Zhuoran Yang

Title: Reinforcement Learning Meets Bilevel Optimization: Learning Leader-Follower Games with Sample Efficiency

Abstract: In this talk, I will introduce methods that modify the optimism principle for reinforcement learning in leader-follower games, especially when the follower's reward function is unknown. Such problems generally face statistical challenges due to the ill-posed nature of the best response function. I will discuss two cases that overcome these challenges. The first involves a fully rational follower with a separable reward function, where we use an algorithm combining optimism with pessimistic binary search to identify the follower's indifference curve. In the second case, for a boundedly rational follower defined by entropy regularization, we directly estimate the response model and establish a bonus function for estimation uncertainty. This approach leads to optimism-based online reinforcement learning algorithms that achieve sublinear regret upper bounds, effectively learning the leader's optimal policy in both scenarios.