

Mathématiques pour l'intelligence artificielle

M2 informatique parcours IAAA, Aix-Marseille Université

Thomas Schatz (thomas.schatz@univ-amu.fr)

Mardi 1 septembre 2026

Questions pratiques

- Calendrier des cours sur le site IAAA + ENT AMU
- Questions, etc. : thomas.schatz@univ-amu.fr
- Matériel de cours sur mon site web (<https://thomas.schatz.cogserver.net/teaching/>)
- Evaluation: $\text{Note finale} = 0.5 \cdot \text{CC} + 0.5 \cdot \text{ET}$
 - CC contrôle continu :
 - 2 point : exercices à rendre (correction d'une sélection aléatoire)
 - 8 points : oral, probablement mardi 15 septembre, à confirmer
 - ET examen terminal : oral, probablement mardi 6 Octobre, à confirmer
 - Oraux au format "colle" : 1h, 3 ou 4 étudiants en parallèle, chacun résout un exercice différent sur sa partie du tableau

Plan du cours (prévisionnel)

Introduction générale

1. Algèbre linéaire (2h)

2. Optimisation (4h)

Oral 1

3. Modélisation probabiliste (2h)

4. Probabilités, statistique et théorie de l'apprentissage (4h)

Oral 2

Introduction générale

Introduction générale

Mathématiques pour l'apprentissage automatique ?

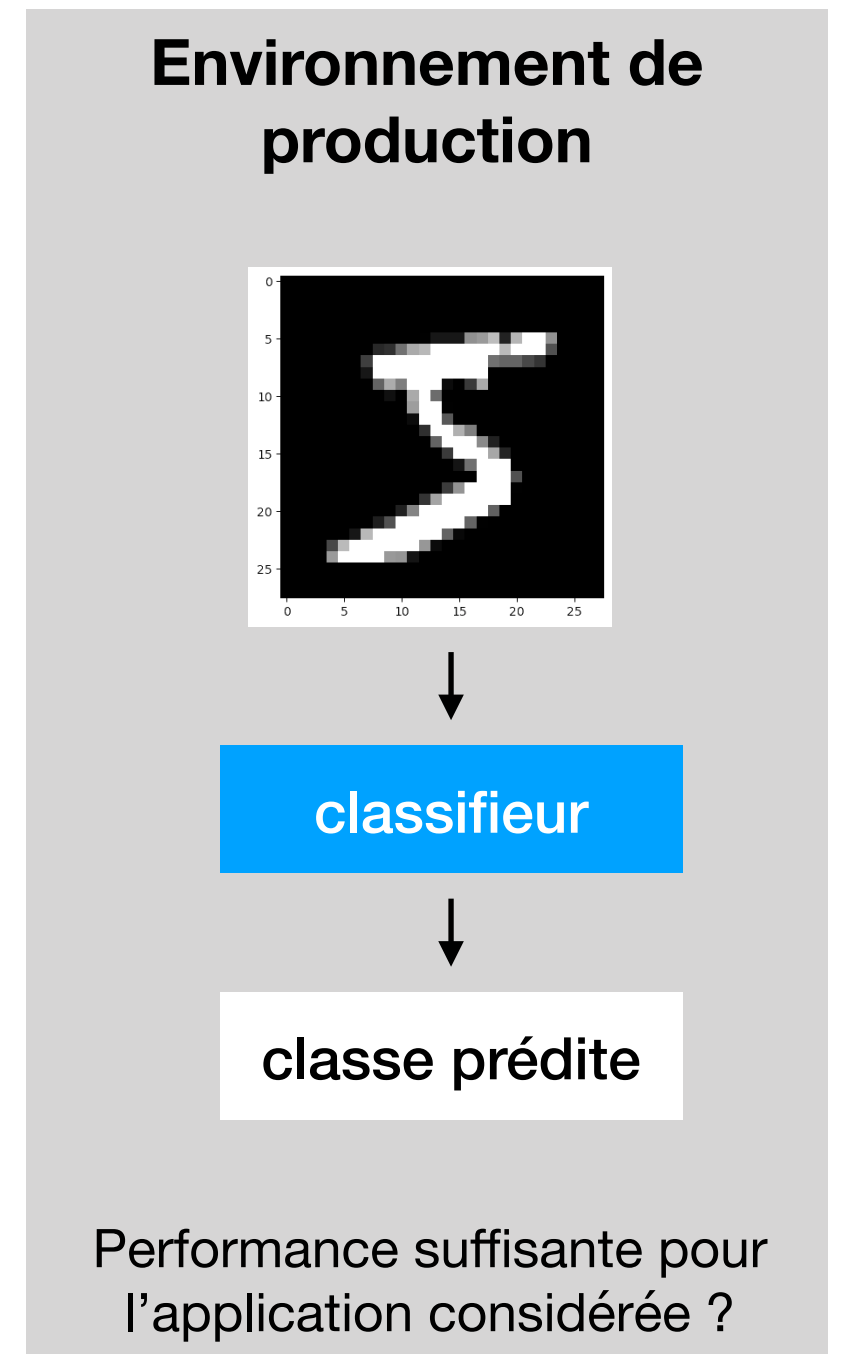
Introduction générale

Mathématiques pour l'apprentissage automatique ?

Apprentissage automatique == Généralisation
+
Calcul intensif

Introduction générale

Structure typique d'un problème d'apprentissage automatique



Introduction générale

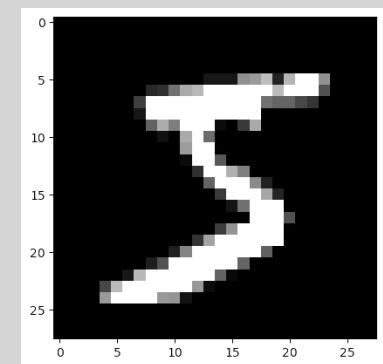
Structure typique d'un problème d'apprentissage automatique

Entraînement (procédure de recherche)

Données d'entraînement



Environnement de production



classifieur

classe prédite

Performance suffisante pour l'application considérée ?

