Список вопросов к зачёту/экзамену по курсу «Обучение с подкреплением», весна-лето 2025

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

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

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

Вопрос 4: Алгоритм DQN и его модификации: Double DQN, приоритезированный буфер, дуэльная архитектура, шумные сети, многошаговый DQN, память. `Источники:` лекция 4, разделы 4.1, 4.2 конспекта [1].

Вопрос 5: Distributional подход в RL. Алгоритм QR-DQN. Алгоритм Implicit Quantile Networks. `Источники:` лекция 5, раздел 4.3 конспекта [1].

Вопрос 6: Внутренняя мотивация: дистилляция случайной сети (RND) и внутренний модуль любопытства (ICM). `Источники:` лекция 6, раздел 8.2 конспекта [1].

Вопрос 7: Подход Policy Gradient. Алгоритм REINFORCE и его модификации. `Источники:` лекция 7, начало 8-ой лекции, разделы 5.1, 5.2 конспекта [1].

Вопрос 8: Подход Policy Gradient. Алгоритм A2C и его модификации, оценка GAE. `Источники:` лекция 8, разделы 5.1, 5.2 конспекта [1].

Вопрос 9: Метод Trust-Region Policy Optimization (TRPO), его теоретическое обоснование. `Источники:` лекция 9, раздел 5.3 (в первую очередь подразделы 5.3.1-5.3.3) конспекта [1].

Вопрос 10: Bias-variance trade-off в обучении с подкреплением. Оценка GAE. Алгоритм Proximal Policy Optimization (PPO). `Источники:` лекция 10, раздел 5.3 (в первую очередь подразделы 5.3.4-5.3.6) конспекта [1].

Вопрос 11: Детерминированный градиент по политике. Off-policy алгоритмы для задач непрерывного управления: DDPG, Twin Delayed DDPG (TD3). `Источники:` лекция 11, раздел 6.1 конспекта [1].

Вопрос 12: Обучение с подкреплением с добавлением энтропии. Алгоритм Soft Actor-Critic. `Источники:` лекция 12, раздел 6.2 конспекта [1].

Вопрос 13: Имитационное обучение и обратное обучение с подкреплением. Схема Guided Cost Learning. Генеративно-состязательное имитационное обучение (GAIL). `Источники:` лекция 13, раздел 8.1 конспекта [1].

Вопрос 14: Задача многоруких бандитов, UCB-бандиты, алгоритм сэмплирования по Томпсону. `Источники:` лекция 14, разделы 7.1, 7.2 конспекта [1].

Вопрос 15: Monte Carlo Tree Search в общем виде. `Источники:` лекция 15, глава 7 (в первую очередь раздел 7.3) конспекта [1].

Вопрос 16: Линейно-квадратичный регулятор и его итеративная версия. Общая схема Model-based RL. `Источники:` лекция 15, глава 7 (в первую очередь раздел 7.4) конспекта [1].

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