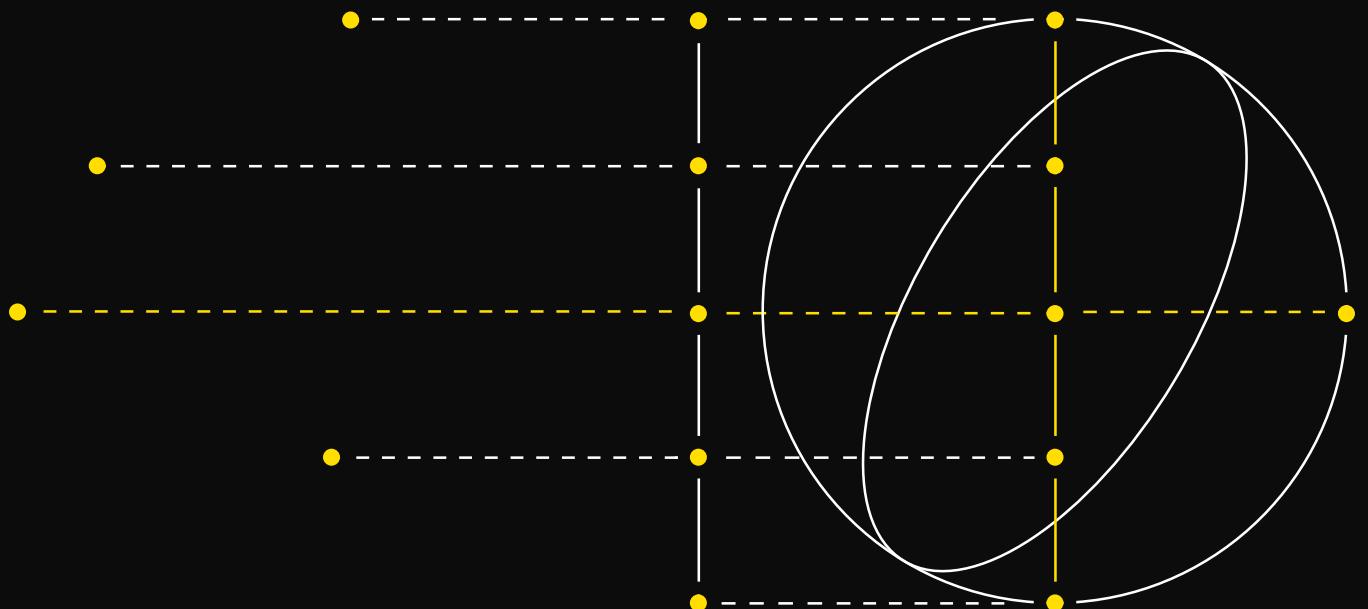


Обучение агента



ИЗБРАННЫЕ ТЕМЫ ИССЛЕДОВАНИЙ В АІ

ДОМАШНЕЕ ЗАДАНИЕ

НЕДЕЛЯ 1



Задача: имплементировать основные алгоритмы, которые мы обсудили на занятии.
Возьми окружение [CartPole](#) из gymnasium и запускайся на нём. Нагугли или спроси GPT, как пользоваться джимами.

REINFORCEMENT LEARNING

Твоей политикой должна быть небольшая нейронка типа MLP (можно поиграться с гиперпараметрами).
Дальше определи несколько функций потерь и обучи с ними политику.

- Vanilla Policy Gradient.
- PG с бейзлайном:
 - Средняя награда, наблюдаемая за время обучения.
 - Велью функция (тоже небольшая нейронка, её тоже нужно обучать параллельно с обучением политики со своим лоссом).
 - RLOO-style бейзайн.
- Добавь к лоссу регуляризацию на энтропию, поиграйся с гиперпараметром. Возможно, есть смысл его заскедулить (то есть, например, линейно изменять по ходу обучения).
- Проанализируй полученные результаты и сравни их между собой. Какие преимущества и недостатки есть у каждого подхода? Выходом этого пункта должен быть репорт а-ля секция экспериментов в статьях.

BEHAVIOUR CLONING

- Возьми лучшую из обученных политик (должна сойтись к оптимуму, идеально решать задачу). Это будет твоим экспертом. Теперь ты можешь нагенерировать много траекторий, запуская из разных начальных состояний. Важно убедиться, что эти траектории действительно набирают максимальные награды.
Должен получиться датасет из пар: состояние, действие.
- Возьми (необученную) нейронку из предыдущего задания и обучи её в supervised-стиле на этом датасете. Проверь качество получившейся политики, также опиши всё в репорте.
- Зная о недостатках этого подхода (мы их обсуждали на занятии), задизайни эксперименты, которые отражают эти недостатки. Покажи, в каких случаях обученная политика будет работать хорошо, а в каких — плохо.

ВСПОМОГАТЕЛЬНЫЕ МАТЕРИАЛЫ

Примеры обучения РЛ-алгоритмов можно найти тут:

Ссылка
Ссылка