Theory of Multi-Armed Bandits

CS 728

Instructor – Gunjan Kumar

What this course is about

- Algorithms for sequential decision making under uncertainty:
  - **News website**: A website chooses a headline to display to a new user and tracks if they click. Goal: Maximize clicks.
  - **Dynamic Pricing** : A store sets a price for a digital product when a customer arrives. The customer then decides whether to make a purchase or not. Goal: Maximize profit.
  - **Investment**: You choose and invest $1 into a stock daily. In the end of the day, you observe the change in value for each stock. Goal: maximize the total wealth.
- We will study a variety of algorithms designed for different uncertain environments to maximize rewards over the duration of play.
- This area intersects significantly with online convex optimization, reinforcement learning.

What the course is about (Continued)

- Course is **theoretical** – we will study mathematical abstraction of the above problems, give algorithms and mathematically analyze their performance.

- Course comes under AI/ML bucket.

What this course is not about

- This is not a course on reinforcement learning – we will study RL in last few lectures only (this course will give a solid background for studying reinforcement learning)

- This course is not about practical aspects of multi armed bandits.

Prerequisites

- Not surprisingly, you will need familiarity with basics of probability and mathematical maturity.

- No previous knowledge in machine learning required.

- I will recap probability in first few lectures (but remember probability is a course in itself, I can only do a quick recap) .

Course Logistics

- Grading:
  - Homework assignments + short project = 70%
  - Mid sem + end sem = 30%
  - For assignments, you can discuss with others (no internet) but **you must write in your own words** and also mention the names with whom you discussed
- Project: either theoretical or implementation, based on your interests
  - Theoretical: you need to read some topics and, in the end, make a presentation
  - Practical: implement some algorithms done in class or any other problem that you find interesting

Book

We will roughly follow this book:

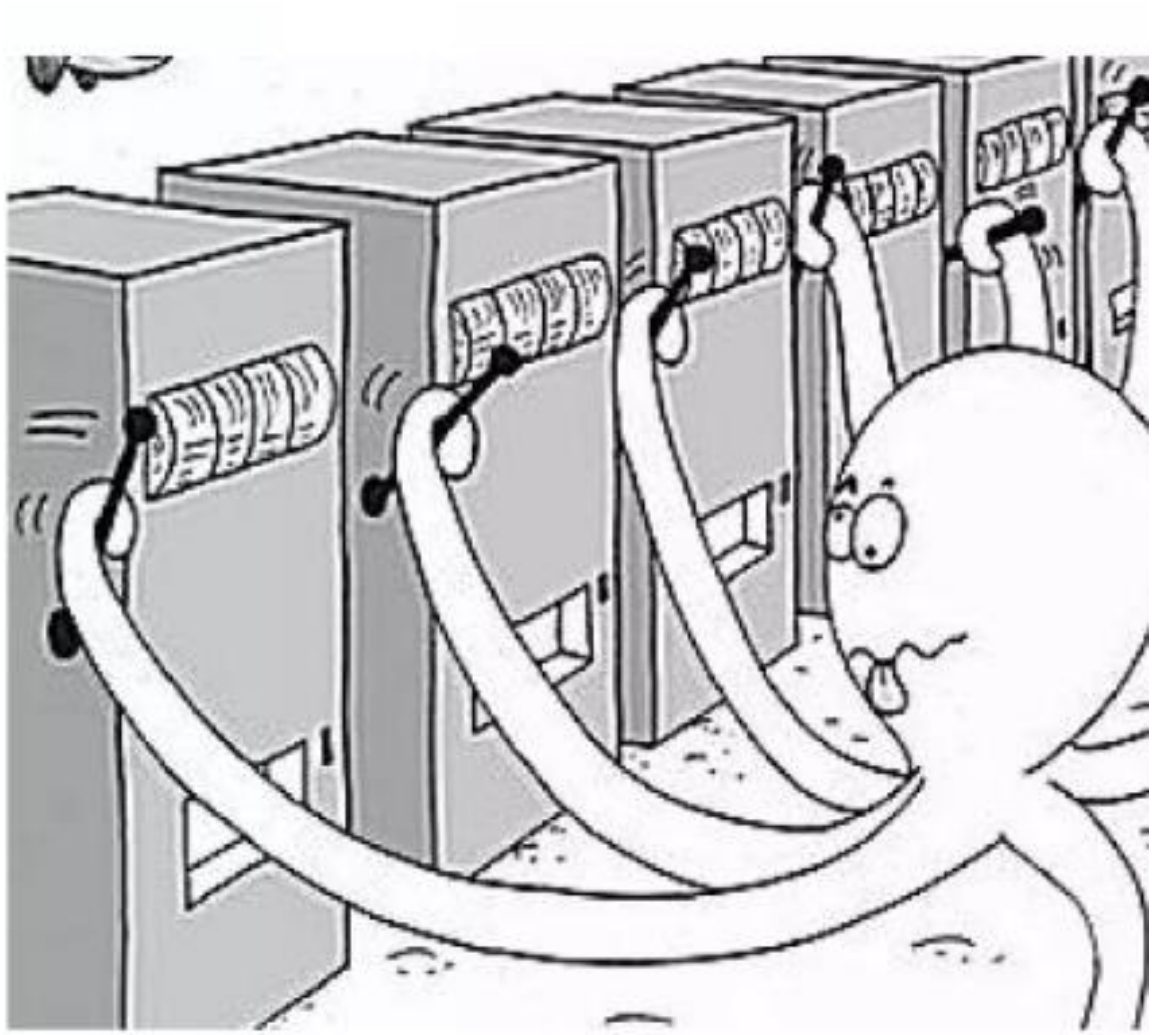
- Introduction to Multi-Armed Bandits by Aleksandrs Slivkins (https://arxiv.org/abs/1904.07272)

Multi-Armed Bandits

- Think of yourself as a gambler standing in front of a row of slot machines, each referred to as an "arm".
- Each time you choose a machine and pull its lever, you receive a payout. The catch?
- You don't know the payout rates of the machines beforehand, and your goal is to maximize your total payout.

Multi-Armed Bandits

Each machine offers a unique chance to win big, but choosing the right one could be the difference between walking away with a fortune or empty pockets.



source: Microsoft Research

Multi Armed Bandits - Examples

Examples	Action	Reward
News Website	An article to display whenever an user arrives (e.g., cricket or bollywood or politics etc.)	1 if user clicks , 0 otherwise
Dynamic Pricing	set price	p if sale, 0 otherwise
Web design	Font colour or page layour	1 if user clicks, 0 otherwise
Recommender systems	Which movie to recommend	1 if recommendation is followed,
Advertisement	Which ads to display	Revenue from the ads
Robot control	Strategy for a given task	Job completion time

Real Life Applications

ONLINE ADVERTISING

RECOMMENDER SYSTEMS

DYNAMIC PRICING

CLINICAL TRIALS

FINANCIAL RISK MANAGEMENT

PORTFOLIO SELECTION

Course Contents

- Stochastic Bandits
- Adversarial Bandits
- Contextual Bandits
- Linear Bandits
- Miscellaneous topics
- Reinforcement Learning