 si h(X)<0,5

- Y=1 si z≥0
- Y=0 si z<0

Exemple

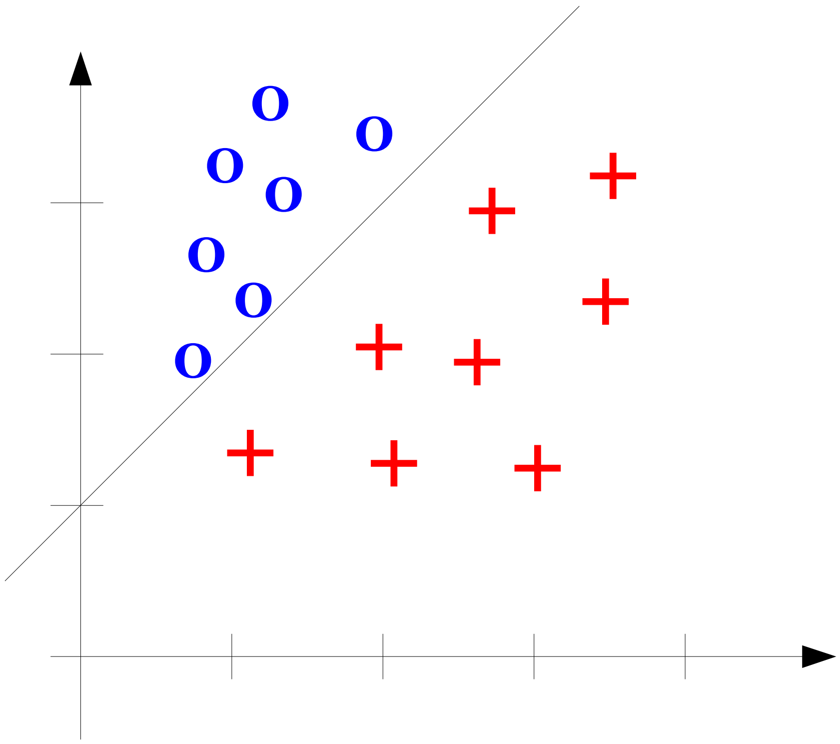


$$h_\theta(X) = g(\theta^T X)$$

$$h_\theta(X) = g(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2)$$

$$\theta = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Exemple



- $Y=1$ si

  $1+x_1+x_2 \geq 0$

- $Y=0$ si

  $1+x_1+x_2 < 0$

- $h(X)=0,5$ si

  $1+x_1+x_2 = 0$

Régression Logistique

Fonction de Coût

Contexte

- Base d'entraînement

$$(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})$$

$$x = \begin{bmatrix} x_0 & x_1 & \ldots & x_n \end{bmatrix}^T \in \mathbb{R}^{n+1}$$

- Hypothèse

$$h_\theta(X) = \frac{1}{1 + e^{-\theta^T X}}$$

Fonction de coût

- Régression :

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X) - Y)^2$$

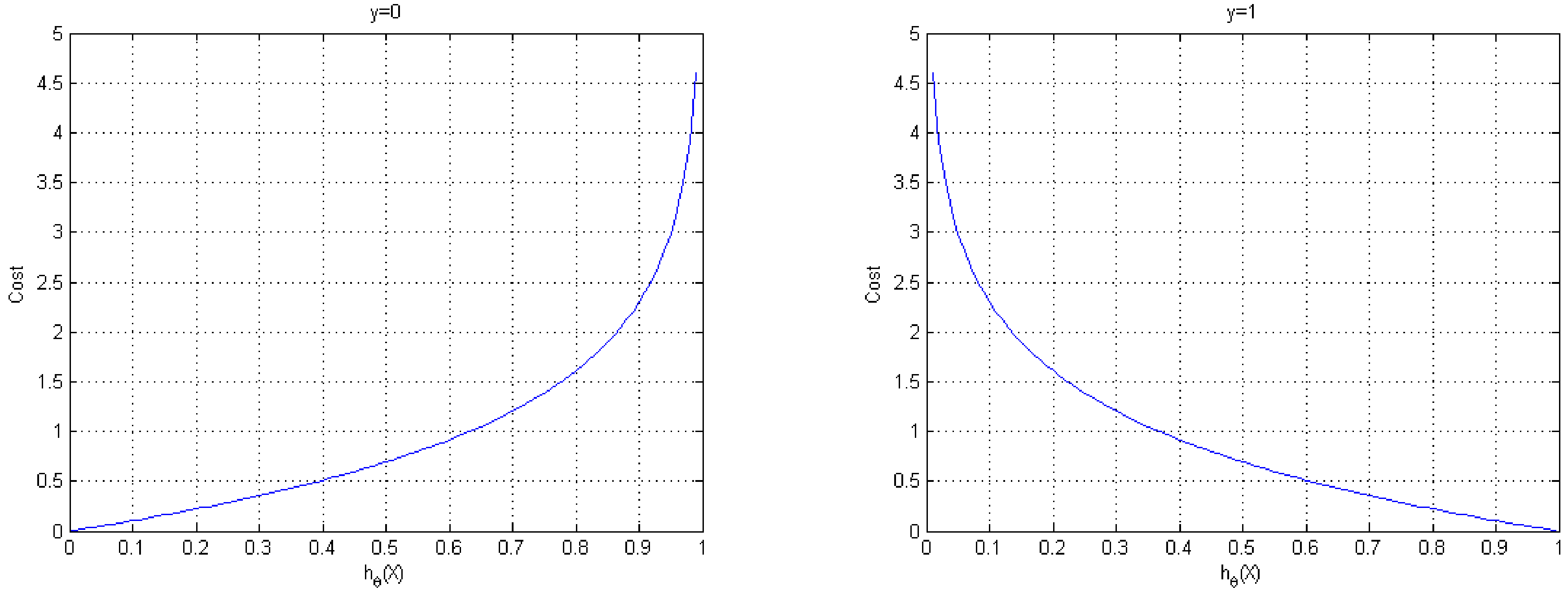
$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} cost(h_\theta(X), Y)$$

