

Список вопросов к экзамену по курсу «Обучение с подкреплением», осень 2020

1. Кросс-энтропийный метод в общем виде. Его применение для решения задач оптимизации и задач обучения с подкреплением.
2. Уравнения Беллмана для функций ценности. Алгоритмы Policy/Value Iteration.
3. Методы временных разностей (Q-learning, Sarsa, Expected Sarsa). Модель DQN.
4. Distributional-подход в RL. Алгоритмы c51 и QR-DQN.
5. Подход Policy gradient. Алгоритмы Reinforce и A2C.
6. Метод Trust-Region Policy Optimization (TRPO), его теоретическое обоснование.
7. Оценка GAE, метод Proximal Policy Optimization (PPO).
8. Обучение с подкреплением с добавлением энтропии. Off-policy алгоритмы для задач непрерывного управления (Soft Actor-Critic, DDPG).
9. Имитационное обучение и обратное обучение с подкреплением. Схема Cost-guided learning.
10. Monte Carlo Tree Search в общем виде. Методы AlphaZero и MuZero.
11. Линейно-квадратичный регулятор и его итеративная версия. Общая схема Model-based RL.

Теоретический минимум

Вопросы из этой части охватывают базовые математические понятия и алгоритмы, которые активно используются в курсе. Незнание ответа на любой вопрос из данной части автоматически влечёт за собой неудовлетворительную оценку по экзамену.

1. Постановка задачи обучения с подкреплением в виде марковского процесса принятия решений. Примеры прикладных задач.
2. Дилемма exploration vs. exploitation. Подходы к её решению.
3. Понятие on-policy и off-policy алгоритмов обучения с подкреплением. Примеры.
4. Уравнения Беллмана для функций ценности. Связь функций ценности между собой.
5. Понятия soft Q-функции и V-функции. Уравнения Беллмана для них. Оптимальная политика для них.
6. Схема метода Q-обучения и DQN.
7. Схема метода A2C.