

Задание на лабораторную работу № 6
по курсу «Методы искусственного интеллекта»
«Обучение с подкреплением»

В рамках лабораторной работы следует ознакомиться с двумя базовыми алгоритмами обучения с подкреплением, основанным на идее временных различий: Sarsa и Q-обучение. В работе используются полуградиентные разновидности этих алгоритмов, адаптированные для работы с линейными аппроксимациями функции ценности действий. За основу рекомендуется взять [блокнот из репозитория курса](#), изменив и доработав его в соответствии с заданием:

1. У реализованного в блокноте алгоритма Sarsa есть несколько параметров: коэффициент дисконтирования, константа обучения, параметр ϵ -жадной стратегии, параметры плиточного кодирования. Исследуйте влияние этих параметров на скорость обучения и характеристики найденной стратегии.
2. Что будет, если в ходе обучения постепенно изменять параметр ϵ , приближая стратегию выбора действия к жадной.
3. На базе класса CartPoleSarsaAgent реализуйте агента, использующего алгоритм Q-обучения. Подсказка: для этого достаточно будет изменить две-три строчки. Исследуйте поведение этого агента.
4. (не обязательно, +3 балла). Одной из причин не особенно впечатляющего поведения реализованного выше агента является то, что он использует только два из параметров, описывающих систему (пусть и самые информативные), из-за чего ему "тяжело" идентифицировать ситуацию скольжения тележки и выезда ее за пределы допустимой области, что также вызывает прерывание эпизода. Добавьте в модель признаки, соответствующие положению тележки. Для этого, например, можно расширить плиточное кодирование на третье измерение, а можно поступить еще проще, предположив аддитивность эффектов этих признаков.