Конфигурация задачи.

В рамках данного проекта была выбрана среда LunarLander-v2. Смысл игры состоит в том, чтобы посадить спускаемый аппарат на посадочную площадку, которая имеет координаты (0,0). В качестве показателя «успешности» посадки выступают очки награды (reward). Так, самое большое количество очков (200) приносит решенная задача – посадка лунного модуля на посадочную площадку. Кроме того, дополнительно начисляется по 10 очков за контакт каждой ноги с землей. Если спускаемый аппарат удаляется от посадочной площадки, то очки теряются. Также очки теряются за «стрельбу» главного двигателя – минус 0,3 балла за один запуск и за «стрельбу» из левого или правого двигателя – минус 0,03 балла за один запуск. Эпизод заканчивается, если посадочный модуль падает или останавливается, получая дополнительные -100 или +100 очков.

Существует несколько возможностей «поиграть» в данной среде: либо человек непосредственно сам отправляет команды посадочному модулю, либо это выполняет искусственный интеллект. В рамках проекта рассматривается второй вариант. Взаимодействие со средой осуществляется при помощи четырех дискретных действий (action): ничего не делать, запустить двигатель левой ориентации, запустить главный двигатель или запустить двигатель правой ориентации. Откликом среды являются очки награды, которые начисляются по окончании эпизода.

Задачей данного проекта является написание алгоритма, который будет осуществлять посадку LunarLander при этом максимизируя количество очков награды. Для этого разрабатывается и реализовывается три модели обучения с подкреплением (Reinforcement Learning) в специфике выбранной среды. Далее происходит исследование и улучшение получившихся алгоритмов с целью максимизации очков награды.