Introduction générale

Structure typique d'un problème d'apprentissage automatique

Entraînement (procédure de recherche)

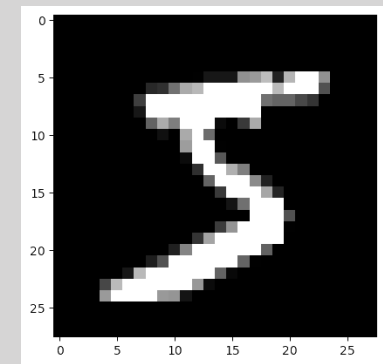
Espace de recherche

Hypothèse/paramètre

Données d'entraînement



Environnement de production



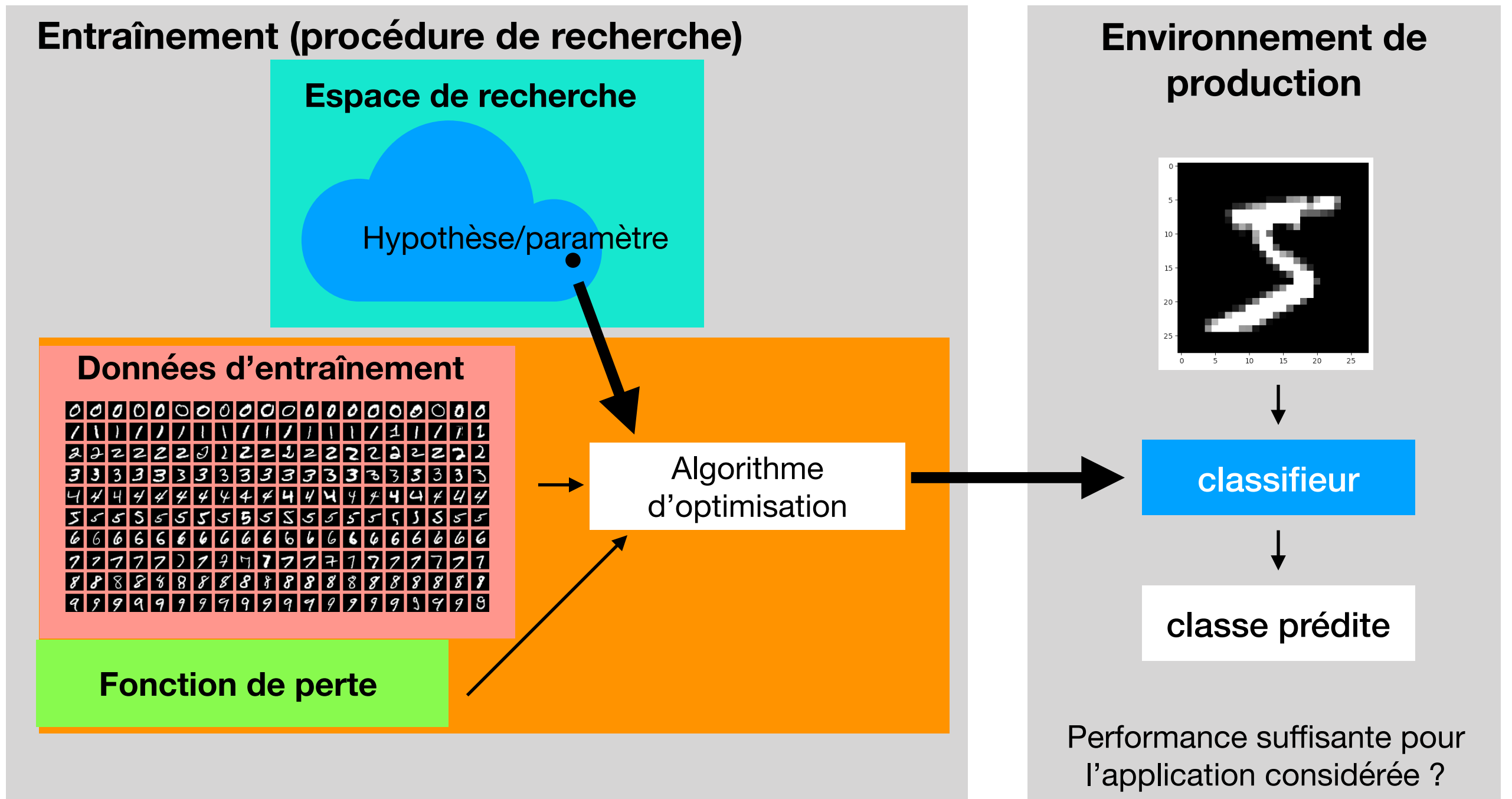
classifieur

classe prédite

Performance suffisante pour l'application considérée ?

