The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits

Authors: John Langford, Tom Zhang

Presented by: Ben Flora

Overview

- Bandit problem
- Contextual bandits
- Epoch-Greedy algorithm

Overview

- Bandit problem
- Contextual bandits
- Epoch-Greedy algorithm

Bandits

- K arms each arm i
  - Wins (reward 1) with probability p_i
  - Looses (reward 0) with probability 1- p_i
- Exploration vs. Exploitation
  - Exploration is unbiased
  - Exploitation is biased by exploration only
- Regret
  - Max return – Actual return

Web Example

- Some number of ads that can be displayed
  - Each ad translates to an arm
- Each ad can be clicked on by a user
  - If clicked reward 1 if not reward 0
- Want to have adds clicked as often as possible
  - This will make the most money

Overview

- Bandit problem
- Contextual bandits
- Epoch-Greedy algorithm

Contextual Bandits

- Add Context to the bandit problem
  - Information aiding in arm choosing
  - Helps know which arm is best
- The rest follows the Bandit problem
- Want to find optimal solution
- More useful than regular bandits

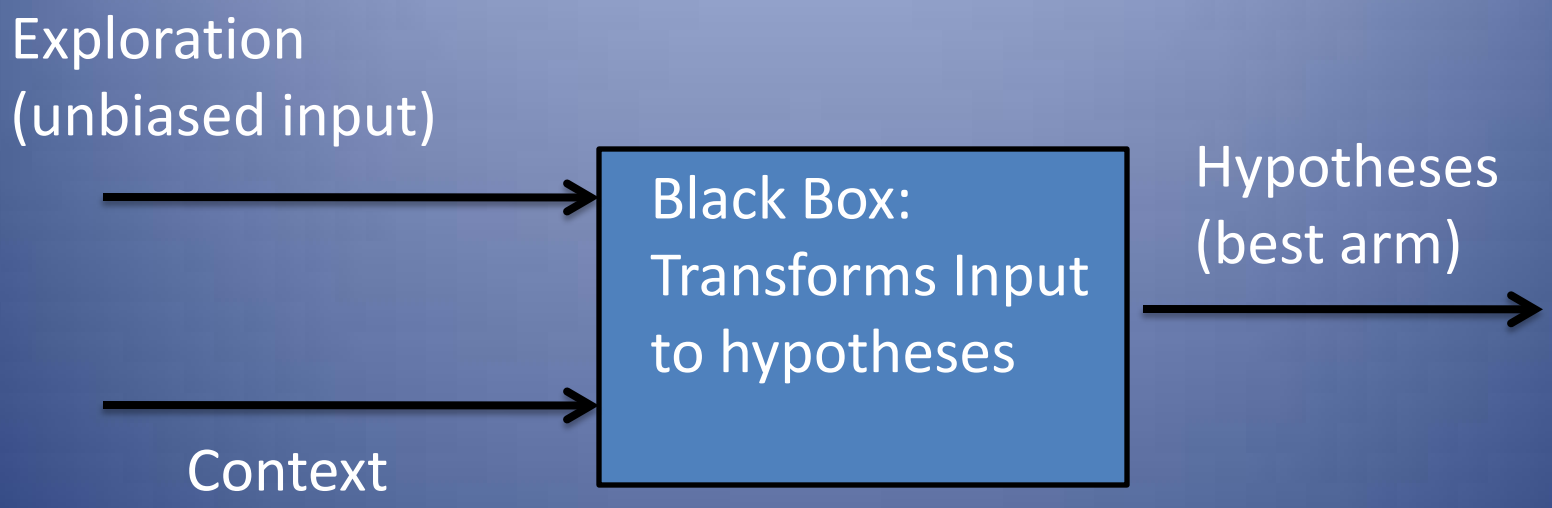
Web Problem

- Now we have user information
  - A user profile
  - Search Query
  - A users preferences
- Use this information to choose an ad
  - Better chance of choosing an ad that is clicked on