Fonction de coût

- Régression logistique:

$$cost(h_\theta(X), Y) = \begin{cases} -log(h_\theta(x)) & si\ Y = 1 \\ -log(1 - h_\theta(x)) & si\ Y = 0 \end{cases}$$

Fonction de coût



Régression Logistique

Descente de Gradient

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h_\theta(X), Y)$$

$$cost(h_\theta(X), Y) = \begin{cases} -log(h_\theta(x)) & si\ Y = 1 \\ -log(1 - h_\theta(x)) & si\ Y = 0 \end{cases}$$

$$cost(h_\theta(X), Y) = -Y \times \log(h_\theta(x)) - (1 - Y) \times \log(1 - h_\theta(x))$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} -ylog(h_\theta(x)) - (1-y)log(1-h_\theta(x))$$

$$min\ J(\theta)$$

Tant que convergence {

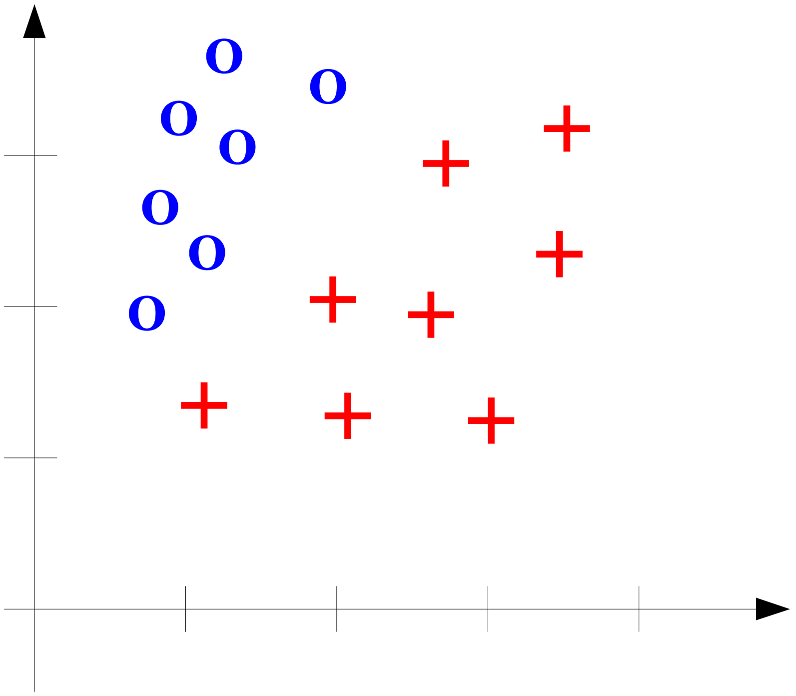
$$\theta_i = \theta_i - \alpha\frac{\partial}{\partial\theta_i}\ J(\theta)$$

}

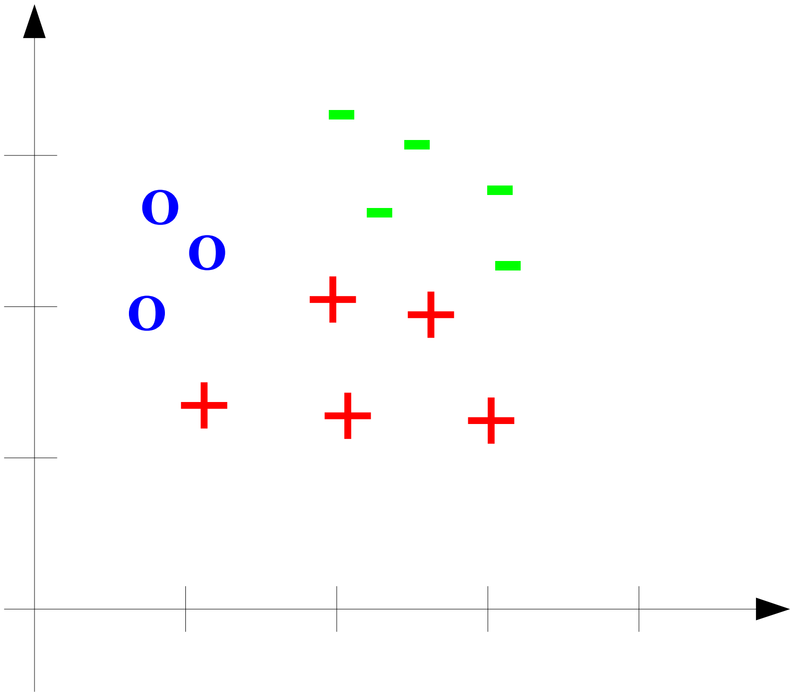
Régression Logistique

Problème Multi-Classes

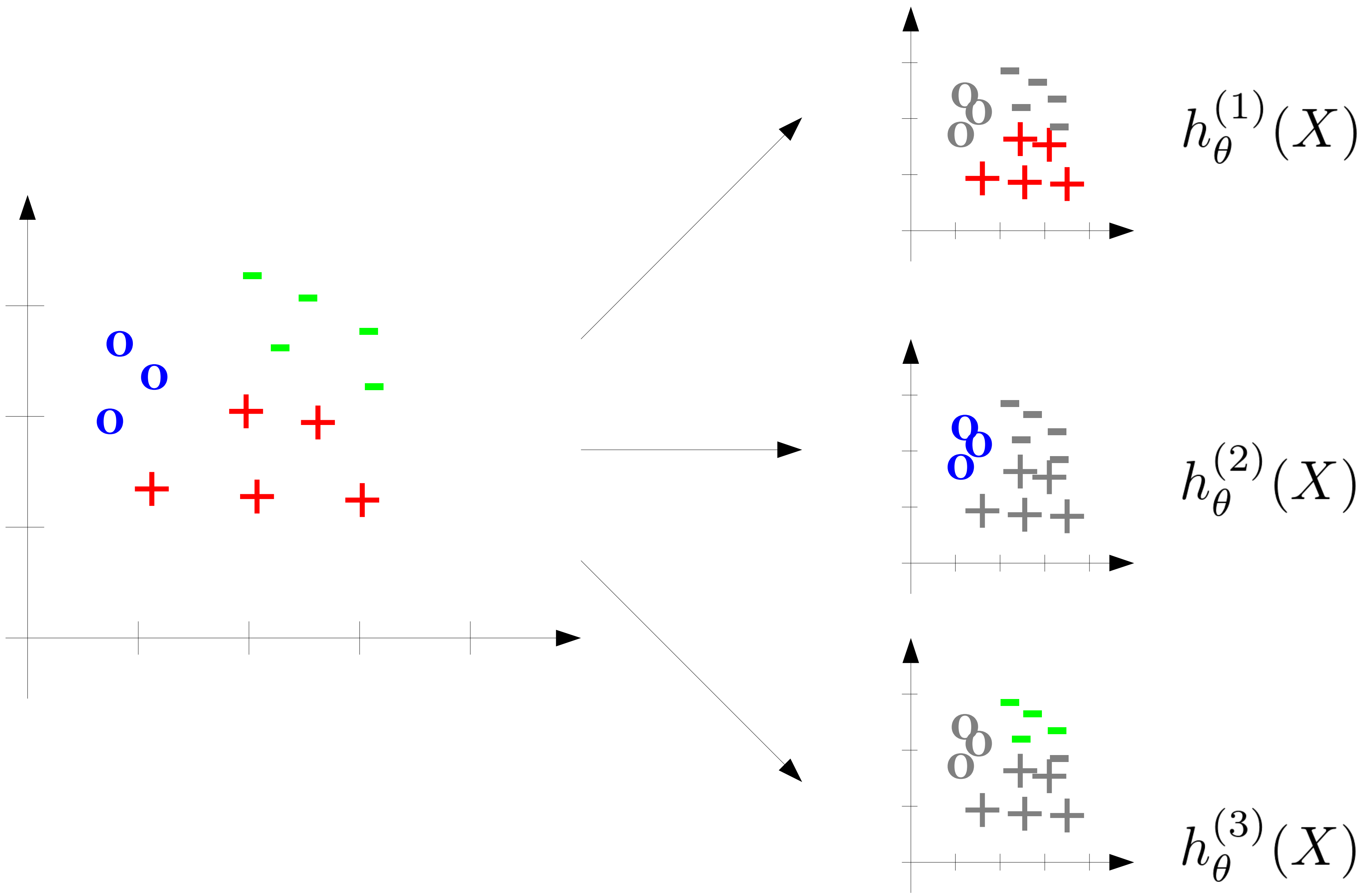
Classification binaire



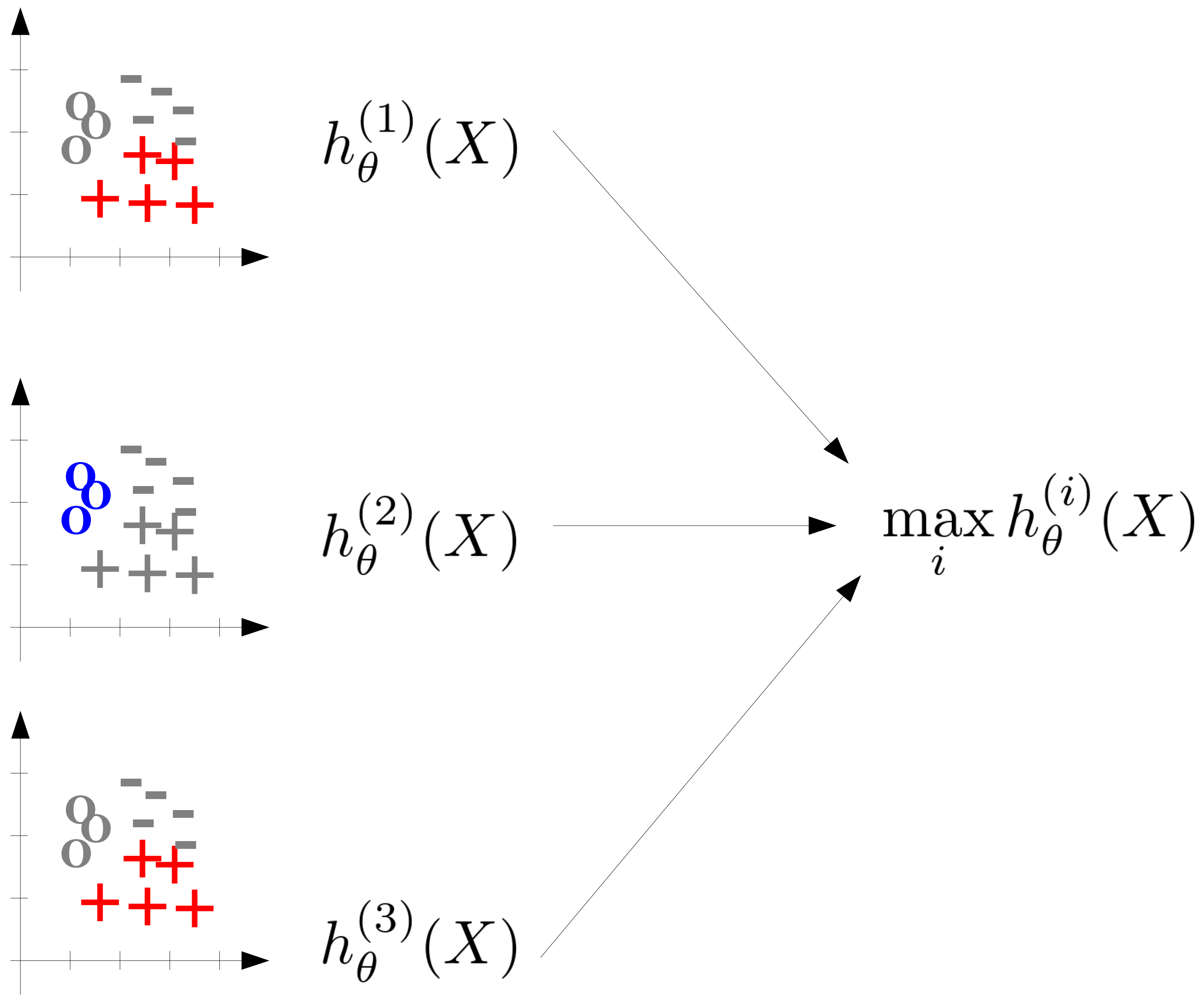
Classification multi-classes



Classification multi-classes



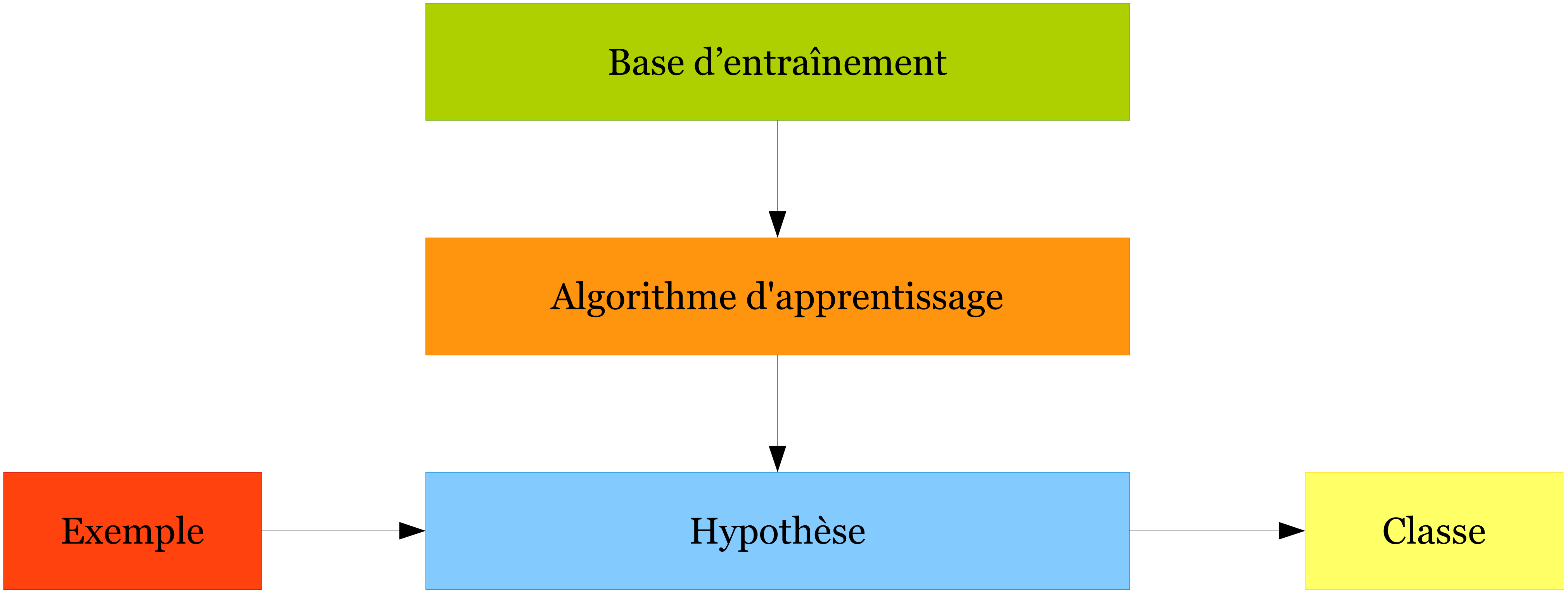
Classification multi-classes



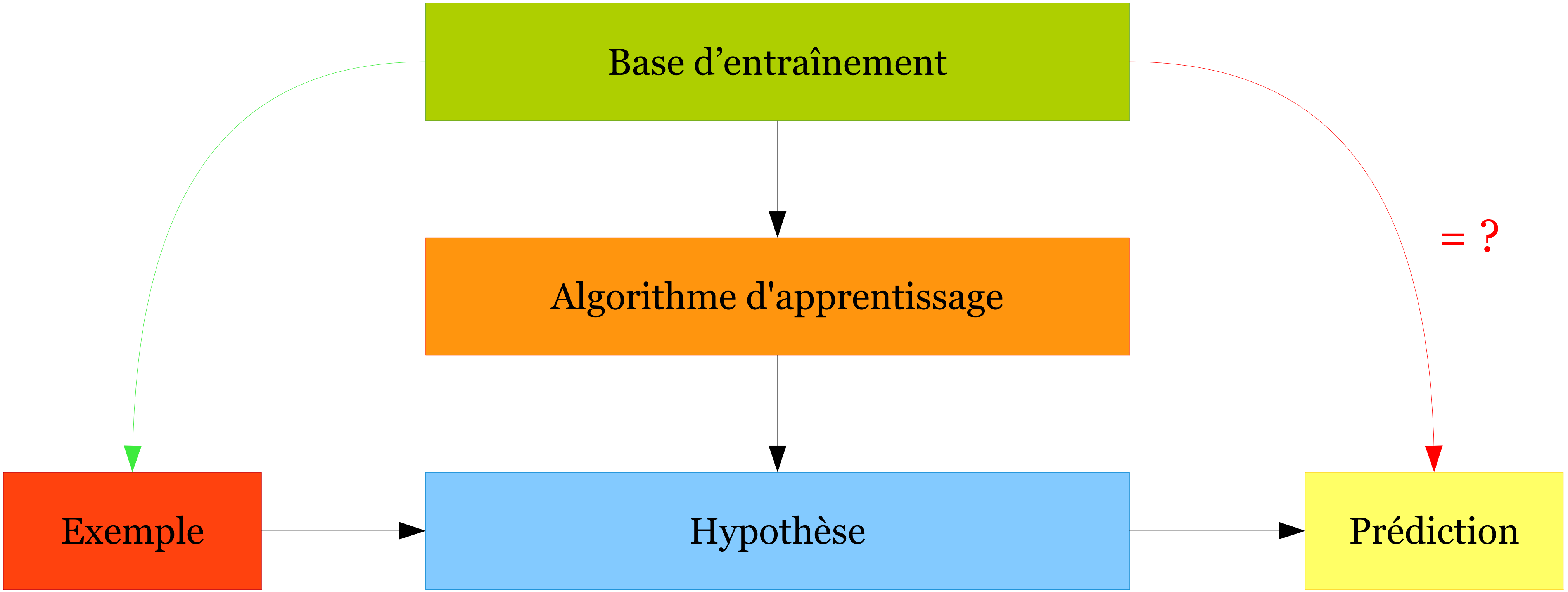
Classification

Évaluation d'Algorithmes

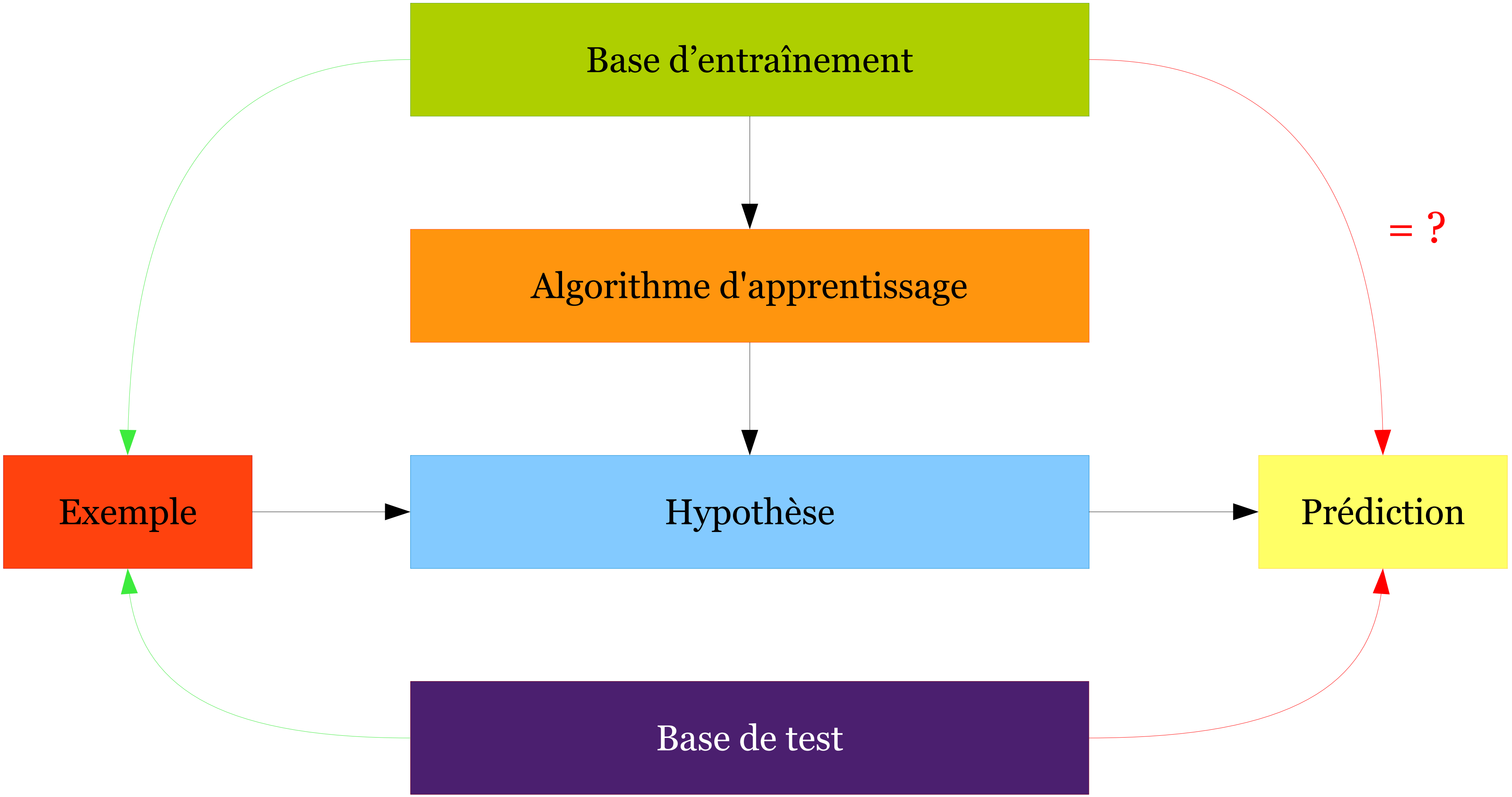