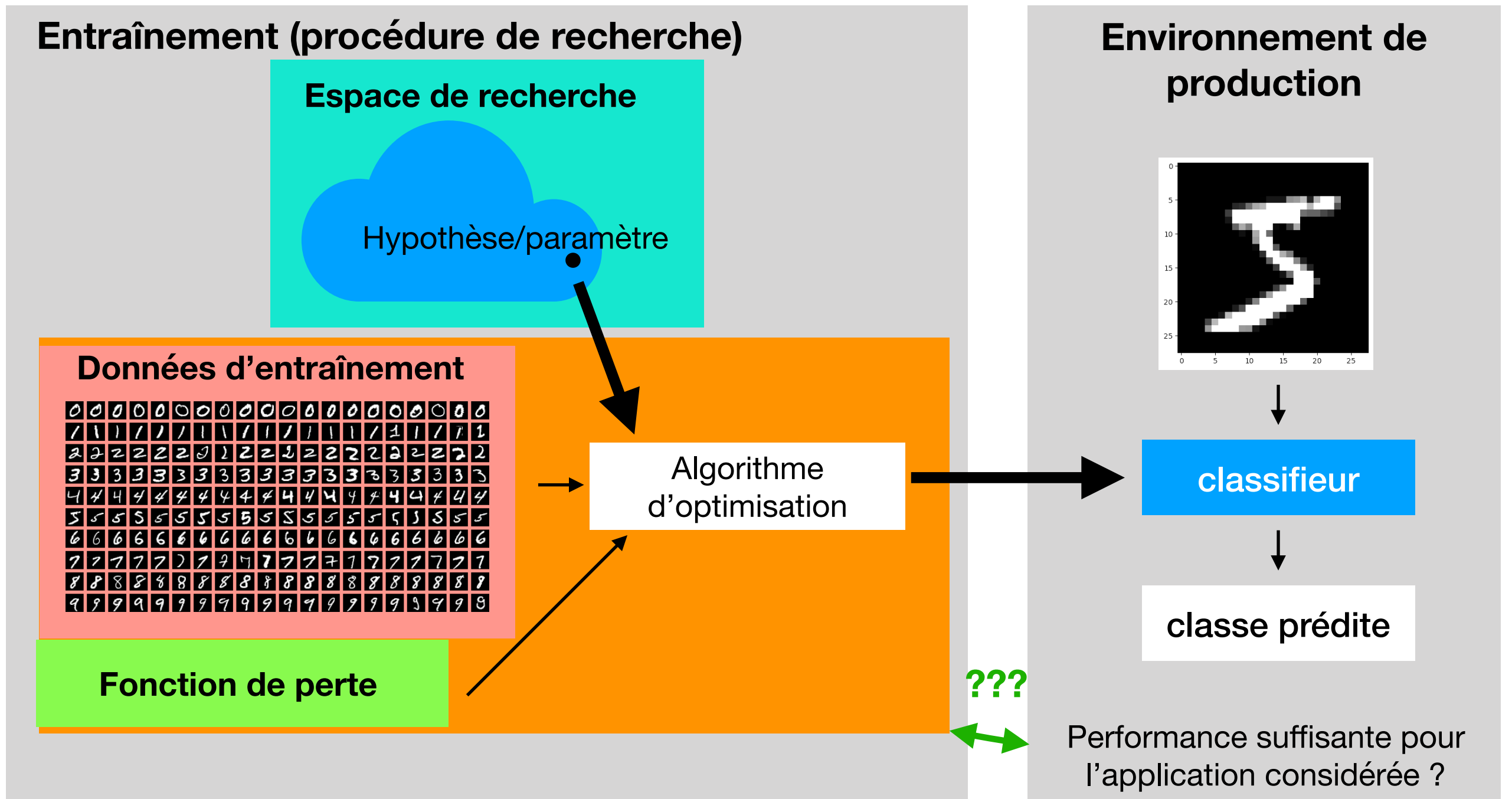
Introduction générale

Structure typique d'un problème d'apprentissage automatique



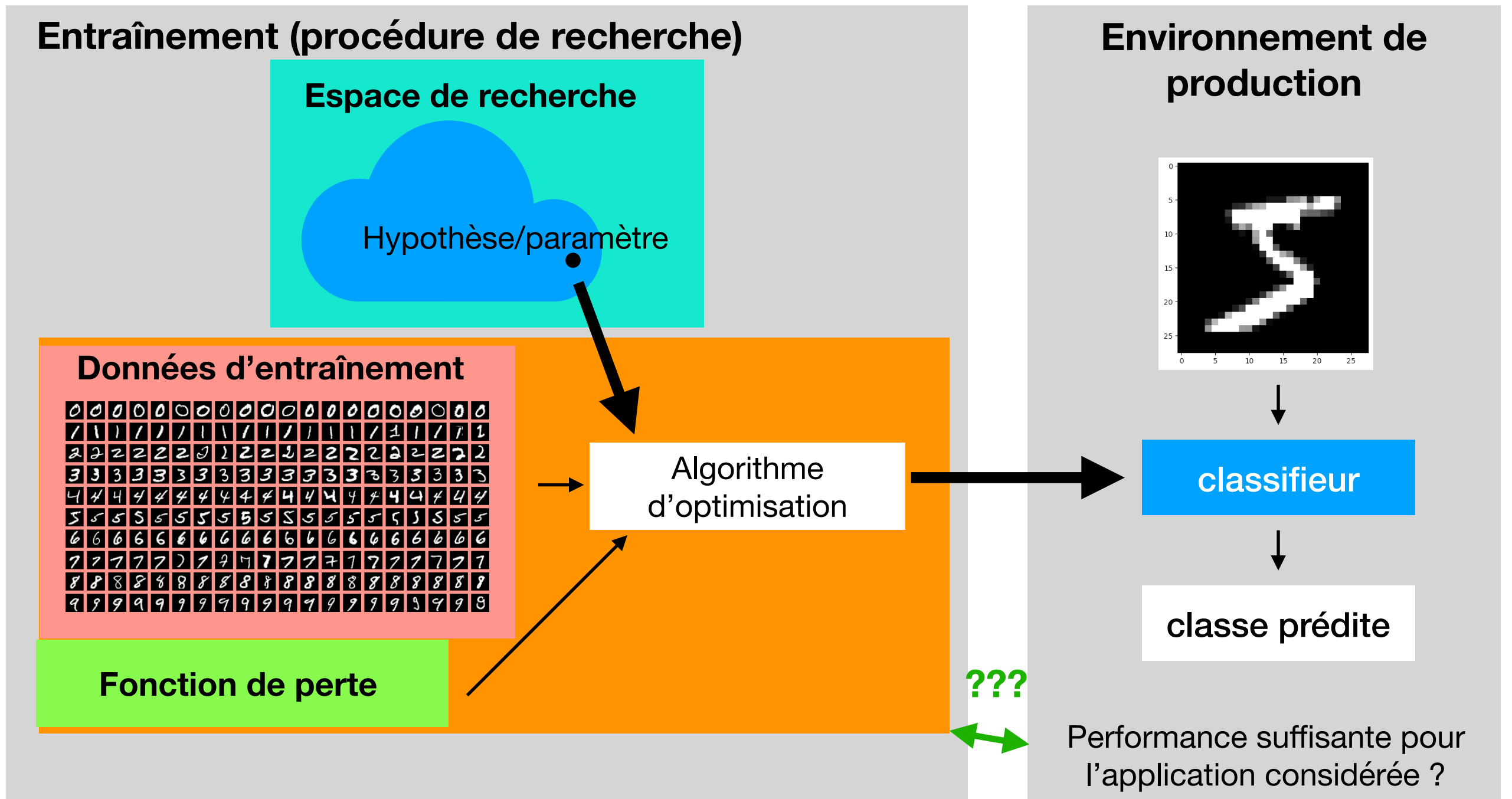
Introduction générale

