

Apprendre *Connect4*

Alexandre Adam - Éric Pfléiderer
MAT3450 - Modélisation mathématique

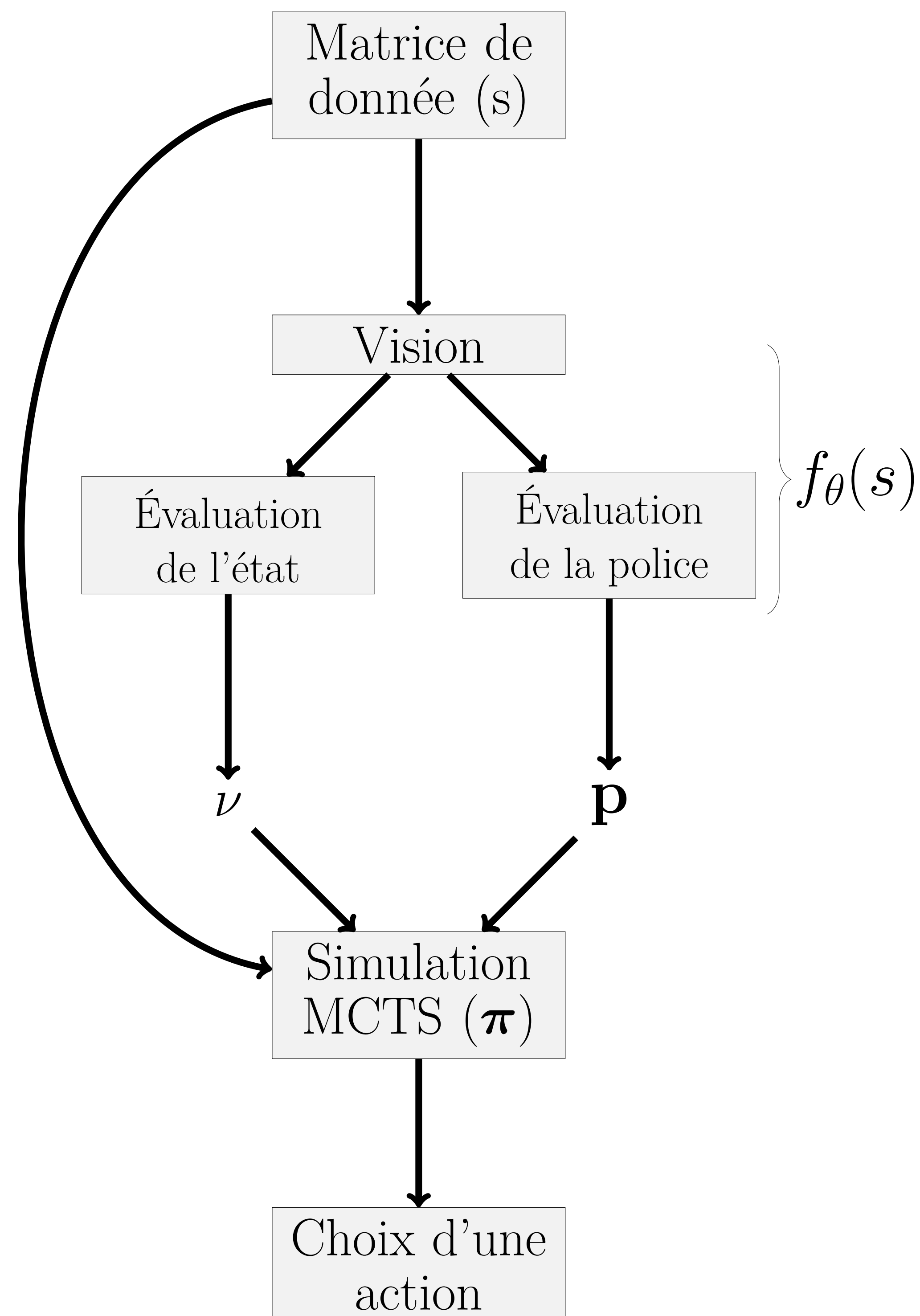
Objectifs pour aujourd'hui

- Introduction au vocabulaire et aux concepts fondamentaux de l'apprentissage machine
- Recherche proposée : *AlphaGo Zero* et *Connect4*
- Méthodologie et concept expérimental
- Résultats

