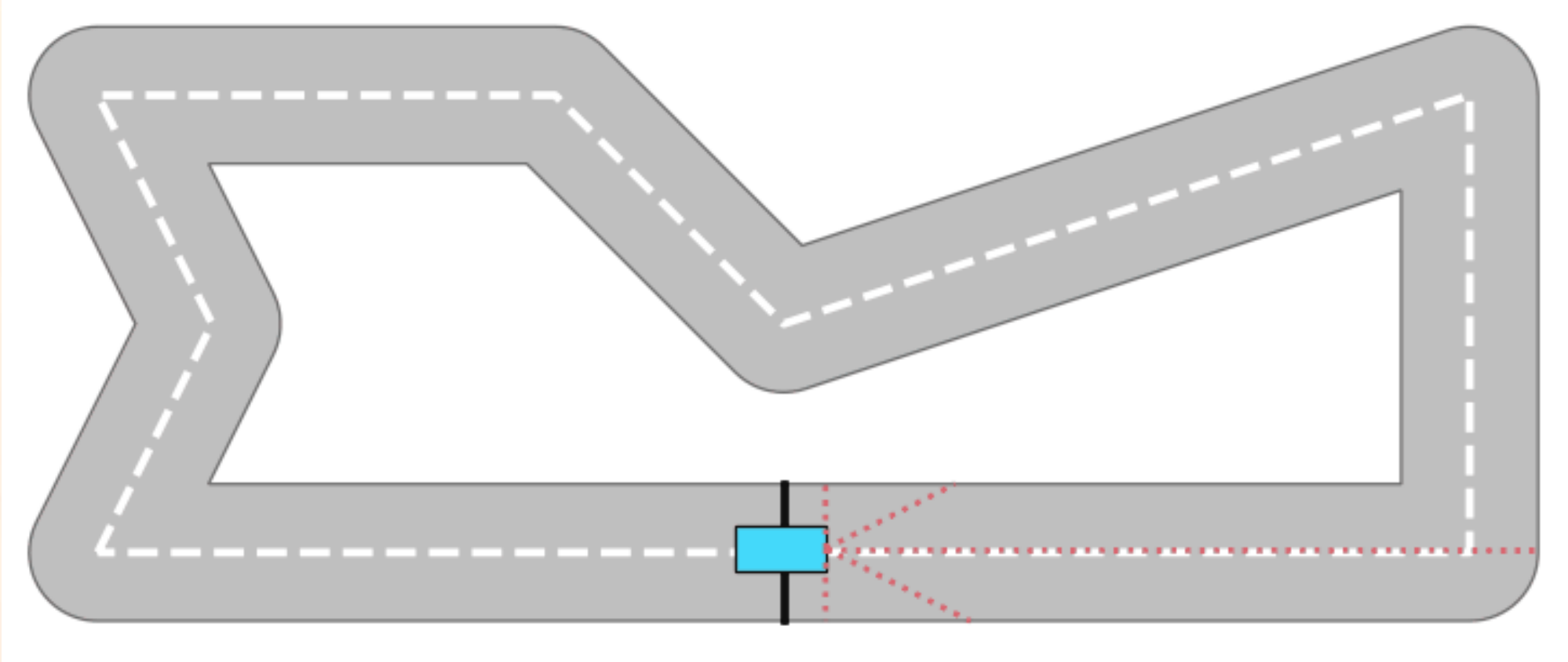
DeepRL

Vous avez dit réseaux de neurones ?

Etats/actions continues !



Espaces infinis

Relation d’ordres

Des approximateurs !

Avec des *fonctions paramétriques*,
on va approximer ces fonctions :

$$q_{\pi,w}(x(s),a) \simeq q_\pi(s,a)$$
$$v_{\pi,w}(x(s)) \simeq v_\pi(s)$$
$$\pi_\theta(a|s) \simeq \pi(a|s)$$

Entrainer ces approximateurs : Critique

$$datas = \{x(S_t), A_t, R_t, x(S_{t+1}), D_t, I_t\}$$

$X = \{x(S_t), A_t\}$

$X = \{x(S_t)\}$

Modèle

$$q_{\pi,w}(s,a)$$

$$\mathcal{L} = \mathbb{E}\left(\left(G_t - \hat{G}_t[A_t]\right)^2\right)$$

$$\nabla_w \mathcal{L} \propto \mathbb{E}\left(\left(G_t - \hat{G}_t[A_t]\right) * \nabla_w q_{\pi,w}(x(S_t))\right)$$

$\hat{G}_t$

$\hat{G}_t[A_t]$

$$G_t = \{R_t + \hat{G}_{t+1}\}$$

$$\hat{G}_{t+1} = (1 - D_t) \max_a(q_{\pi,w}(x(S_{t+1}), a))$$

$$\hat{G}_{t+1} = (1 - D_t) \max_a(\hat{G}_t[a])$$

$$\hat{G}_{t+1} = (1 - D_t) \sum_a \mu(S_{t+1}, a) q_{\pi,w}(S_{t+1}, a)$$

$Sarsa\ TD(0): Off - policy$

Deep Q-learning (DQN)

$$datas = \{x(S_t), A_t, R_t, x(S_{t+1}), D_t, I_t\}$$

$$X = \{x(S_t)\}$$

Modèle

$$q_{\pi,w}(s,a)$$

$$\mathcal{L} = \mathbb{E}\left(\left(G_t - \hat{G}_t[A_t]\right)^2\right)$$

$$\hat{G}_t[A_t] \qquad G_t = \{R_t + \hat{G}_{t+1}\}$$

$$\hat{G}_{t+1} = (1 - D_t)\max_a(q_{\pi,w}(x(S_{t+1}), a))$$

Stabilisation

$$\hat{G}_{t+1} = (1 - D_t)\max_a(q_{\pi,w_{targ}}(x(S_{t+1}), a))$$

Freezed Q-Learning

Polyak averaging $p \in [0,1]$

$$w_{targ} = (1 - p)w_{targ} + p * w$$