Workshop 10

COMP90051 Statistical Machine Learning

Semester 1, 2023

Learning Outcomes

By the end of this workshop you should be able to:

1. Be able to implement epsilon-greedy multi-armed bandits
2. Be able to implement upper confidence bound multi-armed bandits
3. Be familiar with offline evaluation of MABs
4. Develop intuition about exploitation vs. exploration

COMP90051 Statistical Machine Learning

Stochastic MAB setting

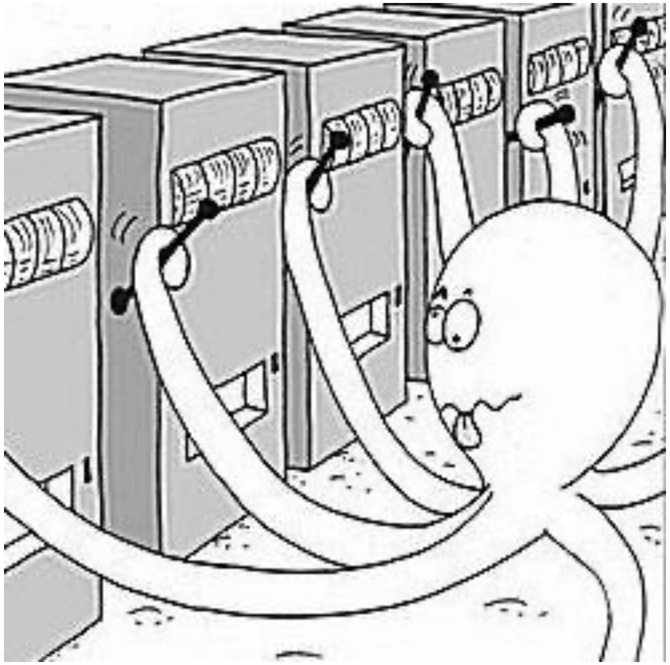
- Possible actions $\{1, \dots, k\}$ called "arms"
  - Arm i has distribution P_i on bounded rewards with mean μ_i
- In round $t = 1 \dots T$
  - Play action $i_t \in \{1, \dots, k\}$ *(possibly randomly)*
  - Receive reward $R_{i_t}(t) \sim P_{i_t}$
- Goal: minimise cumulative regret
  - $\mu^* T - \sum_{t=1}^{T} E\left[R_{i_t}(t)\right]$

    Expected cumulative reward of bandit

    Best expected cumulative reward with hindsight

  where $\mu^* = \max_i \mu_i$
  - Intuition: Do as well as a rule that is simple but has knowledge of the future

Multi-armed bandits



Start

Worksheet 10