Vocabulaire et concepts

- Déterminisme et probabilisme
- Réseau neuronal
- Vision machine : convolutionnel et résiduel
- Apprentissage par renforcement
- Recherche arborescente Monte Carlo
- Descente du gradient

Schéma conceptuel du modèle



Légende

θ : paramètres du réseau neuronal
 \mathbf{p} : vecteur de probabilités (réseau)
 ν : probabilité estimée de victoire
 $\boldsymbol{\pi}$: vecteur de probabilités (MCTS)
 l : fonction de perte
 z : vainqueur (± 1)

Réseau de neurones

$$f_{\theta}(s) = (\mathbf{p}, \nu)$$

$$l = (z - \nu)^2 - \boldsymbol{\pi}^T \log \mathbf{p} + c \|\theta\|^2$$

Simulation MCTS

$$\text{Statistiques d'un noeud} \begin{cases} N(s_t, a_t) = N(s_t, a_t) + 1 \\ W(s_t, a_t) = W(s_t, a_t) + \nu \\ Q(s_t, a_t) = \frac{W(s_t, a_t)}{N(s_t, a_t)} \end{cases}$$

$$\text{Évaluation d'une branche} \begin{cases} P(s, a) = (1 - \epsilon) \mathbf{p}_{\alpha} + \epsilon \eta_{\alpha} \end{cases}$$

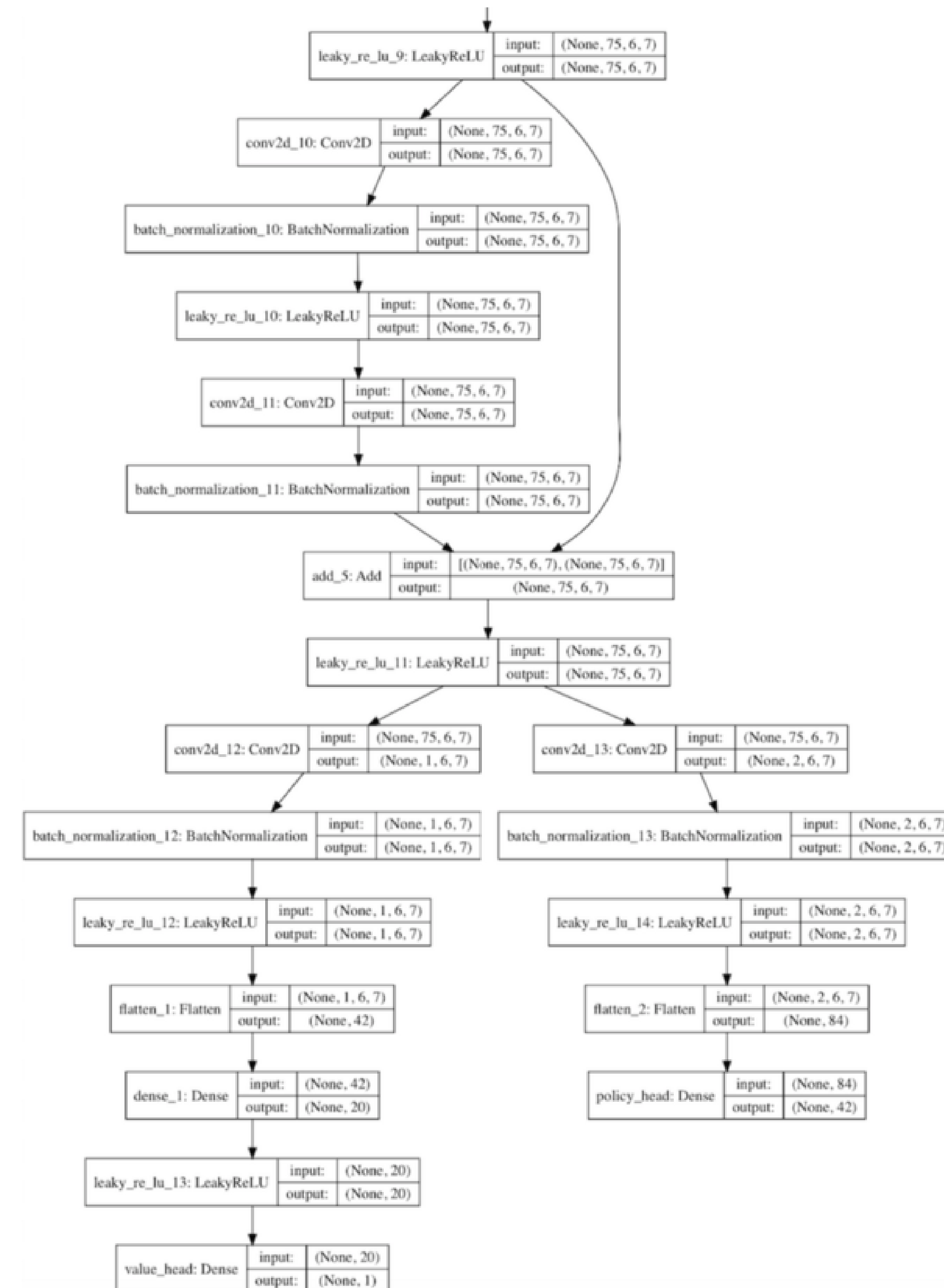
Choix d'une action

$$a_t = \operatorname{argmax} \left[Q(s, a) + c_{puct} P(s, a) \sqrt{\frac{\varepsilon_b N(s, b)}{1 + N(s, a)}} \right]$$

Paramètres

$$\begin{cases} \text{Poids du bruit} & \epsilon = 0.3 \\ \text{Degré d'exploration} & c_{puct} = 1 \\ \text{Bruit stochastique} & \eta \approx \text{Dir}(0.3) \end{cases}$$

