Список вопросов для промежуточной аттестации по дисциплине «Введение в обучение с подкреплением».

1. Общая постановка задачи обучения с подкреплением, основные понятия и обозначения: агент, среда, схема их взаимодействия, траектория, состояния, действия, вознаграждения, доход, коэффициент обесценивания, стратегия, функции ценности V и Q, марковская модель среды. Базовые подходы к решению задач RL.
2. Многорукий бандит. Постановка задачи. Оценка ценностей действий. Жадный и нежадный выбор, eps-жадные стратегии. Способы обновления оценки. Модификации: оптимистичные старты, ВДГ-стратегии, выборка Томпсона
3. Дискретный марковский процесс, марковский процесс вознаграждений. Функция ценности состояний и методы её поиска: точный и итеративный.
4. Марковские процессы принятия решения (МППР). Конечные МППР, функция динамики среды. Связь МППР и МПВ. Уравнение Беллмана для функций ценности. Оценка стратегии, итеративный метод оценки стратегии. Уравнение Беллмана в виде диаграммы.
5. Теорема об улучшении стратегии. Уравнение оптимальности Беллмана и единственность его решения. Определение и существование оптимальных функций ценности. Оптимальные стратегии.
6. Уравнение оптимальности Беллмана, оптимальные функции ценности, оптимальные стратегии и методы их поиска. Метод итерации по стратегиям и по ценностям. Обобщённая итерации по стратегиям.
7. Метод Монте-Карло и его особенности. Оценка стратегии и поиск оптимальной стратегии. Исследовательские старты и итерация по eps-мягким стратегиям
8. TD методы и их особенности. Оценка стратегии. Метод SARSA для поиска оптимальной стратегии: формулировка и обоснование работы алгоритма.
9. Метод Q-learning: формулировка и обоснование работы алгоритма. Особенности. Метод ExpectedSARSA и n-шаговые TD методы.
10. Работа в средах с непрерывным пространством состояний. Методы приближения функций ценности с помощью линейных функций. Формула для обновления параметров при оценке стратегии и при поиске оптимальных стратегий.
11. Использование ИНС для приближения функций ценности. Метод DQN: формулировка и особенности работы. Дополнительные модификации: буфер памяти, использование двух моделей.
12. Метод REINFOCE: формулировка и особенности работы. Теорема о градиенте стратегии.
13. Метод актор-критик, понятие преимущества, формулировки работы метода A2C и особенности. Работа в средах с непрерывным пространством действий. Метод PPO.