

# HW3

Отчёт по домашнему заданию 3.

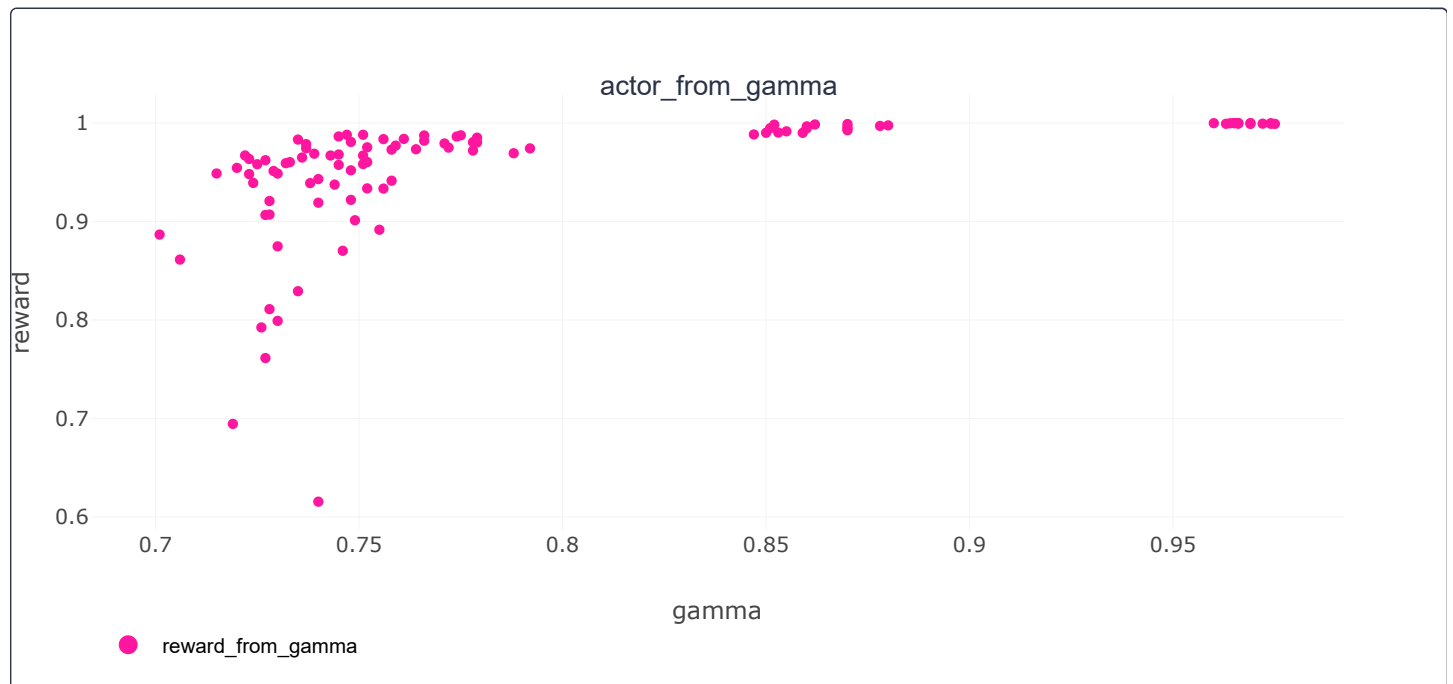
Репозиторий

telegram

\*\* Для первого и второго задания файл агента и файл оптимизации.

