Московский государственный технический университет им. Н.Э. Баумана

Факультет «Головной учебно-исследовательский и методический центр

профессиональной реабилитации лиц с ограниченными возможностями здоровья (инвалидов)»

Кафедра «Системы обработки информации и управления»



**Лабораторная работа 4**

**по дисциплине «Методы машинного обучения в АСОИУ»**

**"** **Реализация алгоритма Policy Iteration "**

СТУДЕНТ:

студент группы ИУ5Ц-21М

Москалик А.А.

ПРЕПОДАВАТЕЛЬ:

Гапанюк Ю.Е.

Москва, 2024

# 

# Цель лабораторной работы: ознакомление с базовыми методами обучения с подкреплением.

# Задание: на основе рассмотренного на лекции примера реализуйте алгоритм Policy Iteration для любой среды обучения с подкреплением (кроме рассмотренной на лекции среды Toy Text / Frozen Lake) из библиотеки Gym (или аналогичной библиотеки).

# Ход работы

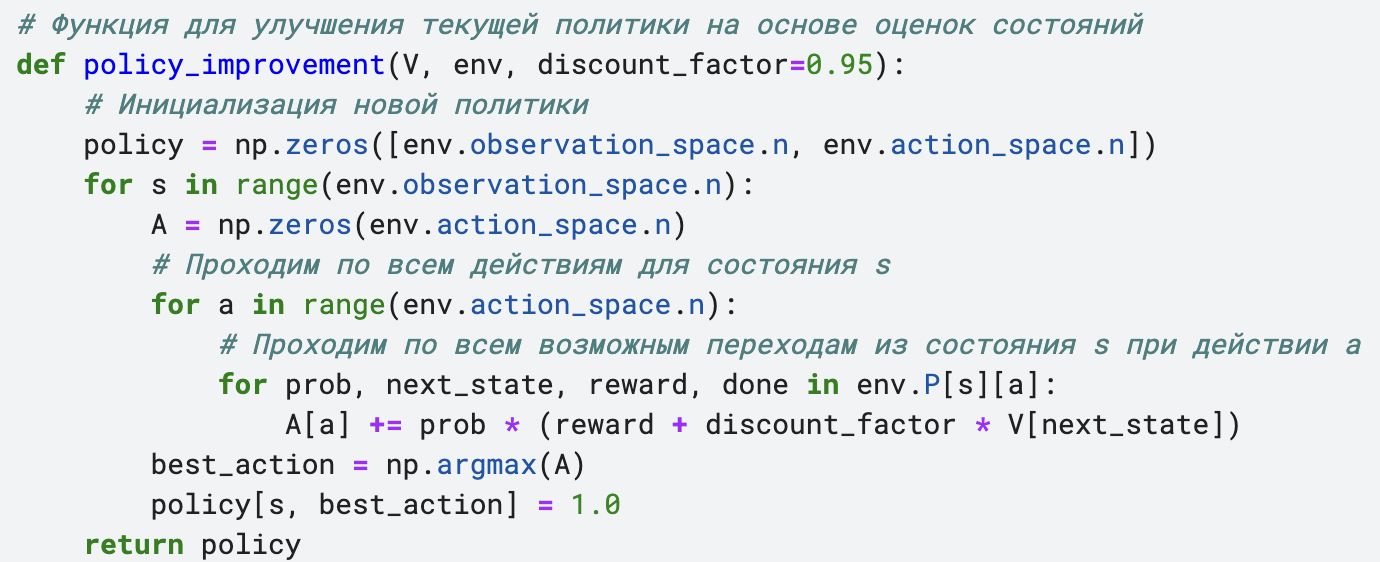
# Импорт необходимых библиотек и определение функции policy\_evaluation

Функция policy\_evaluation — это часть алгоритма итерации политики (Policy Iteration), которая используется для оценки текущей политики. Цель этой функции — вычислить ценности состояний (value function) для данной политики, что позволяет понять, насколько хороша текущая политика в терминах ожидаемой награды. Эта информация затем используется для улучшения политики.

