****Московский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №4  
по дисциплине  
«Методы машинного обучения»  
на тему

**«Разведочный анализ данных. Исследование и визуализация данных»**

Выполнил:  
студент группы ИУ5И-21М  
Ван Чжэн

Москва — 2024 г.

**1. Цель лабораторной работы**

изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

**2. Задание**

1. 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).

**3. Ход выполнения работы**

Среда CartPole-v1 - это классическая среда управления, предоставляемая OpenAI Gym, которая также известна как задача инверсии маятника. Эта среда проста в дизайне, но вызывает вызовы и предназначена для тестирования производительности алгоритмов обучения с подкреплением в непрерывных пространствах состояний и действий.

В среде CartPole-v1 есть маленький вагончик (cart), который может двигаться влево и вправо по горизонтальному пути, сверху находится шест (pole). Задача шеста - поддерживать вертикальное положение, не падая. Игрок может управлять вагончиком, прикладывая силу влево или вправо, чтобы попытаться сохранить баланс шеста.

Пространство состояний среды является непрерывным и состоит из четырех значений:

1. Положение вагончика (cart position)
2. Скорость вагончика (cart velocity)
3. Угол шеста (pole angle)
4. Угловая скорость шеста (pole angular velocity)

Игрок может совершать два дискретных действия:

* 0: Приложить силу влево
* 1: Приложить силу вправо

Награда в среде организована следующим образом:

* На каждом временном шаге, если шест все еще стоит, награда равна 1
* Если шест падает или вагончик выходит за границы (превышает ограничения), награда равна 0, и игра заканчивается

Цель среды CartPole-v1 заключается в том, чтобы с помощью подходящей стратегии управления обеспечить как можно более длительное время вертикальное положение шеста, то есть максимизировать накопленную награду.

