|  |  |
| --- | --- |
| C:\Users\USER\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\616878FB.tmp | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

|  |  |
| --- | --- |
| ФАКУЛЬТЕТ | Информатика, искусственный интеллект и системы управления |
| КАФЕДРА | Системы обработки информации и управления |

**Рубежный контроль №2**

**«Методы обучения с подкреплением»**

Студент группы ИУ5-25М

Зозуля О.А.

Москва, 2023 г.

**Задание**

Для алгоритма временных различий SARSA осуществите подбор гиперпараметров. Критерием оптимизации должна являться суммарная награда.

**Описание выполнения**

Алгоритм SARSA испытывается в среде CliffWalking-v0, в котором необходимо найти кратчайший безопасный маршрут от старта до финиша.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Номер набора*** | ***1*** | ***2*** | ***3*** | ***4*** |
| eps | 0.4 | 0,1 | 0,1 | 0,1 |
| lr | 0.1 | 0,1 | 0,15 | 0,15 |
| gamma | 0.98 | 0.98 | 0.96 | 0,99 |
| num\_episodes | 20000 | 20000 | 20000 | 5000 |
| Суммарная награда | - 485840 | -325612 | -332116 | -94469 |

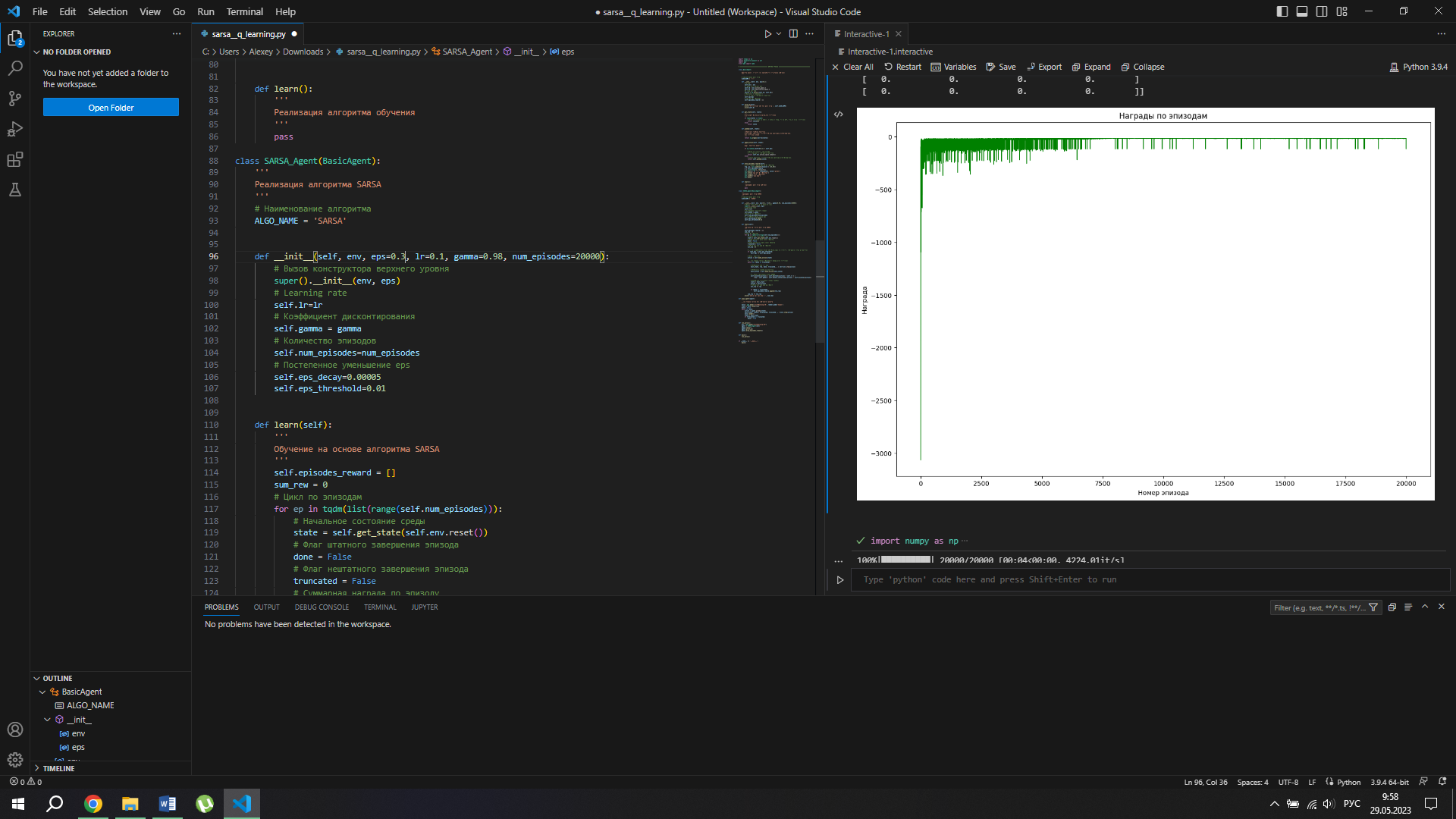


Рисунок 1 – Набор 1

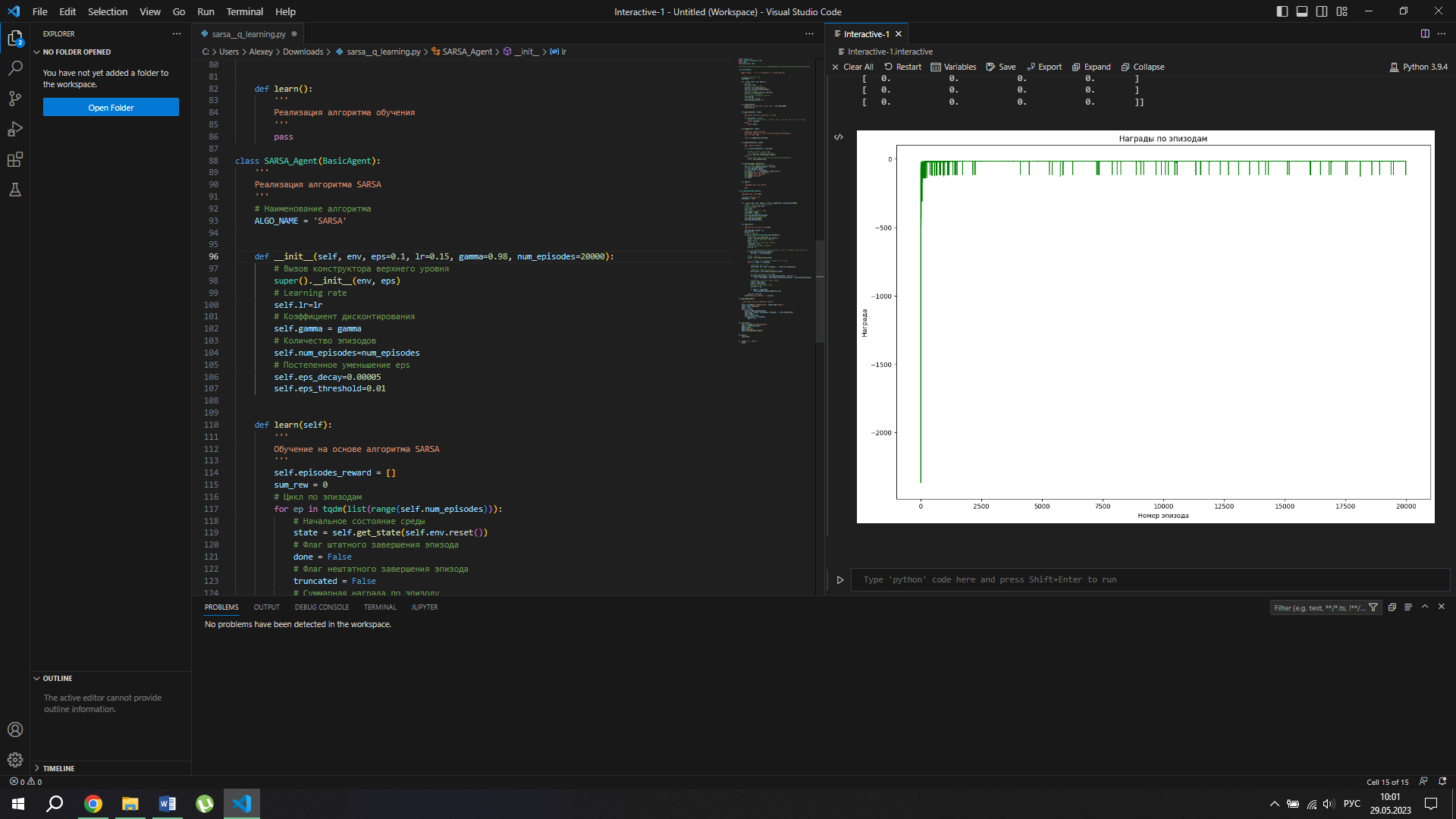


Рисунок 2 – Набор 2

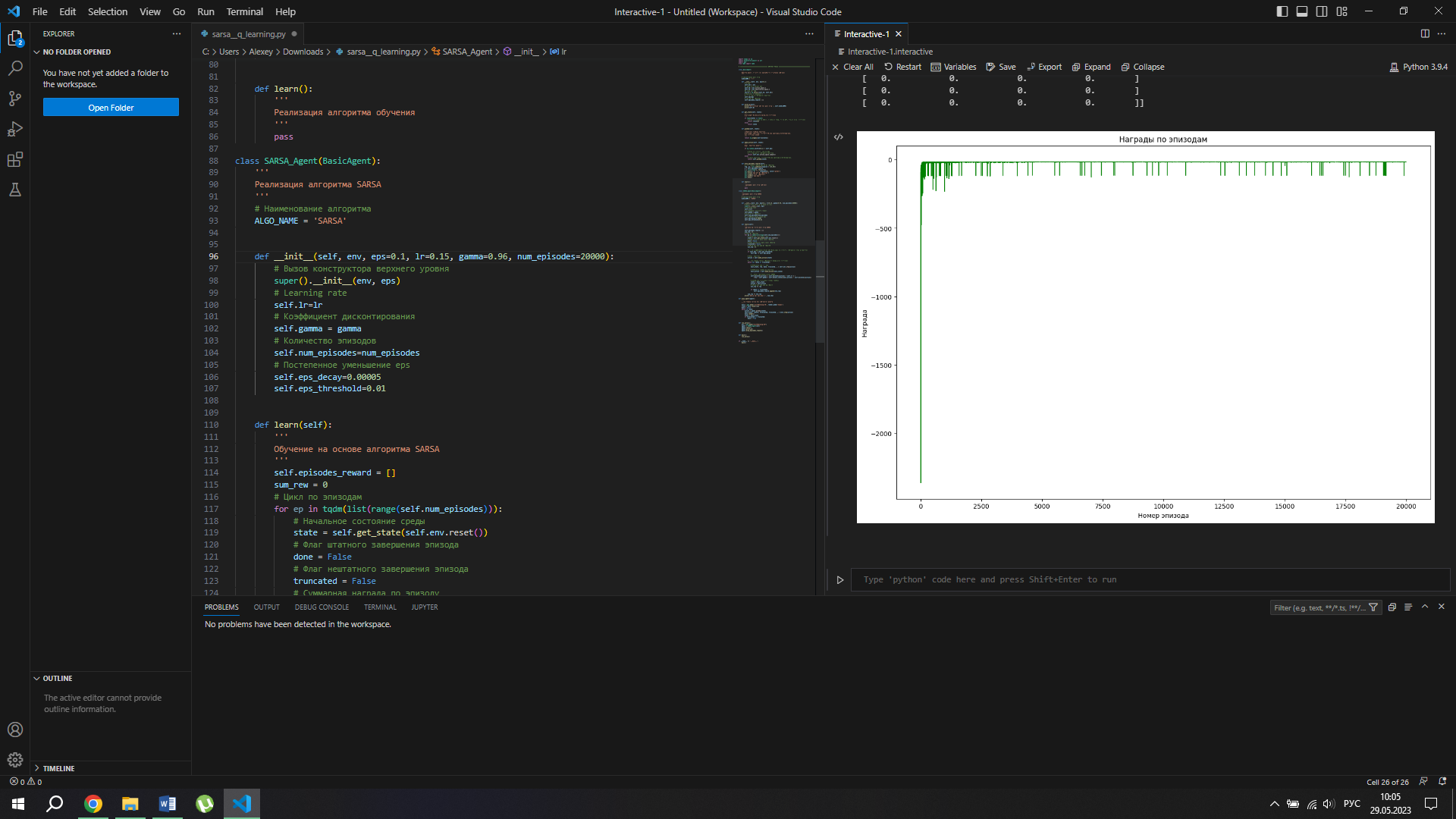


Рисунок 3 – Набор 3

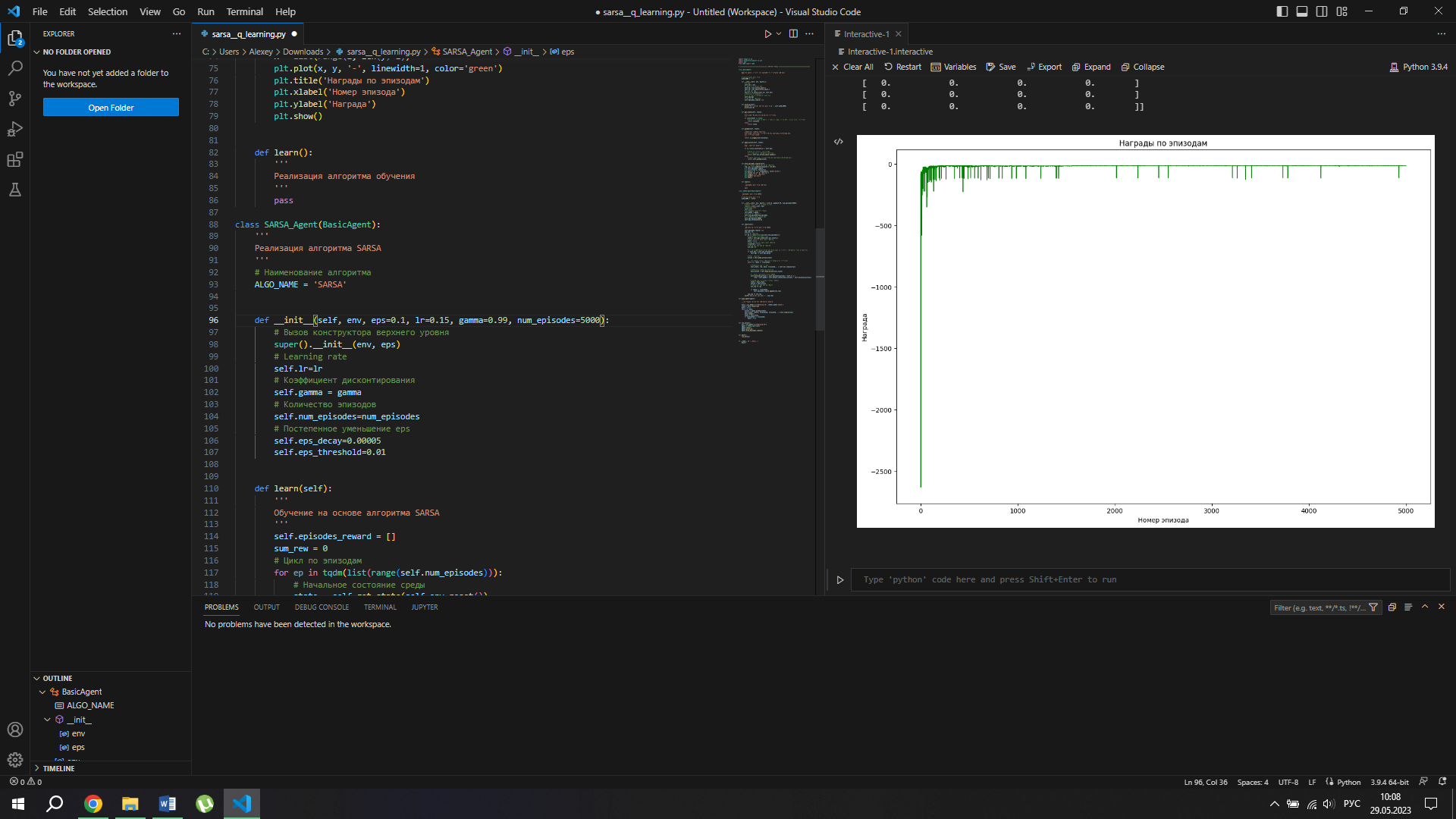


Рисунок 4 – Набор 4

**Вывод**

Таким образом, при выполнении данной работы были произведены эксперименты, на которых алгоритм с определенными гиперпараметрами определял различную суммарную награду.

Награду удалось увеличить за счет снижения параметра eps, на основе которого алгоритм выбирал чаще всего наилучшее действие, небольшого увеличения скорости обучения – lr, также небольшого увеличении параметра, который отвечает за ценность долгосрочной награды-gamma. Затем было сильно сокращено количество эпизодов, так как алгоритм достаточно быстро обучался в районе четверти от исходного количества эпизодов, следовательно, дальнейшее обучения было неэффективно и более времязатратно.

Из всего вышеперечисленного можно сделать вывод, что подбор гиперпараметров помог увеличить суммарную награду, выдаваемую в ходе обучения алгоритма, почти в 5 раз.