Лекция 6: Policy Gradient

Антон Романович Плаксин

Markov Decision Process

Markov Property

$$\mathbb{P}[S_{t+1}|S_t, A_t] = \mathbb{P}[S_{t+1}|S_1, A_1, S_2, A_2 \dots, S_t, A_t]$$

$$\mathbb{P}[R_t|S_t, A_t] = \mathbb{P}[R_t|S_1, A_1, S_2, A_2 \dots, S_t, A_t] = 1$$

Markov Decision Process $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$

- $\mathcal{S}$ — бесконечное (конечное) пространство состояний
- $\mathcal{A}$ — бесконечное (конечное) пространство действий
- $\mathcal{P}$ — неизвестная функция вероятностей переходов между состояниями

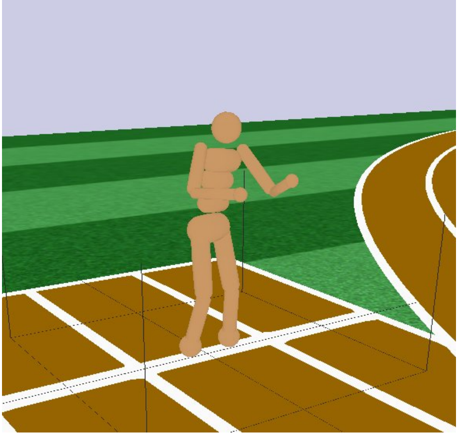
$$\mathcal{P}(s'|s, a) = \mathbb{P}[S_{t+1} = s'|S_t = s, A_t = a]$$

- $\mathcal{R}$ — неизвестная функция вознаграждений

$$\mathcal{R}(s, a) = R_t \quad \Leftrightarrow \quad \mathbb{P}[R_t|S_t = s, A_t = a] = 1$$

- $\gamma \in [0, 1]$ — коэффициент дисконтирования

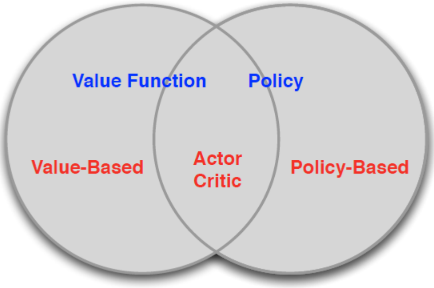
Пример



- Состояния: $\mathbb{R}^{26}$
- Действия: $\mathbb{R}^{6}$
- Награда: $+1$ в каждый момент времени

Value-Based and Policy-Based RL

- Value-Based
  - Обучение Value Function
  - Неявные Policy
- Policy-Based
  - Без Value Function
  - Обучение Policy
- Actor-Critic
  - Обучение Value Function
  - Обучение Policy



Основная идея

Аппроксимация Policy

$$\pi^{\eta}(a|s) \approx \pi_*(a|s),$$

где $\eta \in \mathbb{R}^N$ — вектор параметров

Основная идея

Аппроксимация Policy

$$\pi^\eta(a|s) \approx \pi_*(a|s),$$

где $\eta \in \mathbb{R}^N$ — вектор параметров

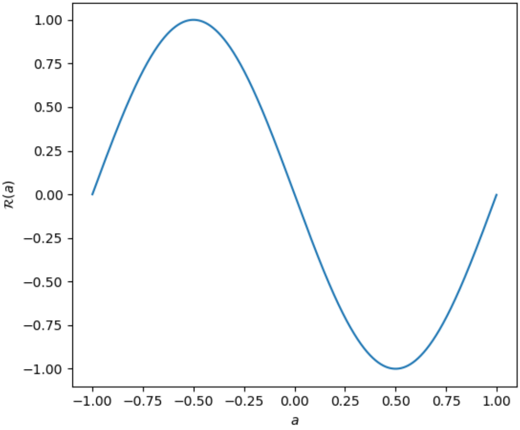
Критерий качества

$$J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] \to \max_\eta,$$

где

$$G_0 = \sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t)$$

Пример: Непрерывный Бандит



- $\mathcal{S} = \{0, 1\}$
- $\mathcal{A} = [-1, 1]$
- $\mathcal{R}(0, a) = \mathcal{R}(a)$

