Quiz : Compréhension de PPO et DQL

Instructions

Répondez aux questions suivantes en sélectionnant la meilleure réponse. Chaque question teste votre compréhension des concepts fondamentaux de **Proximal Policy Optimization (PPO)** et **Deep Q-Learning (DQL)**.

—

Questions

Question 1 : Objectifs de PPO et DQL Quelle est la différence fondamentale entre PPO et DQL?

1. PPO utilise une politique probabiliste tandis que DQL estime les valeurs $Q(s, a)$.
2. DQL est conçu pour les environnements continus, tandis que PPO est conçu pour les environnements discrets.
3. PPO est basé sur les valeurs et DQL est basé sur les politiques.
4. DQL est plus récent et plus stable que PPO.

—

Question 2 : Exploration dans PPO Comment PPO encourage-t-il l'exploration?

1. En ajoutant un terme d'entropie dans la fonction objectif.
2. En utilisant un réseau Q avec un facteur ϵ-greedy.
3. En ajustant les politiques de manière agressive avec clipping.
4. En maximisant directement la récompense immédiate.

—

Question 3 : Applications typiques Dans quel contexte DQL est-il le plus efficace?

1. Contrôle continu de robots.
2. Jeux avec des actions discrètes comme Pac-Man ou Space Invaders.

3. Gestion des ressources dans des data centers.

4. Trading algorithmique avec des décisions continues.

—

Question 4 : Rôle du clipping dans PPO Pourquoi PPO utilise-t-il le clipping dans sa fonction objectif?

1. Pour limiter les mises à jour excessives de la politique.

2. Pour forcer l'agent à explorer des actions rares.

3. Pour améliorer la précision des valeurs $Q(s, a)$.

4. Pour réduire la complexité computationnelle de l'algorithme.

—

Question 5 : Environnements continus Quel algorithme est le mieux adapté aux environnements continus?

1. Deep Q-Learning (DQL).

2. Proximal Policy Optimization (PPO).

3. Q-Learning traditionnel.

4. Aucune des réponses ci-dessus.

—

Question 6 : Réseaux utilisés dans PPO Quels réseaux de neurones PPO utilise-t-il?

1. Un réseau Q pour estimer les valeurs $Q(s, a)$.

2. Un réseau de politique et un réseau de valeur.

3. Deux réseaux Q séparés pour l'exploration et l'exploitation.

4. Aucun réseau de neurones n'est utilisé.

—

Question 7 : Objectif général des algorithmes d'apprentissage par renforcement Quel est l'objectif principal des algorithmes d'apprentissage par renforcement comme PPO et DQL?

1. Maximiser la récompense cumulée sur le long terme.

2. Minimiser les erreurs entre les états prédits et les valeurs réelles.

3. Maintenir une politique constante tout au long de l'entraînement.

4. Explorer toutes les actions possibles de manière égale.

—

Question 8 : Avantage dans PPO Quelle est l'utilité de l'avantage $A(s_t, a_t)$ dans PPO?

1. Comparer la nouvelle politique avec l'ancienne.
2. Évaluer la qualité d'une action par rapport à la valeur actuelle.
3. Calculer directement la récompense cumulative.
4. Remplacer les valeurs $Q(s, a)$ dans la fonction objectif.

—

Question 9 : Fonction objectif dans DQL Quelle est la principale fonction objectif dans DQL?

1. Minimiser l'erreur quadratique entre $Q(s, a)$ et sa cible.
2. Maximiser les probabilités d'actions optimales.
3. Maximiser l'entropie de la politique.
4. Réduire le ratio $r_t(\theta)$ entre anciennes et nouvelles politiques.

—

Question 10 : Ratio dans PPO Le ratio $r_t(\theta)$ dans PPO est défini comme :

1. $r_t(\theta) = \frac{\pi_{\text{old}}(a_t|s_t)}{\pi_\theta(a_t|s_t)}$.
2. $r_t(\theta) = \pi_\theta(a_t|s_t) - \pi_{\text{old}}(a_t|s_t)$.
3. $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\text{old}}(a_t|s_t)}$.
4. $r_t(\theta) = \pi_\theta(a_t|s_t) \cdot \pi_{\text{old}}(a_t|s_t)$.

—

Réponses correctes (masquées pour le test)

1. **1.** A
2. **2.** A
3. **3.** B
4. **4.** A
5. **5.** B
6. **6.** B

7. **7.** A
8. **8.** B
9. **9.** A
10. **10.** C