****

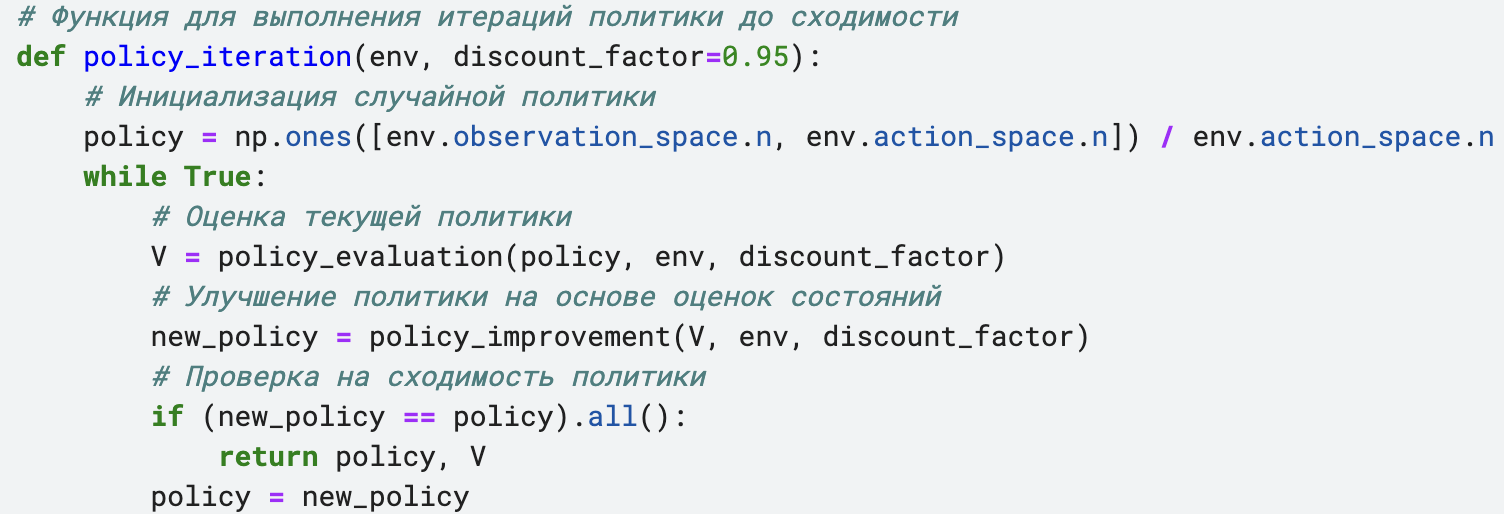
**Определение функции policy\_improvement**

Функция policy\_improvement — это часть алгоритма итерации политики, которая используется для улучшения текущей политики на основе оценок состояний. Основная идея этой функции заключается в выборе наилучших действий для каждого состояния, которые максимизируют ожидаемую награду.

****

**Определение функции policy\_iteration**

Функция policy\_iteration — это ключевая часть алгоритма итерации политики (Policy Iteration) в обучении с подкреплением. Этот алгоритм используется для нахождения оптимальной политики в детерминированных и стохастических средах. Алгоритм состоит из двух основных шагов: оценка текущей политики (policy evaluation) и улучшение политики (policy improvement). Эти шаги повторяются до тех пор, пока политика не перестанет изменяться, что свидетельствует о достижении оптимальной политики.



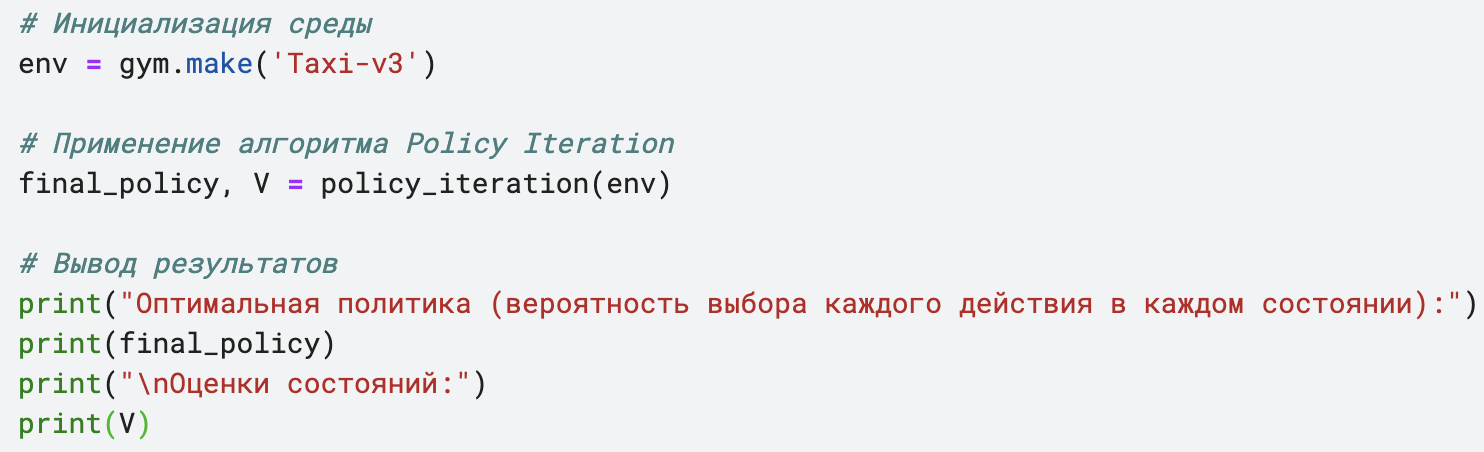
**Инициализация среды, запуск алгоритма и вывод результатов**