Structure typique d'un problème d'apprentissage automatique



Introduction générale

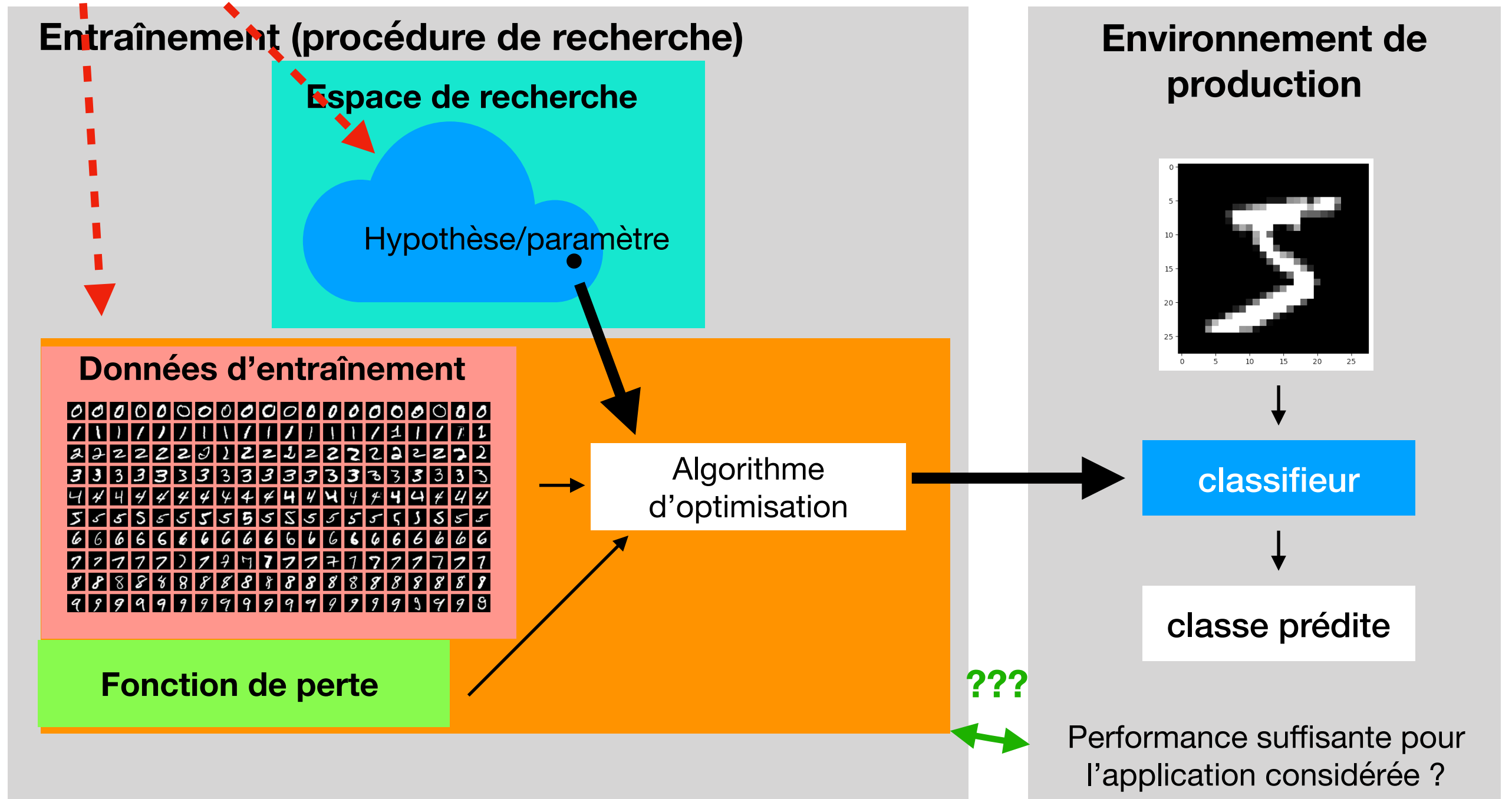
Structure typique d'un problème d'apprentissage automatique



