BANDIT ALGORITHM (ONLINE MACHINE LEARNING)

PROF. MANJESH HANAWAL

Department of Industrial Engineering and Operation Research
IIT Bombay

PRE-REQUISITES : Basics of Probability Theory and Optimization

INTENDED AUDIENCE : Computer Sceince, Electrical Engineering, Operations Research, Mathematics and Statistics

INDUSTRIES APPLICABLE TO : All companies related to Internet Technologies (ex. Google, Microsoft, Flipkart, Ola, Amazon, etc.)

COURSE OUTLINE :
In many scenarios, one faces uncertain environments where a-priori the best action to play is unknown. How to obtain best possible reward/utility in such scenarios. One natural way is to first explore the environment and to identify the `best' actions and exploit them. However, this give raise to an exploration vs exploitation dilemma, where on hand hand we need to do sufficient explorations to identify the best action so that we are confident about its optimality, and on the other hand, best actions need to exploited more number of times to obtain higher reward. In this course we will study many bandit algorithms that balance exploration and exploitation well in various random environment to accumulate good rewards over the duration of play. Bandit algorithms find applications in online advertising, recommendation systems, auctions, routing, e-commerce or in any filed online scenarios where information can be gather in an increment fashion.

ABOUT INSTRUCTOR :
Prof. Manjesh hanawal received the M. S. degree in ECE from the Indian Institute of Science, Bangalore, India, in 2009, and the PhD degree from INRIA, Sophia Antipolis, France, and the University of Avignon, France, in 2013. After two years of postdoc at Boston University, he is now an Assistant Professor in Industrial Engineering and Operations Research at the IIT Bombay, India. His research interests include performance evaluation, machine learning and network economics. He is a recipient of Inspire Faculty Award from DST and Early Career Research Award from SERB.

COURSE PLAN :

Week 1**:** Introduction to Bandit Algorithms. From Batch to Online Setting

Week 2**:** Adversarial Setting with Full information (Halving, WM Algorithm)

Week 3**:** Adversarial Setting with Bandit Information

Week 4**:** Regret lower bounds for adversarial Setting

Week 5**:** Introduction to Stochastic Setting and various regret notions

Week 6**:** A primer on Concentration inequalities

Week 7**:** Stochastic Bandit Algorithms UCB, KL-UCB

Week 8**:** Lower bounds for stochastic Bandits

Week 9**:** Introductions to contextual bandits

Week 10**:** Overview of contextual bandit algorithms

Week 11**:** Introduction to pure exploration setups (fixed confidence vs budget)

Week 12**:** Algorithms for pure explorations (LUCB, KL-LUCB, lil'UCB)