




pdf файл с вашим решением присылайте на grishanov.av@phystech.edu
с темой **recsys mipt 2024 hw2 <Фамилия>**.

1 (1.5 + 1* балл) Многорукие бандиты

Рассмотрим задачу 3-рукого бандита с изображения (действие — выбор конкретного товара, награда — полученный рейтинг).

У вас есть информация о средней награде $D = \{(1, 4.6), (2, 4.3), (3, 4.7)\}$, а также о числе кликов по каждой ручке (arm).

 <p>Original Apple iPhone 6S 6SP Smartphone 4.7"/5.5" 2GB RA...</p> <p>★ 4,6 181 bought</p> <p>8 490,40 ₽</p> <p>TOP CPE_Original mobile p...</p>	 <p>Original Apple Iphone 8 8P 8 Plus 3GB RAM 64GB/256GB...</p> <p>★ 4,3 21 bought</p> <p>13 824,80 ₽</p> <p>High Tip Mobile_Brand orig...</p>	 <p>CN/RU Unlocked Used Apple iPhone 7 / iPhone 7 Plus Quad...</p> <p>★ 4,7 384 bought</p> <p>8 997,60 ₽</p> <p>True Mobile Phone Store</p>
--	---	---

Здесь и далее будем использовать $[p_1, p_2, p_3]^T$ как обозначение стратегии (policy).

1. (0.5 балла) Найдите ϵ -greedy стратегию π_ϵ (положите $\epsilon = 0.01$).
2. (1 балл) Найдите UCSB стратегию π_{UCB} (α выберите сами, например из $\{0.1, 0.5, 1\}$).
Обратите внимание: неравенство Хёффдинга работает не только для бернуллевских бандитов, но и для произвольных $r \in [0, 1]$, поэтому вы можете нормировать награды в $[0, 1]$ и использовать формулы с лекции.
3. (1* балл) Что нужно чтобы применить здесь томсоновское сэмплирование?

2 (2.5 балла) Counterfactual evaluation

Используя многорукого бандита из [задачи 1](#):

1. посчитайте logging policy π_0
2. оцените (evaluate) стратегию $\pi_1 = [0.3, 0.04, 0.66]^T$
(оцените ожидаемый средний рейтинг при использовании π_1 : $\hat{V}(\pi_1, \mathcal{D}) = \mathbb{E}_{p(x)\pi_1(a|x)p(r|x,a)}[r]$)
3. оцените стратегию $\pi_2 = [0.3, 0.66, 0.04]^T$
4. выберите 1 стратегию из [task 1](#) и оцените ее.
5. Проанализируйте результаты.
Возможно ли оценить стратегии из 3 предыдущих пунктов с адекватной точностью?
Если да — опишите как, иначе обоснуйте почему.

3 (1 балл) Несмещенность IPS

1. (0.5 балла) Докажите что оценивание стратегий через [IPS](#) несмещенное, т.е.

$$\mathbb{E}_{\mathcal{D}} [\hat{V}_{\text{IPS}}(\pi; \mathcal{D})] = V(\pi) = \mathbb{E}_{p(x)\pi(a|x)p(r|x,a)}[r]$$

2. (0.5 балла) При каких необходимых условиях выполняется несмещенность?