Introduction générale

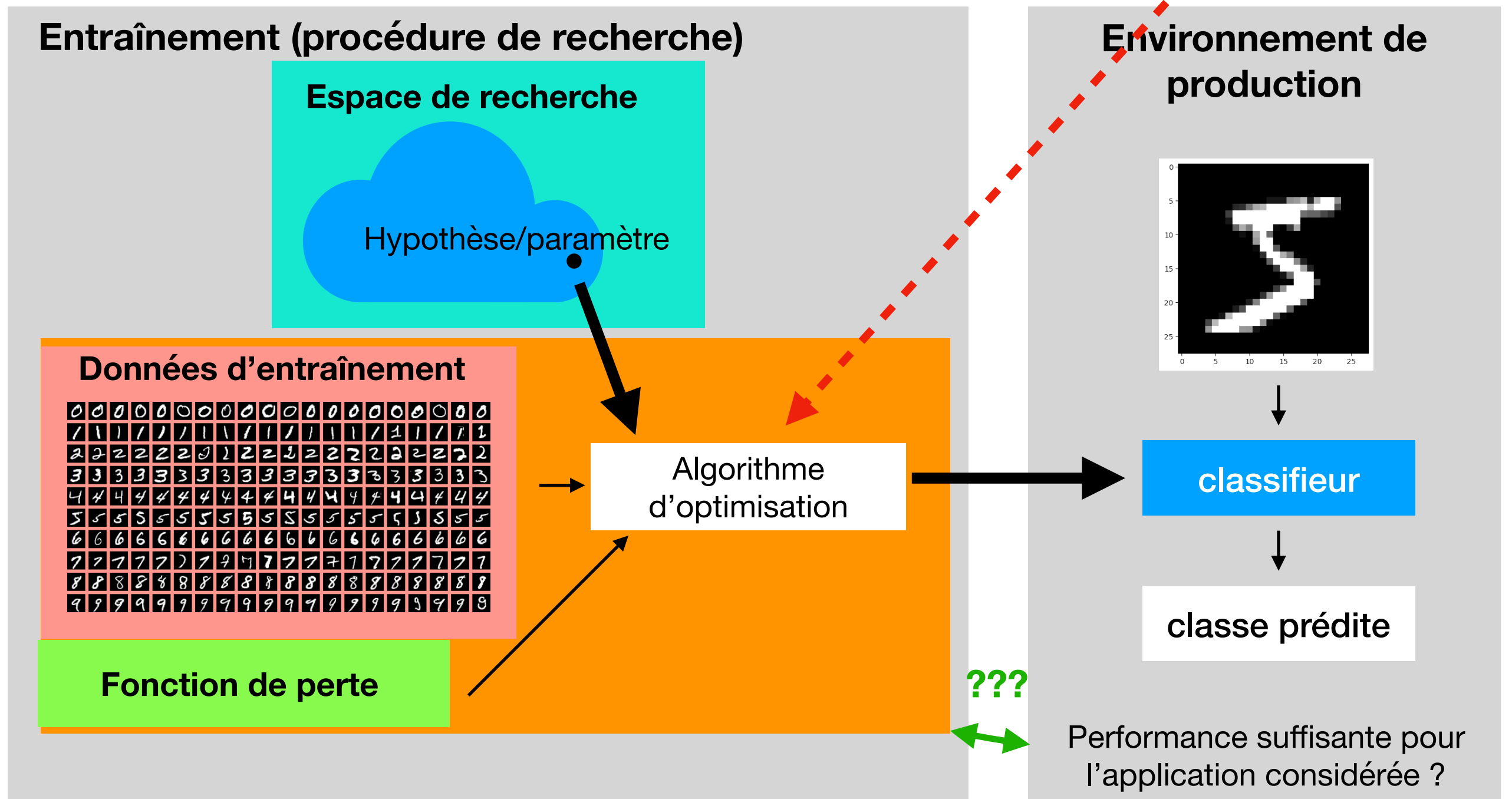
Algèbre linéaire

Structure typique d'un problème d'apprentissage automatique



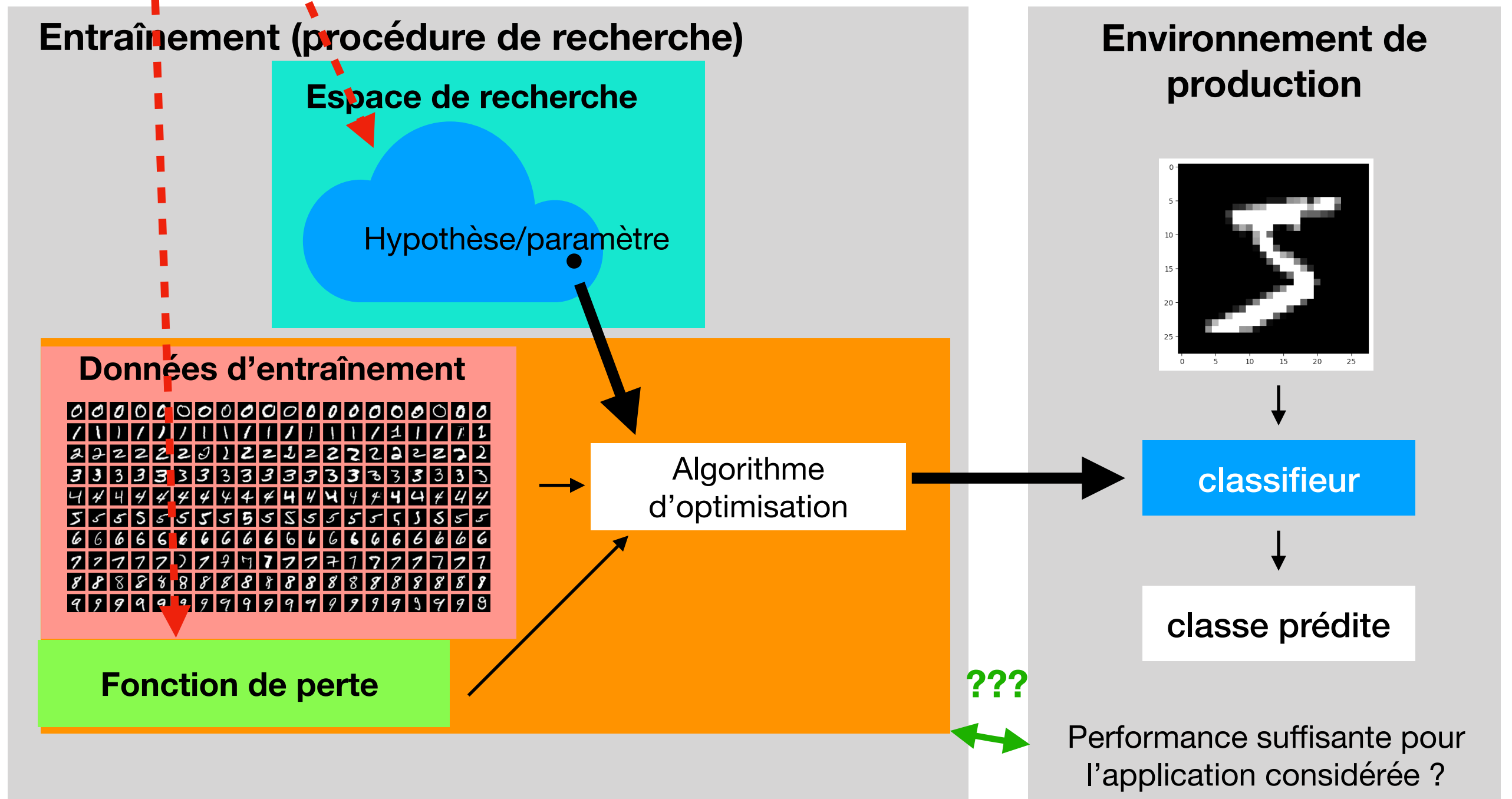
Introduction générale

Structure typique d'un problème d'apprentissage automatique



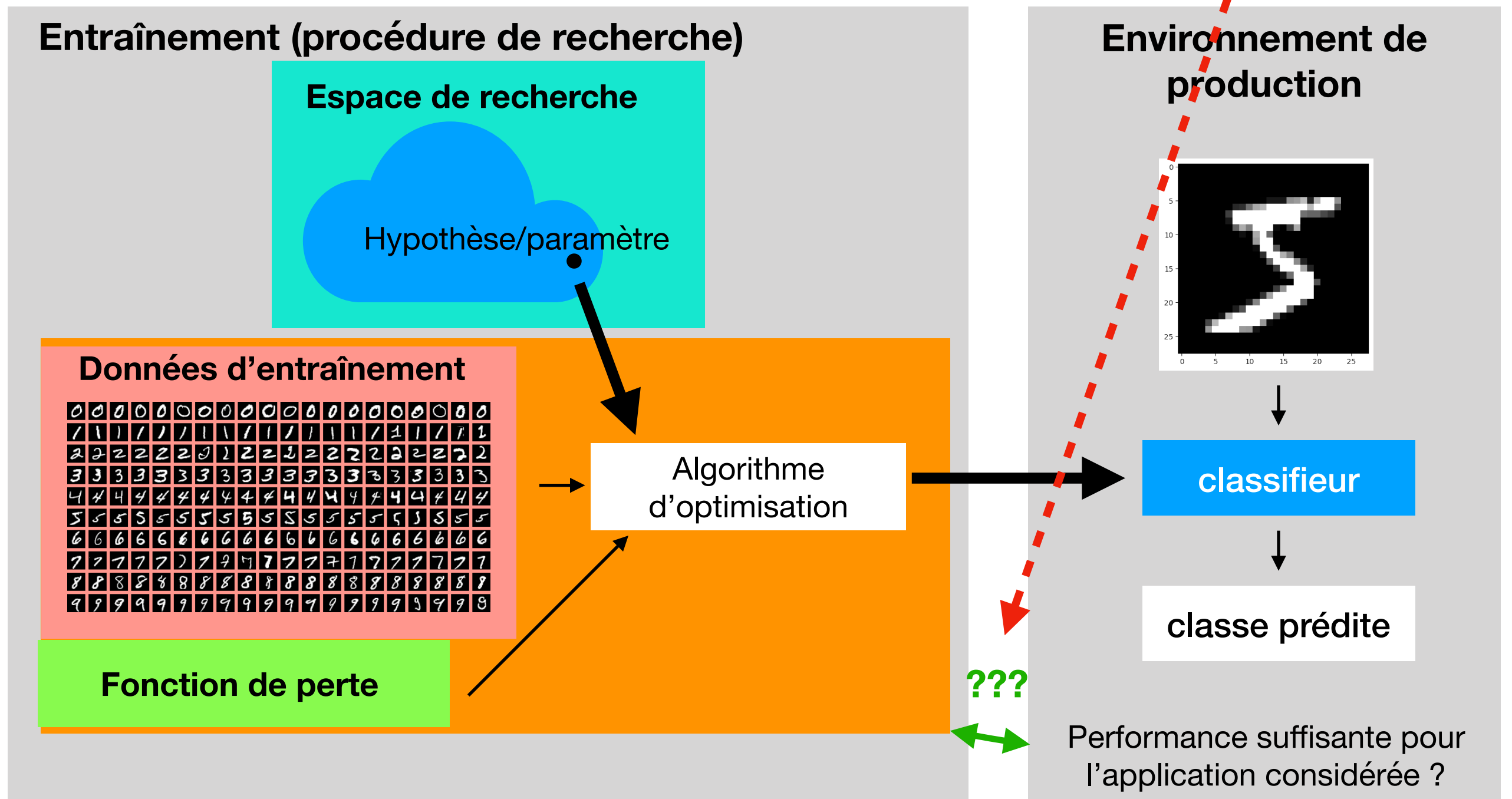
Introduction générale

Structure typique d'un problème d'apprentissage automatique



Introduction générale

Structure typique d'un problème d'apprentissage automatique



Algèbre linéaire

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières
- Théorème spectral, positivité, déterminant, trace

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières
- Théorème spectral, positivité, déterminant, trace

Interpretation des opérations matricielles

Multiplication matrice/vecteur colonne: $(Mv)_i := \sum_{j=1}^n m_{ij} v_j$

Multiplication matrice-matrice: $(AB)_{ij} := \sum_{k=1}^n a_{ik} b_{kj}$

Interpretation des opérations matricielles

Notation par bloc

La matrice

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \end{bmatrix}$$

peut être partitionnée en quatre blocs

$$\mathbf{P}_{11} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{P}_{12} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \mathbf{P}_{21} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}, \mathbf{P}_{22} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

On peut alors écrire la matrice par bloc comme :

$$\mathbf{P}_{\text{partitionnee}} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}.$$