Пример: Непрерывный Бандит

$$\pi^\eta(a) = \pi^\eta(a|0), \quad J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \int\limits_{a \in \mathcal{A}} \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a$$

Пример: Непрерывный Бандит

$$\pi^\eta(a) = \pi^\eta(a|0), \quad J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \int\limits_{a \in \mathcal{A}} \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a$$

Вычисление градиента методом Monte-Carlo

$$\nabla_{\eta_i} J(\eta) \approx \frac{J(\eta + \delta e^i) - J(\eta)}{\delta}, \quad \delta > 0, \quad i \in \overline{1, N},$$

где $e_i \in \mathbb{R}^N$: $e^i_j = 1$, если $j = i$ и $e^i_j = 0$, если $j \neq i$,

$$J(\eta) \approx \frac{1}{n}\sum_{k=1}^{n} \mathcal{R}(a_k), \quad a_k \sim \pi^\eta(\cdot)$$

Пример: Непрерывный Бандит

$$\pi^\eta(a) = \pi^\eta(a|0), \quad J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \int\limits_{a \in \mathcal{A}} \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a$$

Вычисление градиента с помощью Log Derivative Trick

$$\begin{aligned} \nabla_\eta J(\eta) &= \int\limits_{a \in \mathcal{A}} \nabla_\eta \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a \\ &\approx \frac{1}{n}\sum_{k=1}^{n} \mathcal{R}(a_k), \quad a_k \sim \nabla_\eta \pi^\eta(\cdot) \quad ? \end{aligned}$$

Пример: Непрерывный Бандит

$$\pi^\eta(a) = \pi^\eta(a|0), \quad J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \int\limits_{a\in\mathcal{A}} \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a$$

Вычисление градиента с помощью Log Derivative Trick

$$\begin{aligned}
\nabla_\eta J(\eta) &= \int\limits_{a\in\mathcal{A}} \nabla_\eta \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a = \int\limits_{a\in\mathcal{A}} \pi^\eta(a)\frac{\nabla_\eta \pi^\eta(a)}{\pi^\eta(a)}\mathcal{R}(a)\mathrm{d}a \\
&= \int\limits_{a\in\mathcal{A}} \pi^\eta(a)\nabla_\eta \ln \pi^\eta(a)\mathcal{R}(a)\mathrm{d}a = \mathbb{E}_{\pi^\eta}[\nabla_\eta \ln \pi^\eta(a)\mathcal{R}(a)] \\
&\approx \frac{1}{n}\sum_{k=1}^{n} \nabla_\eta \ln \pi^\eta(a_k)\mathcal{R}(a_k), \quad a_k \sim \pi^\eta(\cdot)
\end{aligned}$$

Policy Gradient Theorem

Discounted State Distribution

$$\rho_\pi(s) = \sum_{t=0}^{\infty} \gamma^t \mathbb{P}[S_t = s | S_0, \pi]$$

Policy Gradient Theorem

Discounted State Distribution

$$\rho_\pi(s) = \sum_{t=0}^{\infty} \gamma^t \mathbb{P}[S_t = s | S_0, \pi]$$

$$\begin{aligned} \mathbb{E}_\pi[G_0] &= \int_{s \in \mathcal{S}} \rho_\pi(s) \int_{a \in \mathcal{A}} \pi(a|s) r(s,a) \mathrm{d}a \mathrm{d}s \\ &= \mathbb{E}_{s \sim \rho_\pi, a \sim \pi}[r(s,a)] \end{aligned}$$

Policy Gradient Theorem

$$J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[r(s, a)]$$

Policy Gradient Theorem

$$J(\eta) = \mathbb{E}_{\pi^\eta}[G_0] = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[r(s,a)]$$

Теорема

Пусть $\pi^\eta(a|s)$ дифференцируема по η
и $\pi^\eta(a|s) \neq 0$ для любых $s \in \mathcal{S}$, $a \in \mathcal{A}$. Тогда

