

SESSION 7 - REINFORCEMENT LEARNING

Problem Statement

Instructor: Manu K. Gupta
Email: manu.gupta@ms.iitr.ac.in

Maximizing Revenue of an Online Retail Business

Imagine an Online Retail Business that has millions of customers. These customers are the only people buying some products on the website from time to time, getting them delivered at home. The business is doing good, but the board of executives has decided to take some action plan to maximize revenue even more. This plan consists of offering to the customers the option to subscribe to a premium plan, which will give them some benefits like reduced prices, special deals, etc. This premium plan is offered at a yearly price of ₹1000, and the goal of this online retail business is of course to get the maximum customers to subscribe to this premium plan. Let's do some quick maths to give us some motivation for building an AI agent to maximize the revenues of this business. Let's say that this online retail business has 100 million customers. Now consider two conversion strategies trying to convert the customers to the premium plan: a bad one, with a conversion rate of 1%, and a good one, with a conversion rate of 11%. If the business deploys the bad strategy, it will make in one year a total extra revenue coming from the premium plan subscription of: $100,000,000 \times 0.01 \times 1000 = ₹1,000,000,000$. On the other hand, if the business deploys the good strategy, it will make in one year a total extra revenue coming from the premium plan subscription of $100,000,000 \times 0.11 \times 1000 = ₹11,000,000,000$. Thus, by figuring out the good strategy to deploy, the business maximized its extra revenues by making 10 Billion extra Rupees.

Imagine we have 9 different marketing strategies as shown in the figure below. AI agent will have no

Strategy 1 <ul style="list-style-type: none">• Form 1• Package 1• Ad 1• Special Deal 1	Strategy 2 <ul style="list-style-type: none">• Form 2• Package 2• Ad 2• Special Deal 2	Strategy 3 <ul style="list-style-type: none">• Form 3• Package 3• Ad 3• Special Deal 3
Strategy 4 <ul style="list-style-type: none">• Form 4• Package 4• Ad 4• Special Deal 4	Strategy 5 <ul style="list-style-type: none">• Form 5• Package 5• Ad 5• Special Deal 5	Strategy 6 <ul style="list-style-type: none">• Form 6• Package 6• Ad 6• Special Deal 6
Strategy 7 <ul style="list-style-type: none">• Form 7• Package 7• Ad 7• Special Deal 7	Strategy 8 <ul style="list-style-type: none">• Form 8• Package 8• Ad 8• Special Deal 8	Strategy 9 <ul style="list-style-type: none">• Form 9• Package 9• Ad 9• Special Deal 9

idea of which strategy is the best one, and absolutely no prior information on any of their conversion rates. However we will make the assumption that each of these 9 strategies does have a fixed conversion rate. These strategies were carefully and smartly elaborated by the marketing team, and each of them has the same goal: convert the maximum clients to the premium plan. However, these 9 strategies are all different. They have different forms, different packages, different ads, and different special deals to convince and persuade the clients to subscribe to the premium plan. Of course, the marketing team has no idea of which among these 9 strategies is the best one. But they want to figure it out as soon as possible, and by saving the maximum costs, which one has the highest conversion rate, because they know how finding and deploying that best strategy can significantly maximize the revenues. Also, the marketing experts choose not to send an email

to their 100 million customers, because that would be costly and they would risk spamming too many customers. Instead they will subtly look for that best strategy through online learning. What is online learning? It will consist of deploying a strategy each time a customer browses the online retail business website to hang out inside or buy some products. Then as the customer navigates the website, he or she will suddenly get a pop-up ad, suggesting him or her to subscribe to the premium plan. And for each customer browsing the website, only one of the 9 strategies will be deployed. Then the user will choose, or not, to take action and subscribe to the premium plan. If the customer subscribes, it is a success, otherwise, it is a failure. The more customers we do this with, the more feed-backs we get, and the better we could get an idea of what is the best strategy. But of course, we will not figure that out manually, visually, or with some simple maths. Instead we want to implement the smartest algorithm that will figure out what is the best strategy in the least amount of time. And that's for the same two reasons: first because deploying each strategy has a cost (e.g. coming from the pop-up ad), and second because the company wants to annoy the least customers with their ad.

Q1: By using `environment.ipynb` file, generate the rewards for 10000 customers. Please use the seed as the sum of digits in your enrollment number. In this problem, we deploy bandit based AI agent for maximizing the revenue and identifying the best marketing strategy.

- (a) Plot the histogram for the number of times each marketing strategy was selected when you deploy an AI agent.
- (b) Plot the regret for AI based agent vs a random agent. Assume a random agent picks one of the nine marketing strategies uniformly at random.