## Задание 1

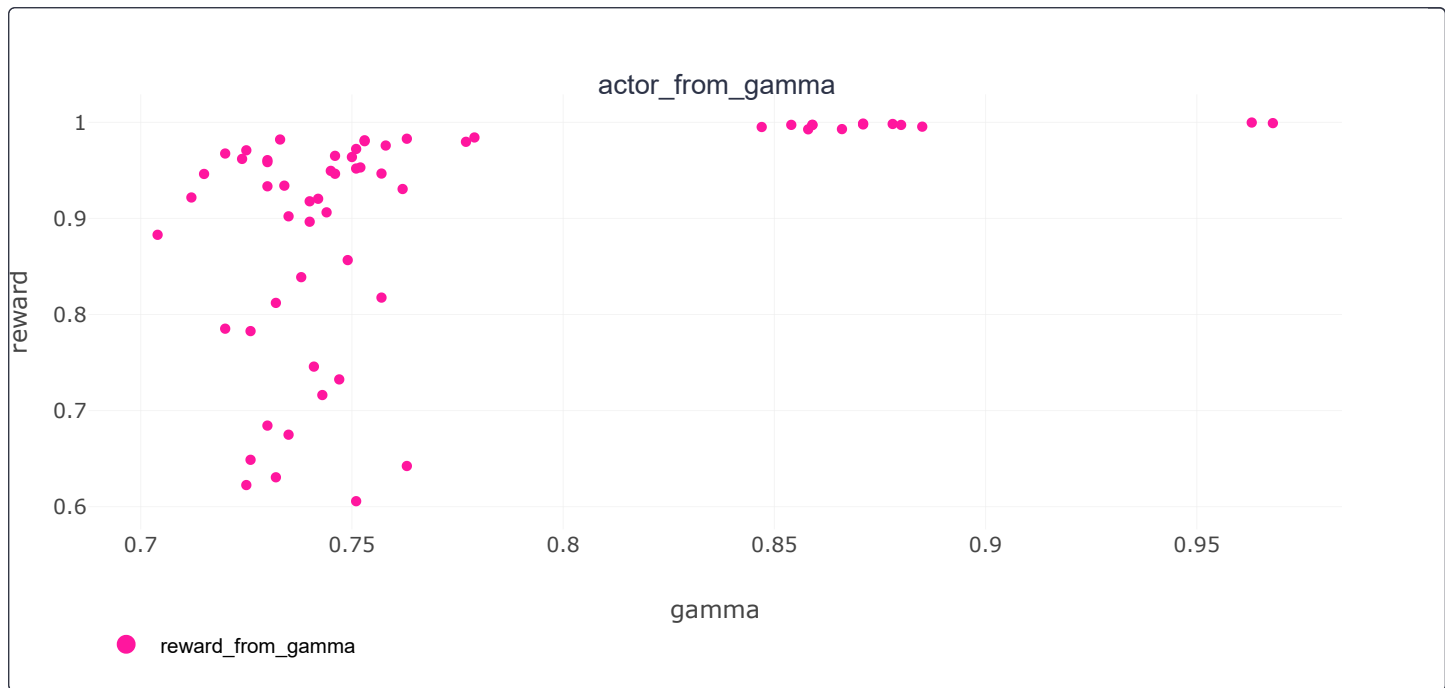
С помощью библиотеки optuna было подобрано оптимальное значение параметра  $\gamma = 0.9990733250956919$ . На графике представлены результаты 100 итераций:



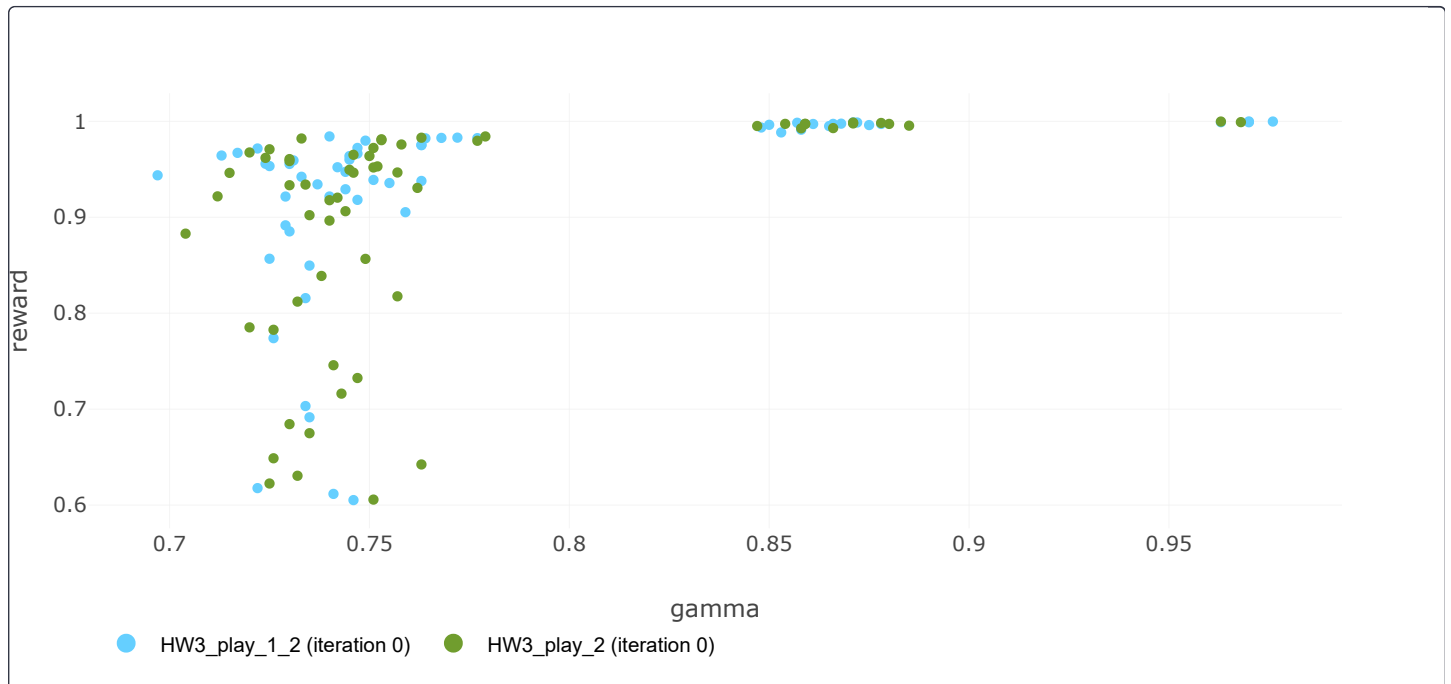
## Задание 2

Был добавлен флаг `warmstart`, при котором агент брал `v_values`, полученные на предыдущем шаге при оценке `q_values`.