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s) q_{\pi^\eta}(s,a)]$$

Reinforce: Основная идея

Аппроксимация Policy

$$\pi^{\eta}(a|s) \approx \pi_{*}(a|s),$$

Reinforce: Основная идея

Аппроксимация Policy

$$\pi^{\eta}(a|s) \approx \pi_*(a|s),$$

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s) q_{\pi^\eta}(s,a)]$$

Reinforce: Основная идея

Аппроксимация Policy

$$\pi^{\eta}(a|s) \approx \pi_*(a|s),$$

Policy Gradient Theorem

$$\nabla_{\eta} J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^{\eta}}, a\sim\pi^{\eta}}[\nabla_{\eta} \ln \pi^{\eta}(a|s) q_{\pi^{\eta}}(s,a)]$$

Monte-Carlo Estimate

$$q_{\pi^{\eta}}(s,a) = \mathbb{E}_{\pi^{\eta}}[G_t | S_t = s,\ A_t = a] \approx G_t = \sum_{i=t}^{T-1} \gamma^{i-t} R_i$$

Reinforce Algorithm

Задаем структуру аппроксимации $\pi^\eta(a|s)$ и начальный вектор параметров η.
Для каждого эпизода делаем:

Reinforce Algorithm

Задаем структуру аппроксимации $\pi^\eta(a|s)$ и начальный вектор параметров η.
Для каждого эпизода делаем:

- Согласно π^η получаем траекторию $\tau = (S_0, A_0, \ldots, S_T)$ и награды $(R_0, \ldots, R_{T-1})$. По ним определяем $(G_0, \ldots, G_{T-1})$:

$$G_t = \sum_{i=t}^{T-1} \gamma^{i-t} R_i$$

Reinforce Algorithm

Задаем структуру аппроксимации $\pi^\eta(a|s)$ и начальный вектор параметров η.
Для каждого эпизода делаем:

- Согласно π^η получаем траекторию $\tau = (S_0, A_0, \ldots, S_T)$ и награды $(R_0, \ldots, R_{T-1})$. По ним определяем $(G_0, \ldots, G_{T-1})$:
$$G_t = \sum_{i=t}^{T-1} \gamma^{i-t} R_i$$
- Для каждого $t \in \overline{0, T-1}$ обновляем η:
$$\eta \leftarrow \eta - \alpha \nabla_\eta \ln \pi^\eta(A_t|S_t)\, G_t$$

On-Policy Actor-Critic: Основная идея

Аппроксимация Policy и Action-Value Function

$$\pi^\eta(a|s) \approx \pi_*(a|s), \quad Q^\theta(s,a) \approx q_{\pi^\eta}(s,a)$$

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s) q_{\pi^\eta}(s,a)]$$

Аппроксимация на каждом шаге

Если $Q^\theta \approx q_{\pi^\eta}$, то

$$\begin{aligned} \nabla_\eta J(\eta) &\approx \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s) Q^\theta(s,a)] \\ &\approx \nabla_\eta \ln \pi^\eta(A_t|S_t) Q^\theta(S_t, A_t) \end{aligned}$$

On-Policy Actor-Critic: Основная идея

Аппроксимация Policy и Action-Value Function

$$\pi^\eta(a|s) \approx \pi_*(a|s), \quad Q^\theta(s,a) \approx q_{\pi^\eta}(s,a)$$

Bellman Expectation Equation для q_π

$$q_{\pi^\eta}(s,a) = \mathbb{E}_{\pi^\eta}[R_t + \gamma q_{\pi^\eta}(S_{t+1}, A_{t+1}) | S_t = s,\ A_t = a]$$

Аппроксимация на каждом шаге

Если

$$Q^\theta(S_t, A_t) \approx R_t + \gamma Q^\theta(S_{t+1}, A_{t+1}), \quad A_{t+1} \sim \pi^\eta(\cdot|S_t),$$

то $Q^\theta \approx q_{\pi^\eta}$

On-Policy Actor-Critic Algorithm

Задаем структуру аппроксимаций $\pi^\eta(a|s)$, $Q^\theta(s,a)$ и начальные вектора параметров η, θ.
Для каждого эпизода делаем:
Пока эпизод не закончен делаем:

- Находясь в состоянии S_t совершаем действие $A_t \sim \pi^\eta(\cdot|S_t)$, получаем награду R_t, переходим в состояние S_{t+1}.
- По (S_t, A_t, R_t, S_{t+1}) выбираем действие $A_{t+1} \sim \pi^\eta(\cdot|S_{t+1})$, определяем Loss функции:

$$Loss_1(\theta) = \left(R_t + \gamma Q^\theta(S_{t+1}, A_{t+1}) - Q^\theta(S_t, A_t)\right)^2$$

$$Loss_2(\eta) = \ln \pi^\eta(A_t|S_t) Q^\theta(S_t, A_t)$$

и обновляем параметры:

$$\theta \leftarrow \theta - \alpha \nabla_\theta Loss_1(\theta), \quad \eta \leftarrow \eta + \beta \nabla_\eta Loss_2(\eta)$$

Deterministic Policy Gradient Theorem

Аппроксимация детерминированной Policy

$$\pi^{\eta}(s) \approx \pi_*(s)$$

Теорема

Пусть $\pi^{\eta}(s)$ дифференцируема по η
и $q_{\pi^{\eta}}(s, a)$ дифференцируема по a. Тогда

$$\nabla_{\eta} J(\eta) = \mathbb{E}_{s \sim \rho_{\pi^{\eta}}} \left[\nabla_{\eta} q_{\pi^{\eta}}(s, \pi^{\eta}(s)) \right]$$

DDPG (Off-Policy Actor-Critic): Основная идея

Deterministic Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s \sim \rho_{\pi^\eta}} [\nabla_\eta q_{\pi^\eta}(s, \pi^\eta(s))]$$

DDPG (Off-Policy Actor-Critic): Основная идея

Deterministic Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s \sim \rho_{\pi^\eta}} [\nabla_\eta q_{\pi^\eta}(s, \pi^\eta(s))]$$

Аппроксимация по батчу

Если $\pi \approx \pi^\eta$ и $Q^\theta \approx q_{\pi^\eta}$, то

$$\nabla_\eta J(\eta) \approx \nabla_\eta \left(\frac{1}{n} \sum_{i=1}^{n} Q^\theta(s_i, \pi^\eta(s_i)) \right)$$

DDPG (Off-Policy Actor-Critic): Основная идея

Deterministic Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s \sim \rho_{\pi^\eta}} [\nabla_\eta q_{\pi^\eta}(s, \pi^\eta(s))]$$

Аппроксимация по батчу

Если $\pi \approx \pi^\eta$ и $Q^\theta \approx q_{\pi^\eta}$, то

$$\nabla_\eta J(\eta) \approx \nabla_\eta \left(\frac{1}{n} \sum_{i=1}^{n} Q^\theta(s_i, \pi^\eta(s_i)) \right)$$

Bellman Expectation Equation для q_π

$$q_{\pi^\eta}(s, a) = \mathbb{E}[R_t + \gamma q_{\pi^\eta}(S_{t+1}, \pi^\eta(S_{t+1})) | S_t = s,\ A_t = a]$$

DDPG (Off-Policy Actor-Critic): Основная идея

Deterministic Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s \sim \rho_{\pi^\eta}} [\nabla_\eta q_{\pi^\eta}(s, \pi^\eta(s))]$$

Аппроксимация по батчу

Если $\pi \approx \pi^\eta$ и $Q^\theta \approx q_{\pi^\eta}$, то

$$\nabla_\eta J(\eta) \approx \nabla_\eta \left(\frac{1}{n} \sum_{i=1}^{n} Q^\theta(s_i, \pi^\eta(s_i)) \right)$$

Bellman Expectation Equation для q_π

$$q_{\pi^\eta}(s, a) = \mathbb{E}[R_t + \gamma q_{\pi^\eta}(S_{t+1}, \pi^\eta(S_{t+1})) | S_t = s,\ A_t = a]$$

Аппроксимация по батчу

Если

$$\frac{1}{n} \sum_{i=1}^{n} \left(r_i + \gamma Q^\theta(s_i', \pi^\eta(s_i')) - Q^\theta(s_i, a_i) \right) \approx 0,$$

то $Q^\theta \approx q_{\pi^\eta}$

DDPG (Off-Policy Actor-Critic) Algorithm

Задаем структуру аппроксимаций $\pi^\eta(s)$, $Q^\theta(s,a)$ и начальные вектора параметров η, θ.
Для каждого эпизода делаем:
  Пока эпизод не закончен делаем:

DDPG (Off-Policy Actor-Critic) Algorithm

Задаем структуру аппроксимаций $\pi^\eta(s)$, $Q^\theta(s,a)$ и начальные вектора параметров η, θ.
Для каждого эпизода делаем:
Пока эпизод не закончен делаем:

- Находясь в состоянии S_t совершаем действие
$$A_t = \pi^\eta(S_t) + Noise,$$
получаем награду R_t переходим в состояние S_{t+1}.
Сохраняем $(S_t, A_t, R_t, S_{t+1}) \to Memory$

DDPG (Off-Policy Actor-Critic) Algorithm

Задаем структуру аппроксимаций $\pi^\eta(s)$, $Q^\theta(s,a)$ и начальные вектора параметров η, θ.

Для каждого эпизода делаем:

Пока эпизод не закончен делаем:

- Находясь в состоянии S_t совершаем действие
$$A_t = \pi^\eta(S_t) + Noise,$$
получаем награду R_t переходим в состояние S_{t+1}.
Сохраняем $(S_t, A_t, R_t, S_{t+1}) \to Memory$
- Берем $\{(s_i, a_i, r_i, s_i')\}_{i=1}^n \leftarrow Memory$, определяем значения
$$y_i = r_i + \gamma Q^\theta(s_i', \pi^\eta(s_i'))$$
функции потерь
$$Loss_1(\theta) = \frac{1}{n}\sum_{i=1}^n \big(y_i - Q^\theta(s_i,a_i)\big)^2, \quad Loss_2(\eta) = \frac{1}{n}\sum_{i=1}^n Q^\theta(s_i, \pi^\eta(s_i))$$
и обновляем вектор параметров
$$\theta \leftarrow \theta - \alpha\nabla_\theta Loss_1(\theta), \quad \eta \leftarrow \eta + \beta\nabla_\eta Loss_2(\eta), \quad \alpha,\beta > 0$$

DDPG (Off-Policy Actor-Critic) Algorithm

Задаем структуру аппроксимаций $\pi^\eta(s)$, $Q^\theta(s,a)$ и начальные вектора параметров η, θ.

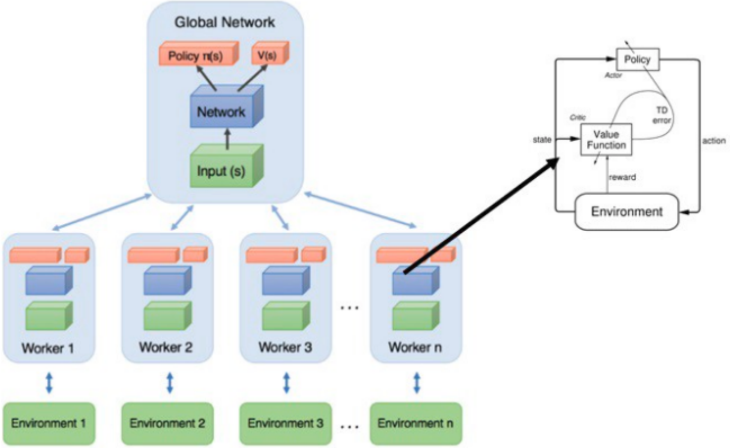
Для каждого эпизода делаем:

Пока эпизод не закончен делаем:

- Находясь в состоянии S_t совершаем действие
$$A_t = \pi^\eta(S_t) + Noise,$$
получаем награду R_t переходим в состояние S_{t+1}.
Сохраняем $(S_t, A_t, R_t, S_{t+1}) \to Memory$
- Берем $\{(s_i, a_i, r_i, s_i')\}_{i=1}^n \leftarrow Memory$, определяем значения
$$y_i = r_i + \gamma Q^\theta(s_i', \pi^\eta(s_i'))$$
функции потерь
$$Loss_1(\theta) = \frac{1}{n}\sum_{i=1}^n \big(y_i - Q^\theta(s_i,a_i)\big)^2, \quad Loss_2(\eta) = \frac{1}{n}\sum_{i=1}^n Q^\theta(s_i, \pi^\eta(s_i))$$
и обновляем вектор параметров
$$\theta \leftarrow \theta - \alpha\nabla_\theta Loss_1(\theta), \quad \eta \leftarrow \eta + \beta\nabla_\eta Loss_2(\eta), \quad \alpha, \beta > 0$$
- Уменьшаем $Noise$