Détails de $f_{\theta}(s)$



Compétitions intra architecture

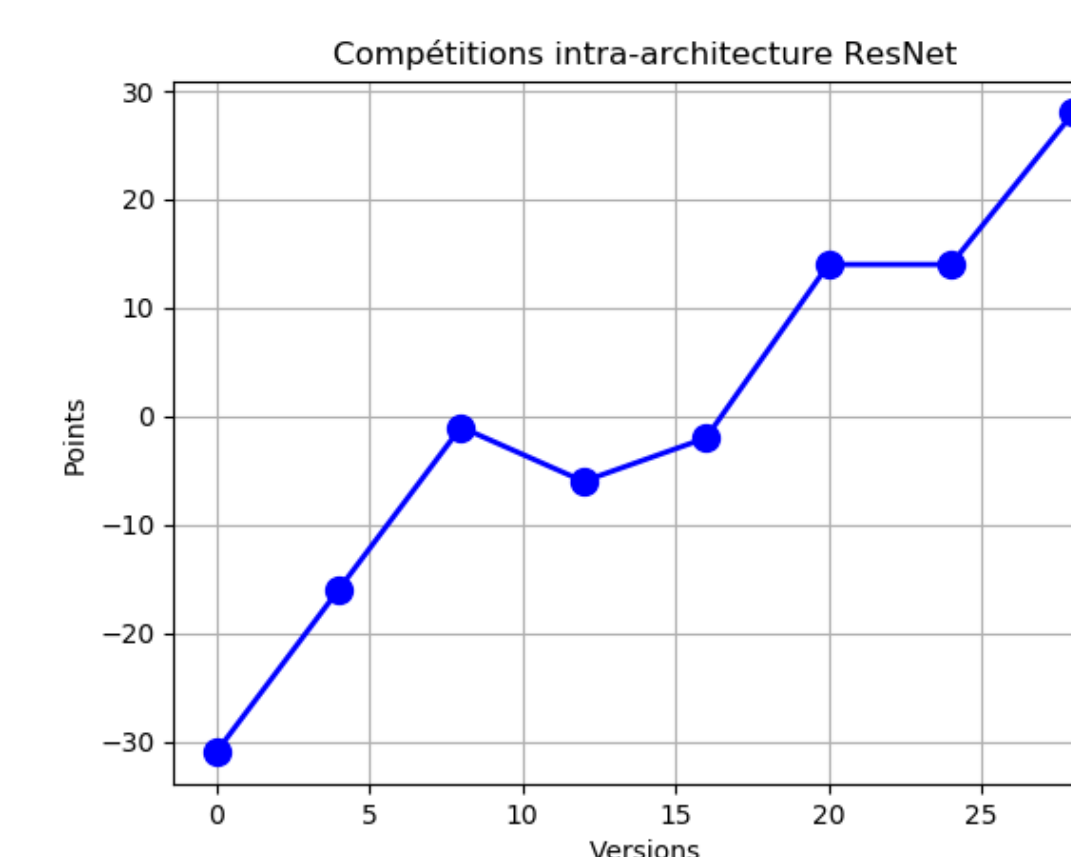


FIGURE – Résultats de la compétition intra architecture ResNet

Techniques et tactiques développées

0	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	32	33	34
35	36	37	38	39	40	41

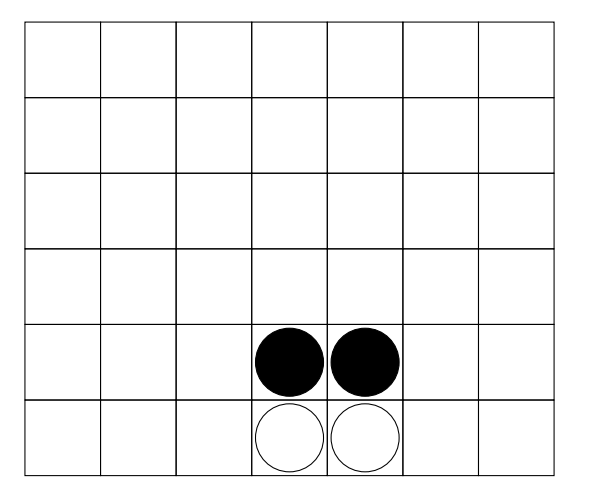
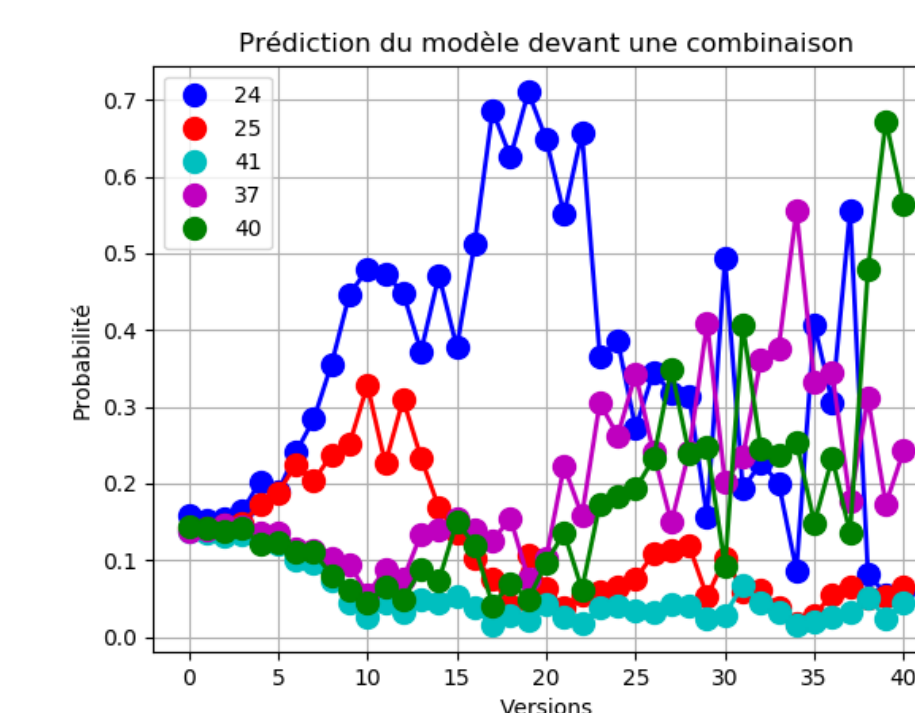


