Reducing contextual bandits to supervised learning

Daniel Hsu
Columbia University

4/27/2016 at 11:00 am
Core A (Room 301)

Abstract

I’ll describe a fast and simple algorithm for the contextual bandit learning problem, where the learner repeatedly takes an action in response to the observed context, observing the reward only for that action. The algorithm assumes access to an oracle for solving cost-sensitive classification problems and achieves the statistically optimal regret guarantee with only about $T^{\frac{1}{2}}$ oracle calls across all $T$ rounds. Joint work with Alekh Agarwal, Satyen Kale, John Langford, Lihong Li, and Rob Schapire.