Таким образом, с помощью optuna был подобран параметр  $\gamma$ . График 60 триалов представлен далее:



Если сравнивать два подхода (с использованием предыдущих значений и без), то не видется существенной разницы. На графике ниже HW3\_play\_1\_2 -- warmstart=False, HW3\_play\_2 -- warmstart=True.



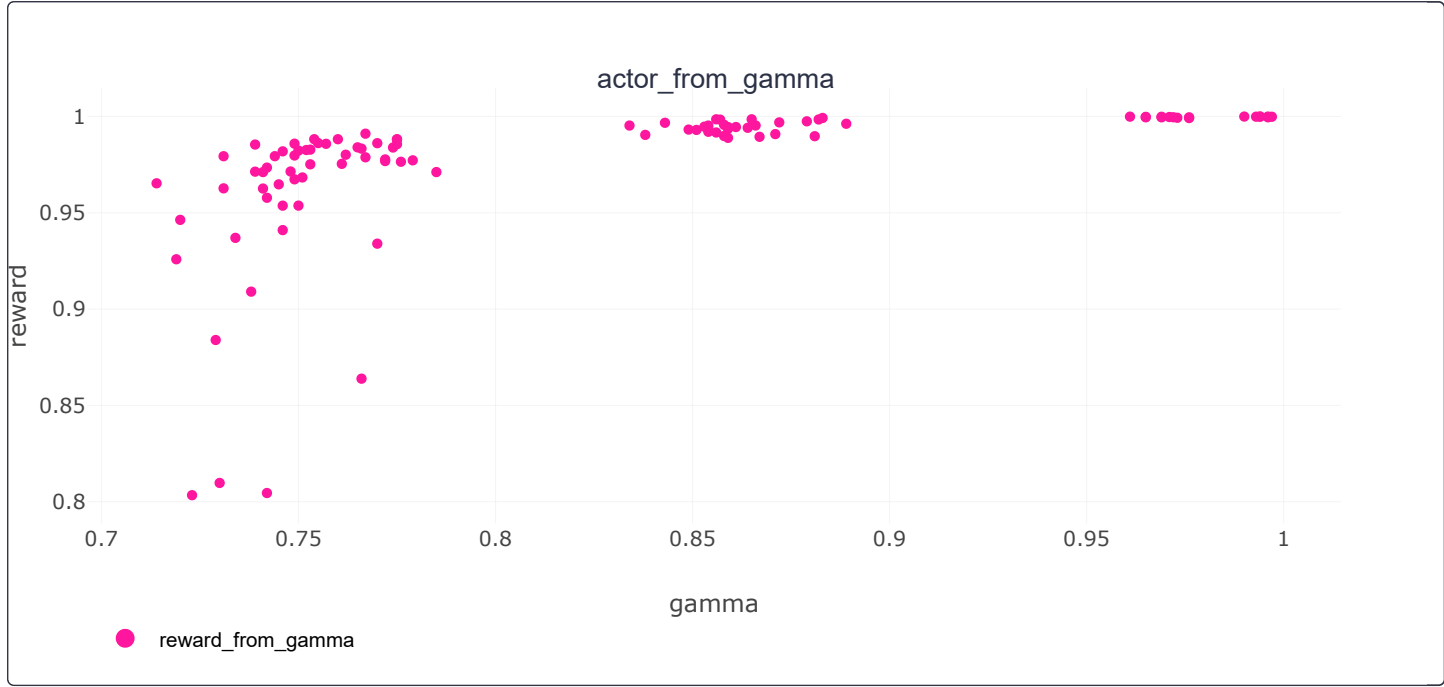
### Задание 3

Была реализована функция value iteration. [Файл репозитория](#).

Функция была оптимизирована в [файле](#) с помощью optuna. Исследовалось три гиперпараметра: гамма, эпсилон (коэффициент ранней остановки) и количество итераций. Лучшие параметры

```
{'n_iterations': 1678, 'gamma': 0.9998483181481228, 'epsilon': 0.0047095883762882235}
```

На следующем графике представлена зависимость средней награды по итерациям:



Как видно на графике, алгоритм не уступает Policy Iteration.

### Сравнение агентов

На следующем графике представлено сравнение трех агентов. Агент с Value\_iteration при меньшем количестве параметров добился лучших результатов при фиксированном  $\gamma$ , алгоритм policy\_iteration с флагом warmstart=True отработал хуже warmstart=False:

