Discussion on "Statistically Efficient Offline Reinforcement Learning"

Chengchun Shi

London School of Economics and Political Science

Summary

- **Theoretically**, derive semiparametric efficiency bounds for off-policy value and policy gradient.
- **Methodologically**, propose efficient off-policy policy evaluation and policy gradient algorithms.
- **Empirically,** show the advantage over classical policy evaluation and policy gradient algorithm.

Contributions

- Reinforcement Learning + Semiparametric Efficiency Theory
- Efficient Off-Policy Policy Evaluation and Policy Gradient Algorithm
  1. Break the Curse of Horizon
  2. Obtain Optimal Rate of Convergence
  3. Achieve Minimum Variance
- Offline Reinforcement Learning
  1. Expensive Data Collection
  2. Sample Efficient Algorithm

Questions

- Challenge in Learning Nuisance Functions
  1. Marginal Density Ratio μ
  2. Derivatives of Q-, μ- and Value functions
  3. Parameters Tuning

- Limitations of Policy Gradient
  1. Local Optimum
  2. Magnitude of Gradient around the Optimal Policy

Thanks!