




Sent pdf file with your solution to [grishanov.av@phystech.edu](mailto:grishanov.av@phystech.edu)  
with the following topic: **recsys mipt 2023 hw3 <Surname>**.

## I Main part (5+1\* points)

### 1 (1.5 + 1\* point) Multi-armed bandits

Consider 3-armed bandit problem as described in picture (action is choosing particular item, reward is a rating received).

You have information  $\mathcal{D}$  about mean reward and number of clicks for each arm.

 <p>Original Apple iPhone 6S 6SP Smartphone 4.7"/5.5" 2GB RA...</p> <p>★ 4,6 181 bought</p> <p><b>8 490,40 ₺</b></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><b>13 824,80 ₺</b></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><b>8 997,60 ₺</b></p> <p>True Mobile Phone Store</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Here and further you may use  $[p_1, p_2, p_3]^T$  notation for policy.

1. (0.5 point) Compute  $\varepsilon$ -greedy policy  $\pi_\varepsilon$  (set  $\varepsilon = 0.01$ ).
2. (1 point) Compute UCB policy  $\pi_{UCB}$  (set  $\alpha$  by yourself, you may choose from  $\{0.1, 0.5, 1\}$ ).  
Note: Hoeffding inequality works not only for bernoulli rewards, but for arbitrary  $r \in [0, 1]$ , so you can scale reward into  $[0, 1]$  to apply formulas from lecture.
3. (1\* point) Explain what is required to use Thompson Sampling here.

## 2 (2.5 points) Counterfactual evaluation

Using problem setup from [task 1](#):

1. compute estimation of logging policy  $\pi_0$
2. evaluate policy  $\pi_1 = [0.3, 0.04, 0.66]^T$   
(get expected mean rating from running  $\pi_1$ :  $\hat{V}(\pi_1, \mathcal{D}) = \mathbb{E}_{p(x)\pi_1(a|x)p(r|x,a)}[r]$ )
3. evaluate policy  $\pi_2 = [0.3, 0.66, 0.04]^T$
4. choose 1 most promising policy from [task 1](#) and evaluate it.
5. Analyze results.

Is it possible to evaluate policies from 3 previous subtasks with adequate precision? If yes describe how, otherwise explain why.

## 3 (1 point) Unbiasedness of IPS

1. (0.5 point) Prove that [IPS estimator](#) is unbiased, e.g.

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

2. (0.5 point) Under which conditions unbiasedness holds?

## II Extra part (up to 5\* points)

Explore using ChatGPT for broadening recommender systems capabilities.

Possible result might be «hypothesis → evidence via experiments → message to audience».

Some particular directions (among many others):

1. Evaluating ChatGPT explainable recommendations
2. Exploring ChatGPT quality in different recommender system domains (films, music, etc.)
3. Benchmarking ChatGPT on ML-1M (feel free to use [github.com/openai/evals](https://github.com/openai/evals)).

We will evaluate this task by novelty, serendipity and coverage of recsys practitioners benefited from your solution. Approximate grading scale:

- 0.25 points — funny ChatGPT prompt related to recommender systems.
- 5 points — beating recsys SOTA with ChatGPT (deadline for [RecSys'23 LBR](#) is 03.07.2023).