Apprentissage

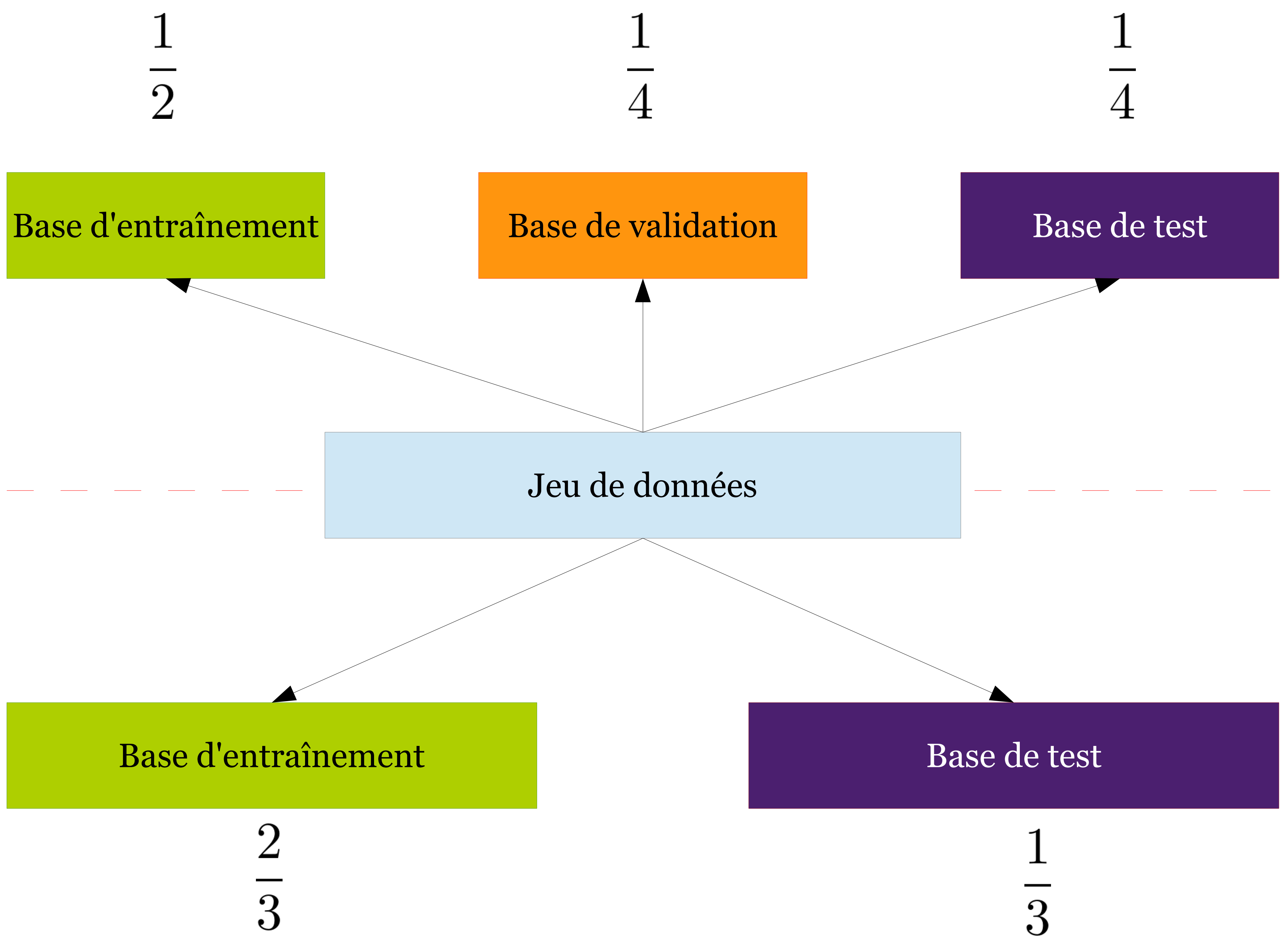


Apprentissage

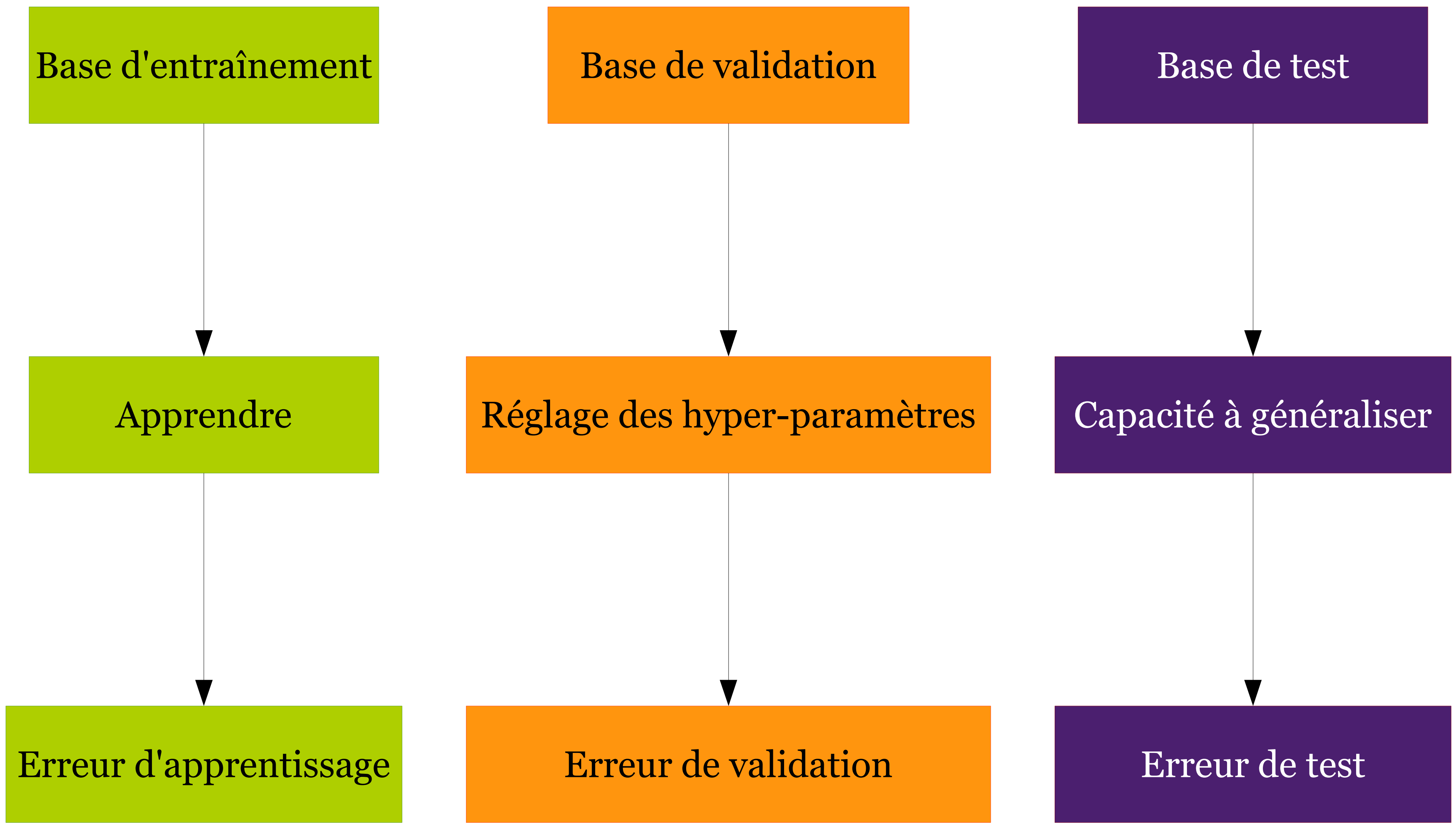


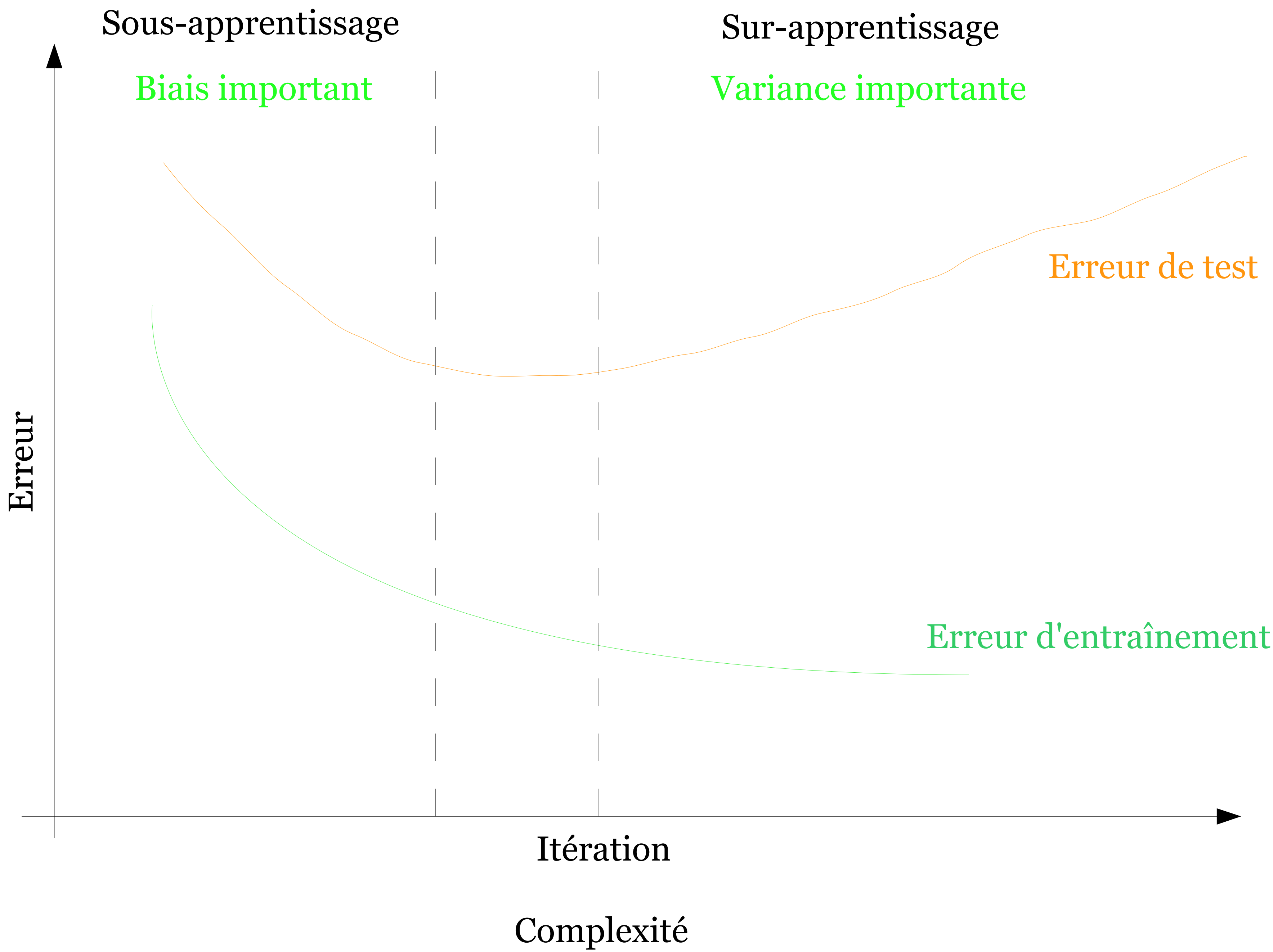
Apprentissage



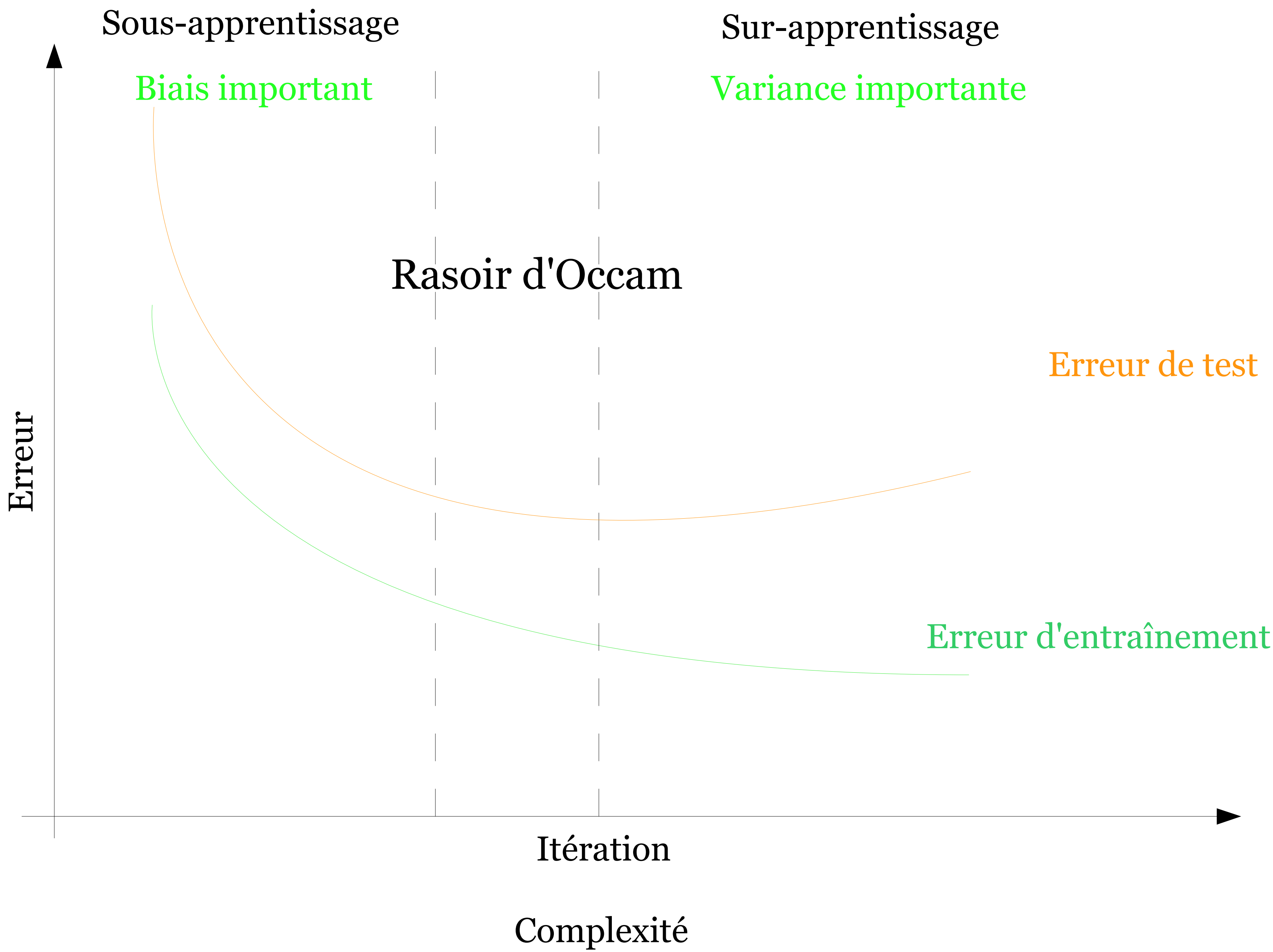


$\frac{1}{2}$
$\frac{1}{4}$
$\frac{1}{4}$
Base d'entraînement
Base de validation
Base de test
Jeu de données
Base d'entraînement
Base de test
$\frac{2}{3}$
$\frac{1}{3}$


Base d'entraînement
Base de validation
Base de test
Apprendre