Overview

- Bandit problem
- Contextual bandits
- Epoch-Greedy algorithm

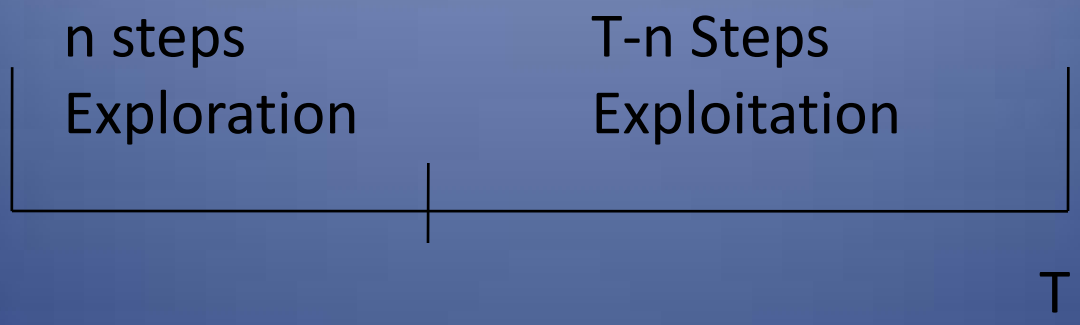
Epoch-Greedy Overview



Similar idea to the papers
we saw on Thursday

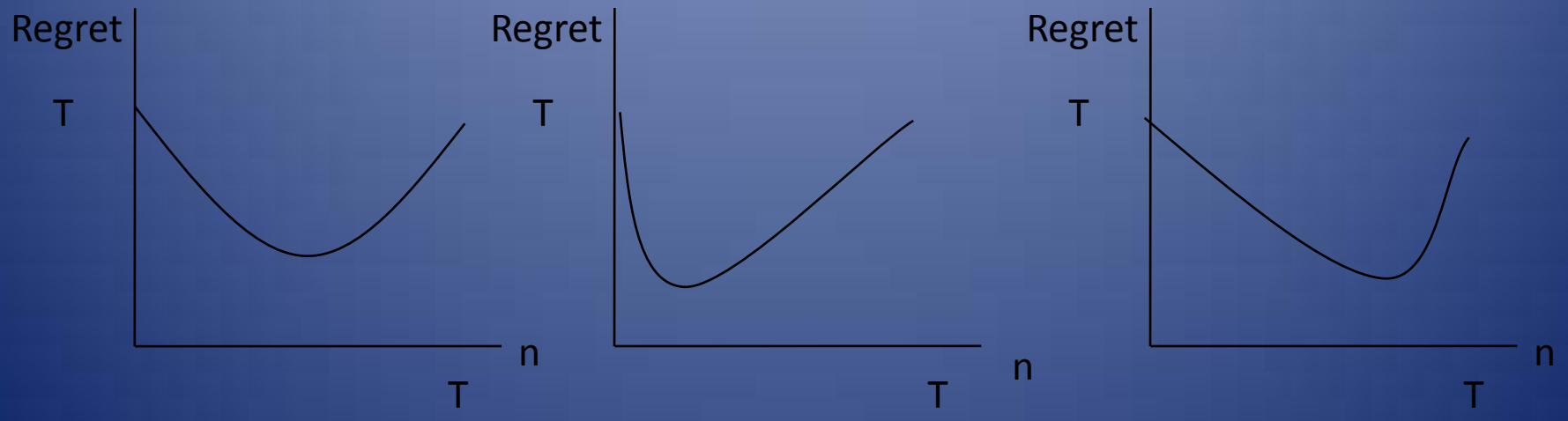
Exploration

- Look at a fixed time horizon
  - Time horizon is the total number of pulls
- Choose a number of Exploration steps



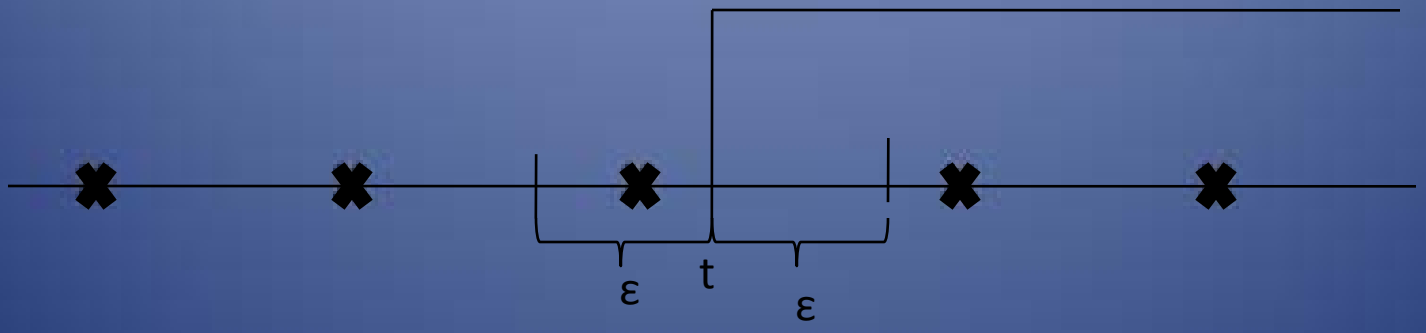
Minimizing Regret

- No explore regret = T
- All exploit regret = T
- Some minimum between those points



Creating a Hypotheses

- Simple two armed case
- Remember binary thresholds
- Want to learn the threshold value



If x < t : pick arm 1

x > t : pick arm 2

Creating a Hypotheses (Cont.)

- Want to be within ε of the threshold
  - Need ≈ O(1/ε)
- As the function gets more complex
  - Need ≈ O((1/ε)*C)
  - C denotes how complex the function is
  - A quick note for those of you who took 156 the C is similar to VC dimension

Epoch

- Don’t always know the time horizon
- Append groupings of known time horizons
  - Repeat until time actually ends
- This specific paper has chosen a single exploration step at the beginning of each epoch

Epoch-Greedy Algorithm

- Do a single step of exploration
  - Begin creating an unbiased vector of inputs to create the hypotheses
  - Observe context information
- Add the learned information to past exploration and create a new hypotheses
  - This uses the contextual data and exploration
- For a set number of steps exploit the hypotheses arm

Review Using Web Example

- Have a variety of ads that can be shown
  - Sports
  - Movie
  - Insurance

Review (Cont)

- Search Query
  - Golf Club Repair
  - Randomly choose
  - Clicked
- Search Query
  - Car Body Repair
  - See Repair and Car
  - Not Clicked

Review (Cont.)

- Search Query
  - Horror Movie
  - Randomly choose
  - Clicked
- Search Query
  - Sheep Movie
  - See Sheep and Movie
  - Clicked