

Список вопросов к зачёту/экзамену по курсу «Обучение с подкреплением», весна 2023

Вопрос 1: Кросс-энтропийный метод в общем виде. Его применение для решения задач оптимизации и задач обучения с подкреплением. **Источники:** главы 1-2 (в первую очередь раздел 2.2) конспекта [1].

Вопрос 2: Уравнения Беллмана для функций ценности. Алгоритмы Policy/Value Iteration. **Источники:** разделы 3.1-3.3 конспекта [1].

Вопрос 3: Табличные методы: Монте-Карло, Q-learning. **Источники:** разделы 3.4, 3.5 конспекта [1].

Вопрос 4: Алгоритм DQN и его модификации: Double DQN, приоритизированный буфер, дуальная архитектура, шумные сети, многошаговый DQN, память. **Источники:** разделы 4.1, 4.2 конспекта [1].

Вопрос 5: Distributional-подход в RL. Алгоритм QR-DQN. **Источники:** раздел 4.3 конспекта [1].

Вопрос 6: Подход Policy gradient. Алгоритмы Reinforce и A2C. **Источники:** разделы 5.1, 5.2 конспекта [1].

Вопрос 7: Метод Trust-Region Policy Optimization (TRPO), его теоретическое обоснование. **Источники:** раздел 5.3 (в первую очередь подразделы 5.3.1-5.3.3) конспекта [1].

Вопрос 8: Bias-variance trade-off в обучении с подкреплением. Оценка GAE. Алгоритм Proximal Policy Optimization (PPO). **Источники:** раздел 5.3 (в первую очередь подразделы 5.3.4-5.3.6) конспекта [1].

Вопрос 9: Детерминированный градиент по политике. Off-policy алгоритмы для задач непрерывного управления: DDPG, Twin Delayed DDPG (TD3). **Источники:** раздел 6.1 конспекта [1].

Вопрос 10: Обучение с подкреплением с добавлением энтропии. Алгоритм Soft Actor-Critic. **Источники:** раздел 6.2 конспекта [1].

Вопрос 11: Имитационное обучение и обратное обучение с подкреплением. Схема Guided Cost Learning. Генеративно-сопоставительное имитационное обучение (GAIL). **Источники:** раздел 8.1 конспекта [1].

Вопрос 12: Задача многоруких бандитов, UCS-бандиты. Внутренняя мотивация: дистилляция случайной сети (RND) и внутренний модуль любопытства (ICM). **Источники:** разделы 7.1, 7.2, 8.2 конспекта [1].

Вопрос 13: Monte Carlo Tree Search в общем виде. Методы AlphaZero и MuZero. **Источники:** глава 7 (в первую очередь раздел 7.3) конспекта [1].

Вопрос 14: Линейно-квадратичный регулятор и его итеративная версия. Общая схема Model-based RL. **Источники:** глава 7 (в первую очередь раздел 7.4) конспекта [1].

[1] *Ivanov, Sergey. Reinforcement Learning Textbook.* arXiv preprint arXiv:2201.09746 (2022).