FIGURE – Évaluation de la position qui émerge après les coups 1. d1 d2, 2. e1 e2. La probabilité du coup 3. d3 (bleue) est dominante jusqu'à la version 20. Les coups 3. c1 (mauve) et 3. f1 (vert) sont progressivement préférés après la version 20.

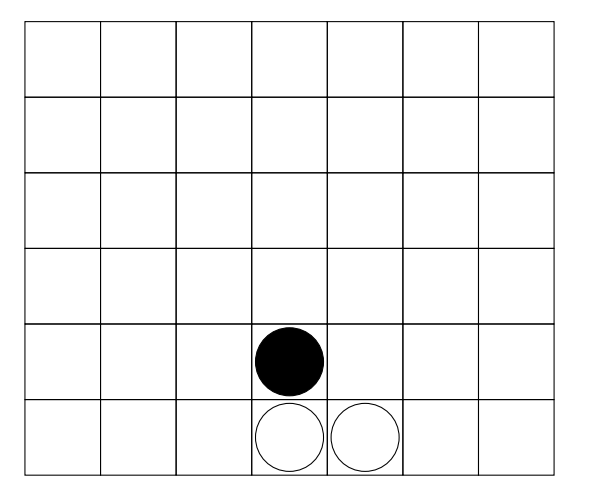
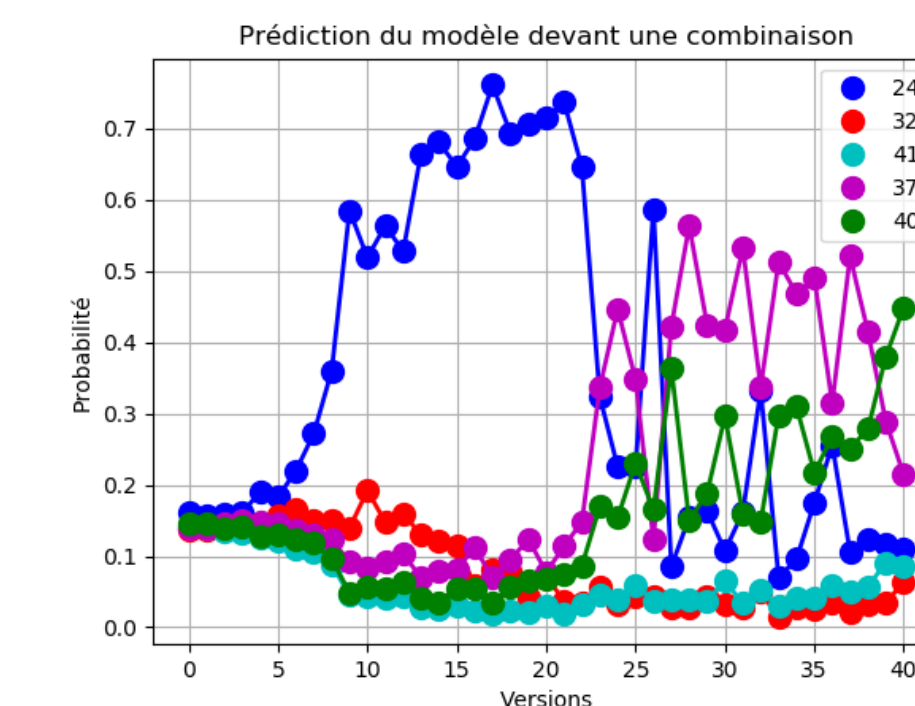