Réglage des hyper-paramètres
Capacité à généraliser
Erreur d'apprentissage
Erreur de validation
Erreur de test



Sous-apprentissage
Sur-apprentissage
Biais important
Variance importante
Erreur de test
Erreur
Erreur d'entraînement
Itération
Complexité


Sous-apprentissage
Sur-apprentissage
Biais important
Variance importante
Rasoir d'Occam
Erreur de test
Erreur
Erreur d'entraînement
Itération
Complexité

Matrice de Confusion

		Classe estimée	
		1	0
Classe réelle	1		
	0		

		Classe estimée		
		1	2	3
Classe réelle	1			
	2			
	3			

Matrice de Confusion

		Classe estimée	
		1	0
Classe réelle	1		
	0		

		Classe estimée		
		1	2	3
Classe réelle	1			
	2			
	3			

Bonnes Classifications

Mauvaises Classifications

Matrice de Confusion

		Classe estimée	
		1	0
Classe réelle	1	Vrais positifs	Faux négatifs
	0	Faux positifs	Vrais négatifs

Indicateurs

		Classe estimée	
		1	0
Classe réelle	1	a	b
	0	c	d

- Sensibilité ou rappel: probabilité que le classe positive soit prédite alors qu'il s'agit en réalité de la classe positive.

$$\frac{a}{a+b}$$

- Spécificité : probabilité que la classe négative soit prédite alors qu'il s'agit en réalité de la classe négative

$$\frac{d}{c+d}$$

Indicateurs

		Classe estimée	
		1	0
Classe réelle	1	a	b
	0	c	d

- Précision : probabilité que, lorsque la classe positive est prédite, il s'agisse en réalité de la classe positive.

$$\frac{a}{a+c}$$

- Taux de faux positifs : probabilité qu'une classe prédite comme positive soit en réalité négative.

$$\frac{c}{c+d}$$

Indicateurs

		Classe estimée	
		1	0
Classe réelle	1	a	b
	0	c	d

- Taux de bonne classification :

$$\frac{a+d}{a+b+c+d}$$

Quiz #1

		Classe estimée	
		1	0
Classe réelle	1	50	10
	0	20	30

En utilisant la matrice de confusion ci-dessus, calculez :

1. Le taux de bonne classification
2. Le rappel
3. La précision

Merci de votre Attention

Questions