# Projet Deep learning (No free lunch) Groupe 2

Erwan Morcet, Maxime Boidin

## Policy Iteration

## Value Iteration

## Monte Carlo

Première remarque : pour les 2 jeux, On-policy est beaucoup plus rapide que off-policy, la mise à jour de la policy directement dans la boucle des simulations retire une boucle d’itération complète ce qui est un gain de temps considérable, on a plus besoin de parcourir l’ensemble des states à chaque boucle pour mettre à jour la policy.

### Grid World

* + Petit 4x4

Off Policy : 200 itérations avec 500 simulations, 100 mouvements par simulation.

Exécuté 22,4928973s

On Policy : 500 simulations, 100 mouvements par simulation.

Exécuté en 0,4370933s.

* + Grand 8x8

Off Policy : 50 itérations avec 1200 simulations, 100 mouvements par simulation.

On Policy : 1200 simulations, 100 mouvements par simulation.

### Sokoban

* + Petit 3x3 avec 2 caisses « complexe » (plus de 2 mouvements)

Off Policy : 100 itérations avec 500 simulations, 100 mouvements par simulation.

Exécuté en 26,2784549s

On Policy : 500 simulations, 100 mouvements par simulation.

Exécuté en 0,7092018s.

## Sarsa

Simple à implémenter comparé à monte Carlo.

### Grid World

* + Petit 4x4

On Policy : 500 simulations, 100 mouvements par simulation. 0.5 pour alpha et 0.1 pour gamma.

Exécuté en 0,007065s.

Grand 8x8

On Policy : 1200 simulations, 100 mouvements par simulation.

Exécuté en 2,12933s

### Sokoban

* + Petit 3x3 avec 2 caisses « complexe » (plus de 2 mouvements)

1200 simulations, 100 mouvements par simulation. 0.5 pour alpha et gamma.

Grand 8x8

700 simulations, 150 mouvements par simulation. 0.5 pour alpha et gamma.

Exécuté en 0,9551858s.

## Qlearning

Peu de différence avec Sarsa, a part une optimisation au moment de l’update de la policy.

### Grid World

* + Petit 4x4

500 simulations, 100 mouvements par simulation. 0.5 pour alpha et 0.1 pour gamma.

Exécuté en 0,0167684s.

Grand 8x8

1200 simulations, 100 mouvements par simulation.

Exécuté en 2,7580465s

### Sokoban

* + Petit 3x3 avec 2 caisses « complexe » (plus de 2 mouvements)

1200 simulations, 100 mouvements par simulation. 0.5 pour alpha et gamma.

Grand 8x8

700 simulations, 150 mouvements par simulation. 0.5 pour alpha et gamma.

Exécuté en 1,3526107s.

L’implémentation de Qlearning et Sarsa est bien plus simple que MonteCarlo pour des temps d’exécution souvent plus rapide dans le cadre de gridworld et sokoban, on peut les rendre d’autant plus rapide en ajustant le nombre d’itération, et les paramètres alpha et gamma.