FIGURE – Évaluation de la position qui émerge après les coups 1. d1 d2, 2. e1. La probabilité du coup 2. ...d3 (bleue) est dominante jusqu'à la version 20. Les coups 2. ...c1 (mauve) et 2. ...f1 (vert) sont préférés après la version 20.

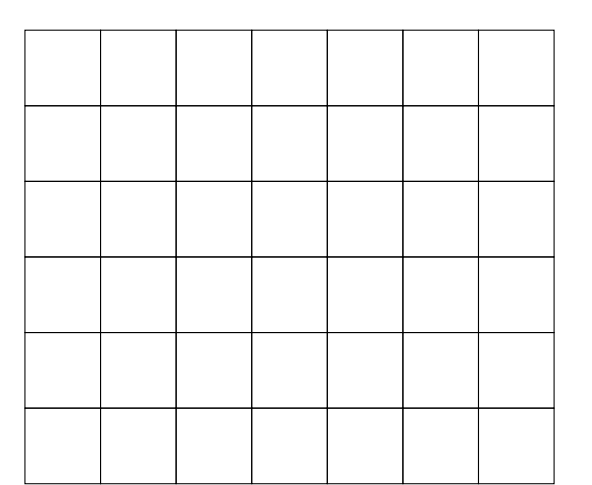
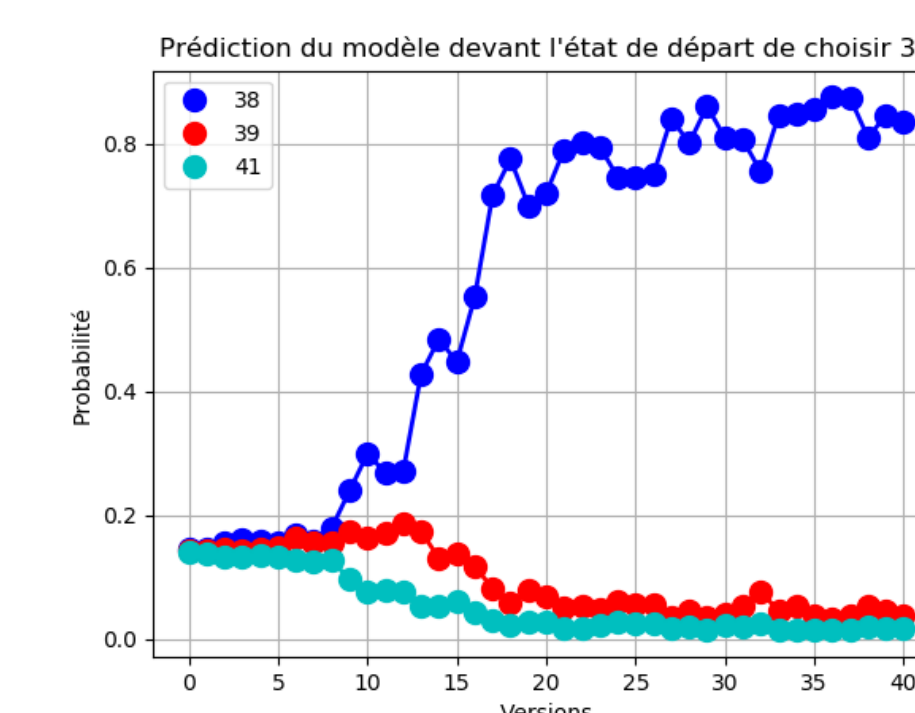


FIGURE – Évaluation de la position initiale. La probabilité du coup 38 (bleue) devient dominante autour de la version 10. La probabilité de jouer les autres coups converge vers 0.