Asynchronous Approach

Asynchronous Approach



Advantage Actor-Critic (A2C): Основная идея

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s)(q_{\pi^\eta}(s,a) - b(s))]$$

Advantage Actor-Critic (A2C): Основная идея

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s)(q_{\pi^\eta}(s,a) - b(s))]$$

Advantage Function

$$b(s) = v_{\pi^\eta}(s), \quad a_{\pi^\eta}(s,a) = q_{\pi^\eta}(s,a) - v_{\pi^\eta}(s)$$

Advantage Actor-Critic (A2C): Основная идея

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s)(q_{\pi^\eta}(s,a) - b(s))]$$

Advantage Function

$$b(s) = v_{\pi^\eta}(s), \quad a_{\pi^\eta}(s,a) = q_{\pi^\eta}(s,a) - v_{\pi^\eta}(s)$$

$$a_{\pi^\eta}(s,a) = \mathbb{E}[R_t + \gamma v_{\pi^\eta}(S_{t+1}) - v_{\pi^\eta}(s)|S_t = s, A_t = a]$$

Advantage Actor-Critic (A2C): Основная идея

Аппроксимация Policy и Action-Value Function

$$\pi^\eta(a|s) \approx \pi_*(a|s), \quad V^\theta(s) \approx v_{\pi^\eta}(s)$$

Policy Gradient Theorem

$$\nabla_\eta J(\eta) = \mathbb{E}_{s\sim\rho_{\pi^\eta}, a\sim\pi^\eta}[\nabla_\eta \ln \pi^\eta(a|s) a_{\pi^\eta}(s,a)],$$

где

$$a_{\pi^\eta}(s,a) = \mathbb{E}[R_t + \gamma v_{\pi^\eta}(S_{t+1}) - v_{\pi^\eta}(s)|S_t = s, A_t = a]$$

Аппроксимация на шаге

Если $V^\theta \approx v_{\pi^\eta}$, то

$$\nabla_\eta J(\eta) \approx \nabla_\eta \ln \pi^\eta(A_t|S_t)\big(R_t + \gamma V^\theta(S_{t+1}) - V^\theta(S_t)\big)$$

Advantage Actor-Critic (A2C): Основная идея

Аппроксимация Policy и Action-Value Function

$$\pi^\eta(a|s) \approx \pi_*(a|s), \quad V^\theta(s) \approx v_{\pi^\eta}(s)$$

Bellman Expectation Equation для v_π

$$v_{\pi^\eta}(s) = \mathbb{E}_{\pi^\eta}[R_t + \gamma v_{\pi^\eta}(S_{t+1})|S_t = s]$$

Аппроксимация на шаге

Если

$$V^\theta(S_t) \approx R_t + \gamma V^\theta(S_{t+1}), \quad S_{t+1} \sim \mathcal{P}(\cdot|S_t, A_t), \quad A_t \sim \pi^\eta(\cdot|S_t)$$

то $V^\theta \approx v_{\pi^\eta}$

Advantage Actor-Critic (A2C)

Задаем структуру аппроксимаций $\pi^\eta(a|s)$, $V^\theta(s)$ и начальные вектора параметров η, θ.

Для каждого эпизода делаем:

Пока эпизод не закончен делаем:

- Находясь в состоянии S_t совершаем действие $A_t \sim \pi^\eta(\cdot|S_t)$, получаем награду R_t, переходим в состояние S_{t+1}.
- По (S_t, A_t, R_t, S_{t+1}) определяем Loss функции:

$$Loss_1(\theta) = \big(R_t + \gamma V^\theta(S_{t+1}) - V^\theta(S_t)\big)^2$$

$$Loss_2(\eta) = \ln \pi^\eta(A_t|S_t)\big(R_t + \gamma V^\theta(S_{t+1}) - V^\theta(S_t)\big)$$

и обновляем параметры:

$$\theta \leftarrow \theta - \alpha \nabla_\theta Loss_1(\theta), \quad \eta \leftarrow \eta + \beta \nabla_\eta Loss_2(\eta)$$

Организационные вопросы

- Пятница, 17:50, аудитория 622
- Отчетность: домашние работы
- Страничка курса: https://github.com/imm-rl-lab/UrFU_course
- E-mail для связи: a.r.plaksin@gmail.com

ВОПРОСЫ?