Interpretation des opérations matricielles

write $A \in \mathbf{R}^{m \times n}$ in terms of its columns:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

then $y = Ax$ can be written as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

write A in terms of its rows:

$$A = \begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_n^T \end{bmatrix}$$

then $y = Ax$ can be written as

$$y = \begin{bmatrix} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_{m_{22}}^T x \end{bmatrix}$$

Interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

$$C = \begin{bmatrix} c_1 & \cdots & c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 & \cdots & Ab_p \end{bmatrix}$$

$$C = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$$

$$C = \sum_i a_i \tilde{b}_i^T$$

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières
- Théorème spectral, positivité, déterminant, trace

Produit scalaire usuel et norme euclidienne

$$(a \mid b) = \sum_{i=1}^n a_i b_i = a^T b = \|a\|_2 \|b\|_2 \cos(\theta)$$

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{(x \mid x)}$$

Angles et orthogonalité

Intérêt ?

Soit E un espace vectoriel de dimension finie n muni d'un produit scalaire $(\cdot | \cdot)$ et $V = (v_1, \dots, v_n)$ une base orthonormale de E . Alors, pour tout $v \in E$,

$$v = \sum_{i=1}^n (v | v_i) v_i.$$

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières
- Théorème spectral, positivité, déterminant, trace

Supposez que vous avez un réseau de neurone avec la matrice de poids

$$W = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

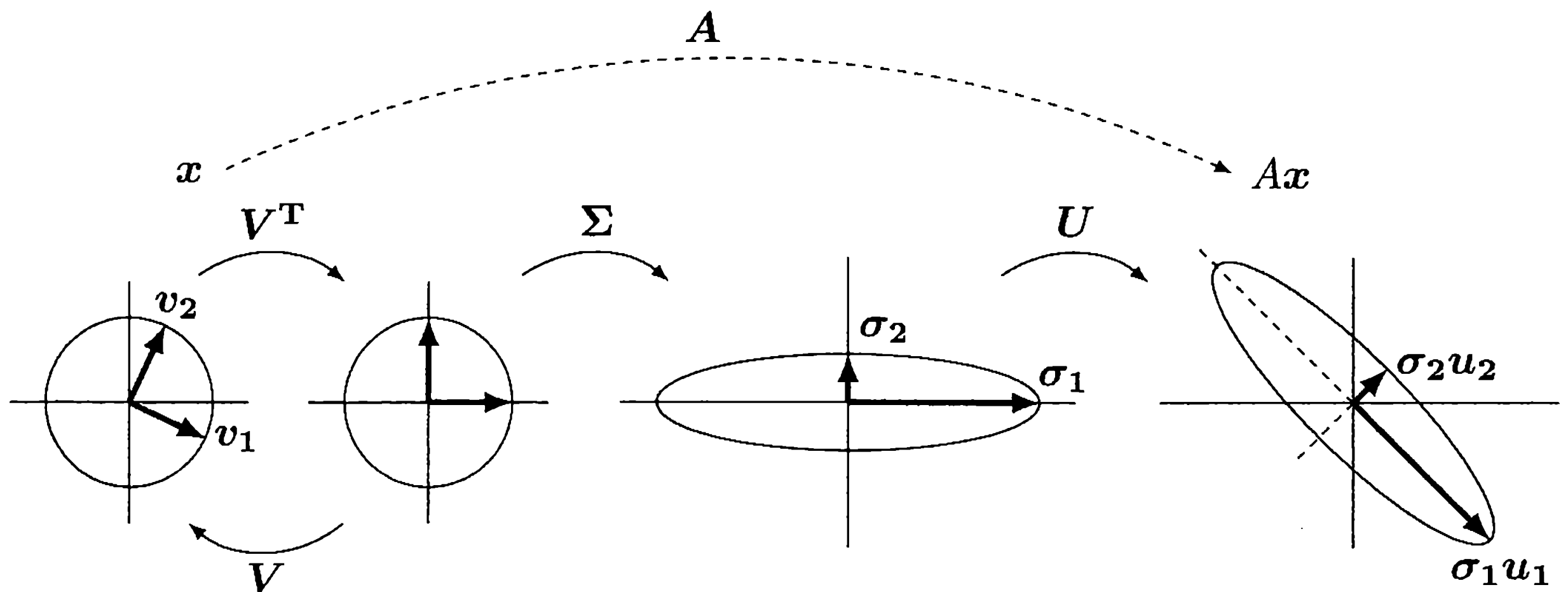
pour la couche L. Comment comprendre comment les activations de la couche L-1 sont transformées au niveau de la couche L dans ce réseau ?

Décomposition en valeurs singulières

$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & 0 & 0 \end{array} \right]$$

Décomposition en valeurs singulières

$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ \hline & & & 0 & \\ & & & & 0 \end{bmatrix}$$



Différentes visions des opérations matricielles : application

LoRA: Low-Rank Adaptation of Large Language Models

Edward Hu* Yelong Shen* Phillip Wallis
Zeyuan Allen-Zhu Yuanzhi Li Shean Wang Weizhu Chen
Microsoft Corporation
{edwardhu, yeshe, phwallis, zeyuana, yuanzhil, swang, wzchen}@microsoft.com

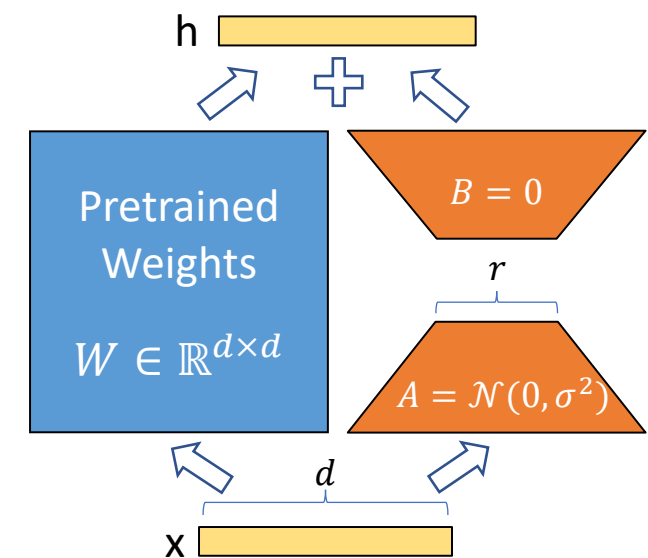


Figure 1: Our reparametrization. We only train A and B .

ICLR 2022, 18179 citations au 9/9/2025

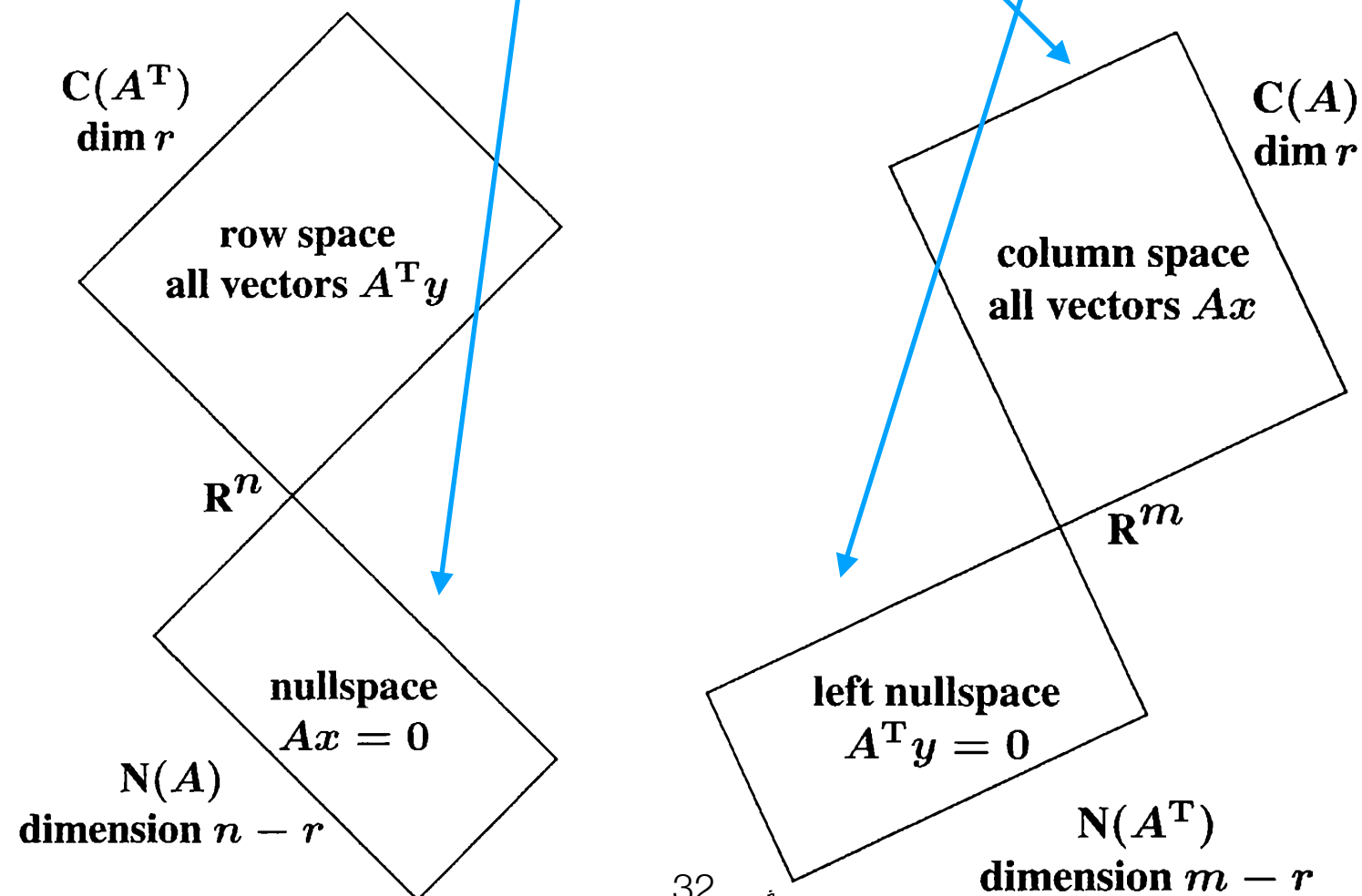
Adaptation *de rang faible* ? Que peut-on dire du rang de $\delta W = AB$?

