The Multi-Armed Bandit Problem

The Multi-Armed Bandit Problem

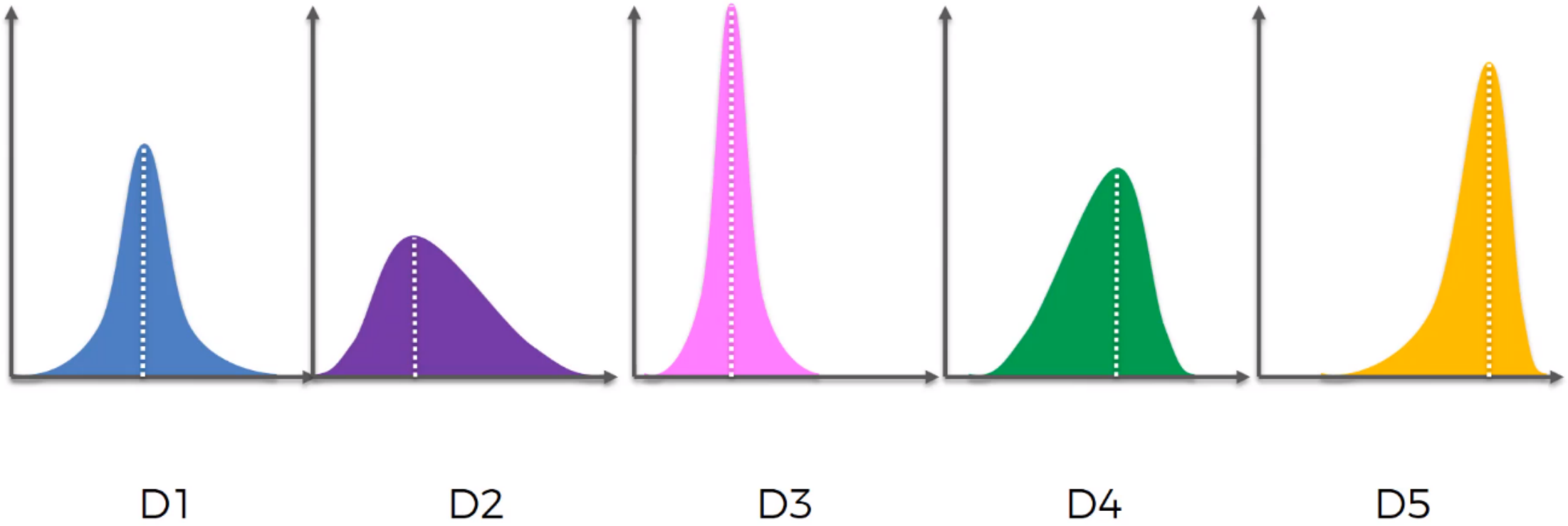
- Reinforcement Learning algorithms:
  - Upper Confidence Bound (UCB)
  - Thompson Sampling

The Multi-Armed Bandit Problem

D1 D2 D3 D4 D5

https://www.udemy.com/course/machinelearning/

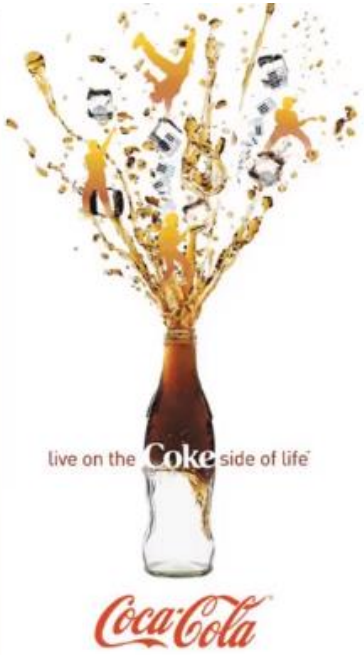
The Multi-Armed Bandit Problem



https://www.udemy.com/course/machinelearning/

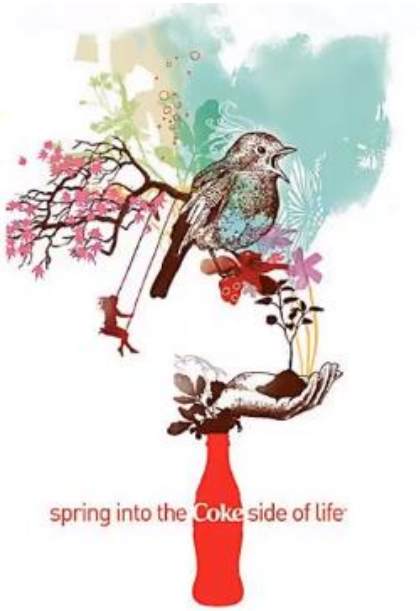
The Multi-Armed Bandit Problem

D1

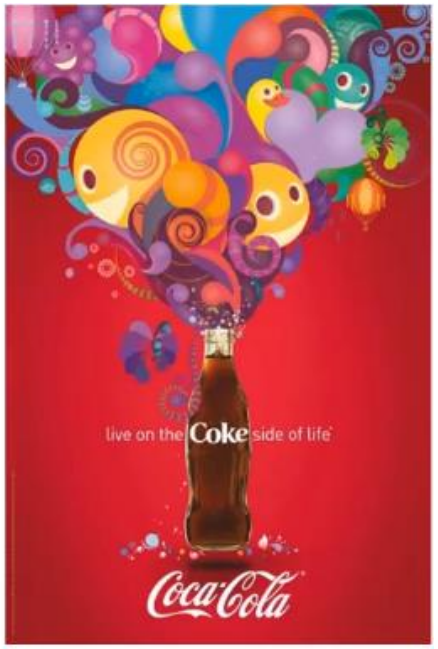


D2

D3



D4



D5

https://www.udemy.com/course/machinelearning/

Notes

- The problem assumes that you have a finite set of machines and there is a distribution which is the best distribution
- It uses **exploration** and **exploitation**
- Tries to the best distribution in order to **exploit** it, but tries to spend the least amount of time **exploring** all of them
- If you don’t spend too much time exploring the machines, you may end up exploiting the non-optimal machine
- It’s better than AB test in the sense that AB test exploits the machines randomly, while UCB tries to exploit the best machine