

Travail Pratique 2 (IFT-4201/IFT-7201)

Présenté à Audrey Durand

Équipe 10 : Adam Cohen, Maxime Genest, Vincent Masse

1. (a)

(b) **Réseau cible**

L'oubli catastrophique est le fait qu'un réseau de neurones oublie son apprentissage lorsqu'il doit apprendre une nouvelle tâche, c'est un peu comme s'il recommençait son apprentissage à 0. Le fait de fixer les poids θ dans la cible pendant plusieurs mises à jours permet de conserver une certaine stabilité pour permettre la convergence.

(c) **Heuristique**

Si τ est trop grand, la cible est trop modifiée, ce qui risque de compromettre la convergence (voir numéro précédent), le cas limite $\tau = 1$ revient à ne pas fixer (même pas partiellement) les targets. Si τ est trop petit, les targets seront trop fixes, et donc les cibles ne sont pratiquement pas améliorées et le réseau apprend donc sur des moins bonnes cibles (non-représentative de ce que l'on doit viser). Le cas limite $\tau = 0$ montre un cas où la cible n'est jamais modifiée et gardera toujours sa valeur initiale, ce qui est bien entendu non-souhaitable.

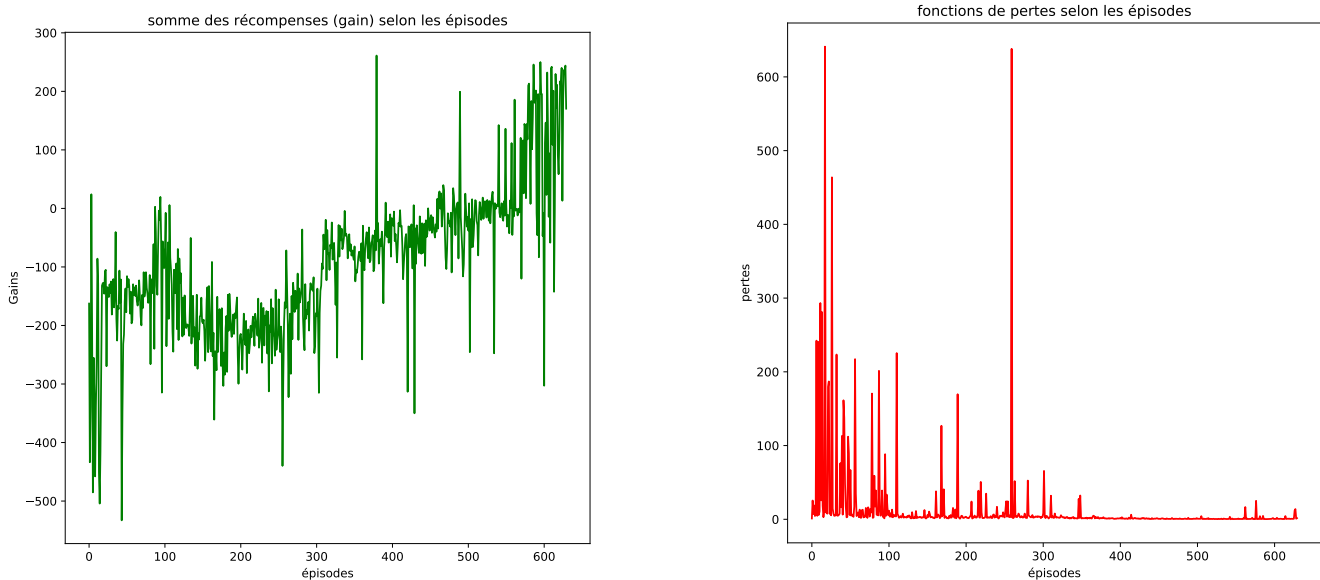
(d) **Replay buffer**

L'avantage de son utilisation est de briser la corrélation entre les échantillons utilisés lors de la mise à jour (le tuple $(s_t, a_t, r_{t+1}, s_{t+1})$ et le tuple suivant $(s_{t+1}, a_{t+1}, r_{t+2}, s_{t+2})$ partage le même état s_{t+1}). Cela compromet l'apprentissage. Le fait de stocker les échantillons et d'échantillonner une mini-batch plusieurs fois au cours de l'épisode permet contourner ce problème. À COMPLÉTER

(e) Apprentissage supervisée

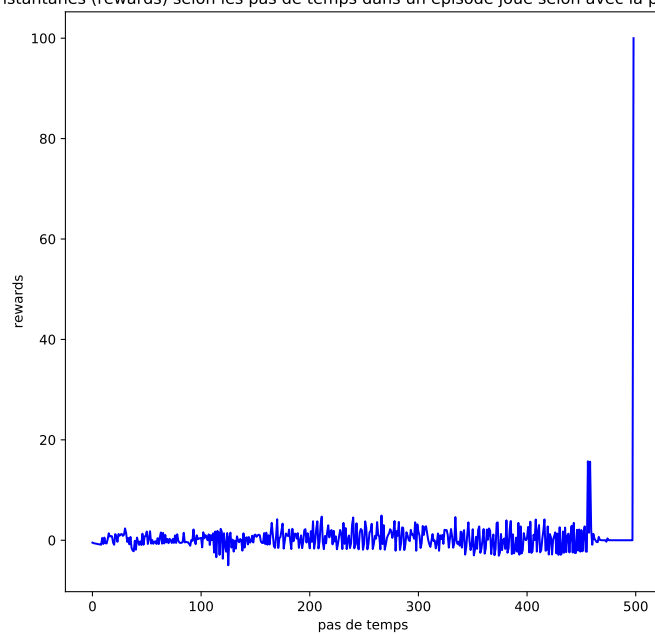
(f) Graphiques...

FIGURE 1 – Évolution des gains et fonction de pertes



FONCTION DE PERTE CI-DESSUS À MODIFIER (Je ne suis pas certain que c'est ça qu'il est demandé)

FIGURE 2 –
gains instantanés (rewards) selon les pas de temps dans un épisode joué selon avec la politique apprise



2. No 2

(a)

(b)