Rappel: le rang d'une matrice est le plus grand nombre de vecteurs linéairement indépendants qu'on peut trouver au sein de l'ensemble des colonnes de la matrice (vues comme des vecteurs).

Espaces associés à une matrice

Bases orthonormées

$$AV = U\Sigma \quad A \begin{bmatrix} \underline{v_1 \dots v_r} & \underline{v_{r+1} \dots v_n} \end{bmatrix} = \begin{bmatrix} \underline{u_1 \dots u_r} & \underline{u_{r+1} \dots u_m} \end{bmatrix} \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & 0 & 0 \end{bmatrix}$$



Espaces associés à une matrice

$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & 0 & 0 \end{array} \right]$$

Par exemple : $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

$$U \approx \begin{pmatrix} -.21 & -.89 & .41 \\ -.52 & -.25 & -.82 \\ -.83 & .39 & .41 \end{pmatrix} \quad V \approx \begin{pmatrix} -.48 & -.57 & -.66 \\ .78 & .08 & -.62 \\ .41 & -.82 & .41 \end{pmatrix}$$

$$(\sigma_1 \quad \sigma_2 \quad \sigma_3) \approx (16.85 \quad 1.07 \quad 0)$$

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières et applications
- Théorème spectral et applications

Théorème spectral

Si S est une matrice symétrique, réelle de taille $m \times m$, alors il existe une matrice orthogonale réelle Q de taille $m \times m$ et une matrice diagonale réelle Λ de taille $m \times m$ telles que $S = Q\Lambda Q^T$.

Positivité

Une matrice symétrique réelle est dite définie positive, noté $S \succ 0$ ssi pour toute matrice colonne u , $u^T S u > 0$.

Une matrice symétrique réelle est dite semi-définie positive, $S \succeq 0$, ssi pour toute matrice colonne u , $u^T S u \geq 0$.

Relation d'ordre sur les matrices symétriques réelles:

$$S_1 \prec S_2 \text{ ssi } S_2 - S_1 \succ 0$$

et

$$S_1 \preceq S_2 \text{ ssi } S_2 - S_1 \succeq 0$$

Déterminant et trace

A matrice carrée $n \times n$

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i, \sigma_i} \right)$$

$\operatorname{sgn}(\sigma)$ est la parité du nombre d'éléments dans une décomposition de σ en une séquence de transpositions (échange de deux éléments).

$$\det(AB) = \det(A) \det(B)$$

$$\operatorname{Tr}(A) = \sum_{i=1}^n a_{i,i} \quad \operatorname{Tr}(AB) = \operatorname{Tr}(BA)$$

$$\operatorname{Tr}(A_1 A_2 \dots A_k) = \operatorname{Tr}(A_2 A_3 \dots A_k A_1) = \dots = \operatorname{Tr}(A_k A_1 A_2 \dots A_{k-1})$$

Optimisation

Optimisation

$$f : \mathcal{D} = \text{dom}(f) \subset \mathbf{R}^n \rightarrow \mathbf{R}$$
$$\min_x f(x) ?$$

- Minimisation d'une fonction sur un domaine
 - Domaine “ouvert” ou non ?
 - Domaine et fonction convexes ?
 - Fonction lisse ou non ?

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Algorithmes d'optimisation
4. Analyse convexe et dualité Lagrangienne
5. Ouverture vers l'optimisation non-lisse

Définitions

$$f : \mathcal{D} = \text{dom}(f) \subset \mathbf{R}^n \rightarrow \mathbf{R}$$

$$\min_x f(x) \text{ ?}$$

x^* est un minimum global de f ssi pour tout $x \in \text{dom}(f)$, $f(x^*) \leq f(x)$

x^* est un minimum local de f ssi il existe un ensemble ouvert $U \subset \text{dom}(f)$ contenant x^* tel que pour tout $x \in U$, $f(x^*) \leq f(x)$

Notions de base de topologie

Dans $\mathbf{R}^n$, La boule fermée $B_f(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 \leq \epsilon$

Dans $\mathbf{R}^n$, La boule ouverte $B_o(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 < \epsilon$

U est un ouvert de $\mathbf{R}^n$ ssi $U \subset \mathbf{R}^n$ et pour tout x dans U , il existe $\epsilon > 0$, tel que $B_f(x, \epsilon) \subset U$.

U est un fermé de $\mathbf{R}^n$ ssi son complément dans $\mathbf{R}^n$ est un ouvert de $\mathbf{R}^n$