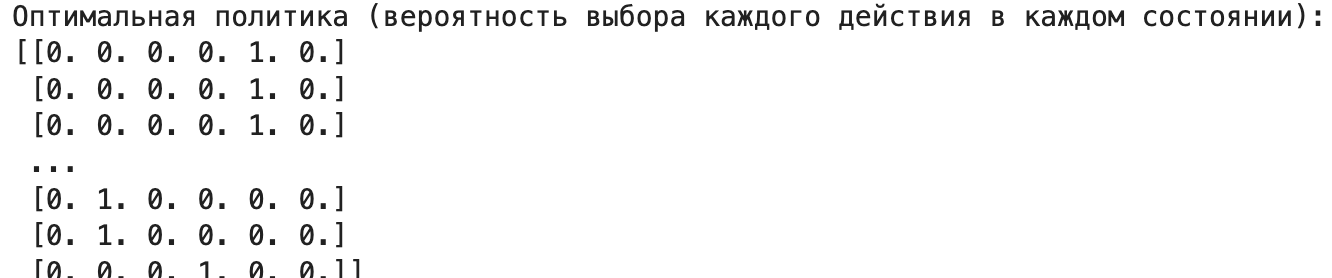
На этом шаге мы инициализируем среду Taxi-v3 из библиотеки Gym. Среда представляет собой задачу, в которой агент (такси) должен забрать пассажира в одном месте и отвезти его в другое. Среда включает дискретное пространство состояний и действий.

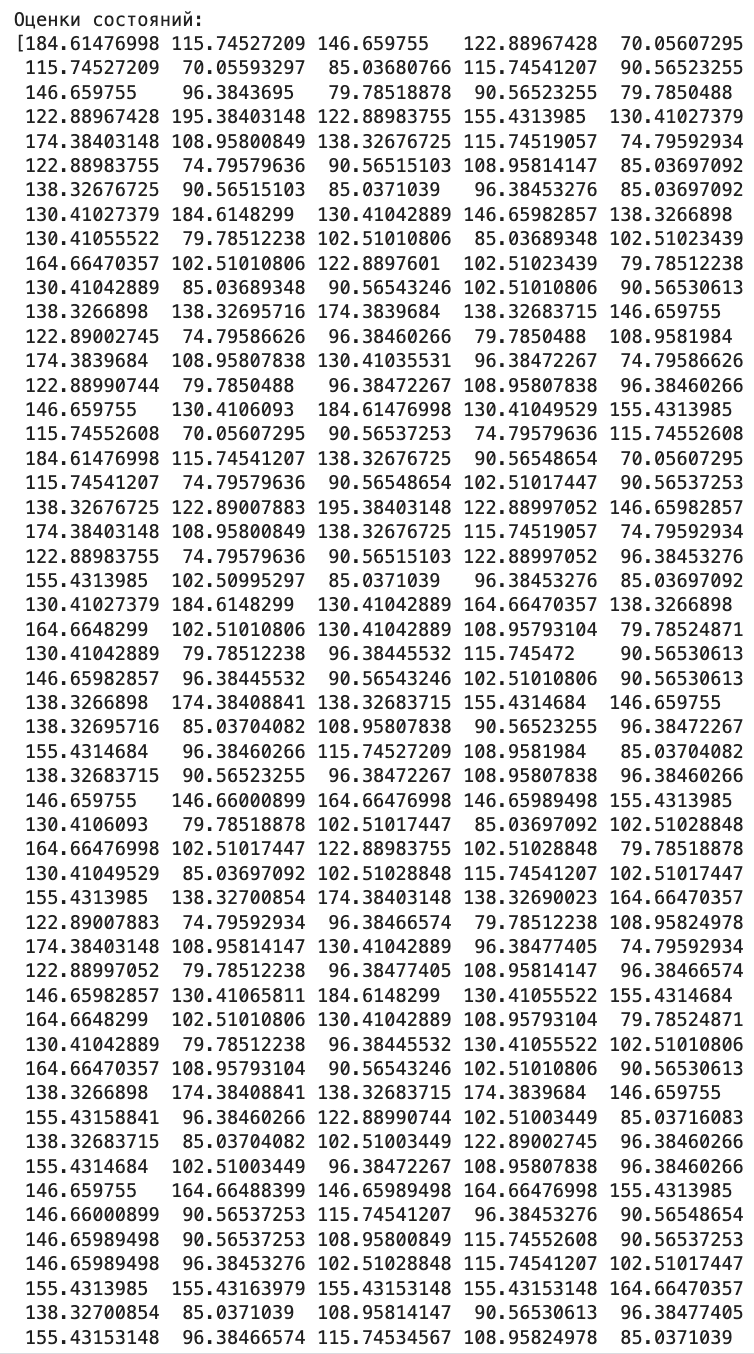
Далее вызывается функция policy\_iteration, которая реализует алгоритм итерации политики. Этот алгоритм проходит через циклы оценки политики и улучшения политики до тех пор, пока не будет достигнута оптимальная политика. В результате мы получаем:

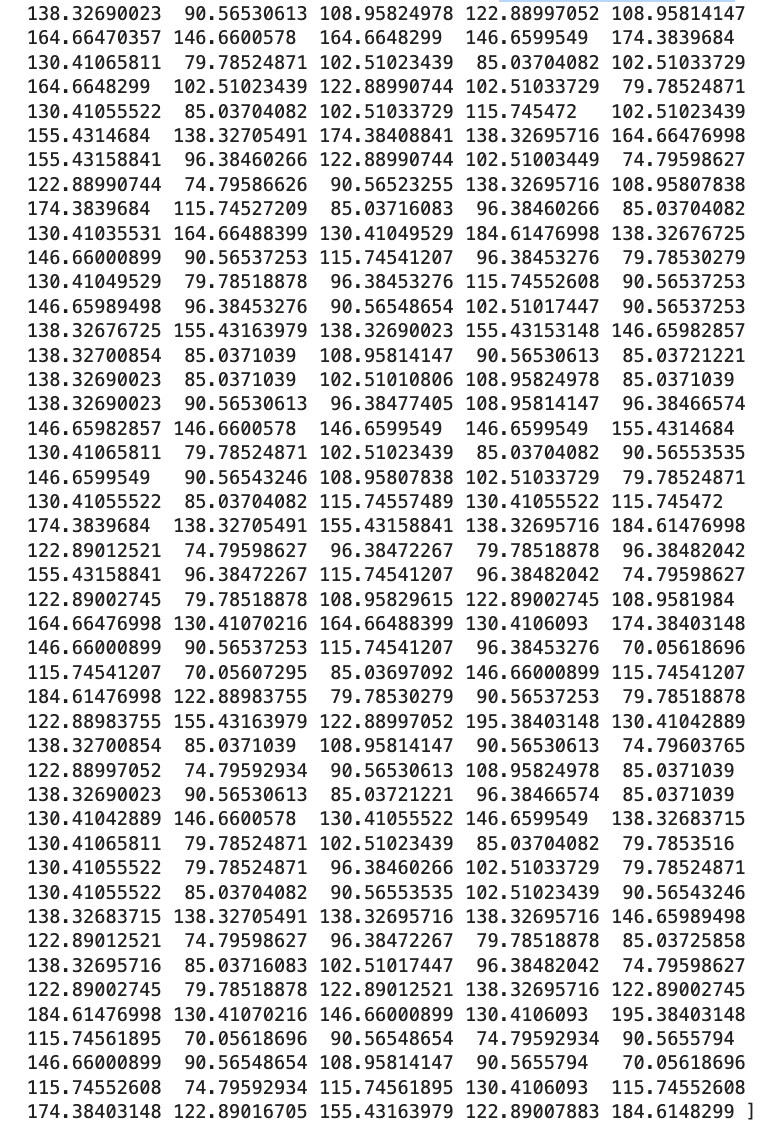
- final\_policy: оптимальная политика, представленная в виде матрицы, где каждая строка соответствует состоянию, а столбцы — действиям. Значение 1.0 в столбце означает, что это действие является оптимальным для данного состояния.

- V: оценки состояний, представляющие собой массив, где каждое значение отражает ожидаемую суммарную награду, начиная из соответствующего состояния и следуя оптимальной политике.









Оптимальная политика: Алгоритм итерации политики успешно определил оптимальные действия для каждого состояния в среде `Taxi-v3`. Эти действия обеспечивают максимизацию ожидаемой награды, что соответствует целям агента.

Оценки состояний: Значения оценок состояний предоставляют полезную информацию о ценности каждого состояния. Это помогает понять, какие состояния являются более предпочтительными для агента с точки зрения получения награды.

**Вывод:**

Алгоритм Policy Iteration продемонстрировал свою эффективность в решении задачи управления такси в среде `Taxi-v3`. Реализация этого алгоритма позволила глубже понять методы обучения с подкреплением и их применение в задачах принятия решений. Полученные результаты показывают, что алгоритмы обучения с подкреплением могут быть успешно применены для решения сложных задач управления и оптимизации.