CSC2626: Assignment 2
Due April 5 at 6pm ET
25 points

March 15, 2021

Instructions

The goal of this assignment is to help you get familiar with offline reinforcement learning, and in particular with the implementation of Conservative Q-Learning `https://arxiv.org/abs/2006.04779`.

Provide a summary, up to 2 pages, of the main ideas and assumptions in the paper. Ensure that you provide a description of the main equations and theorems and describe the significance of the experimental results with respect to related baseline methods. In addition, make sure to address the following questions in your summary, by looking at the reference implementation[1].

- Which equation in the paper was used to implement the CQL update rule in the reference implementation?
- How was this equation derived in the paper?
- How is the equation above implemented for discrete systems and continuous systems?
- Do the CQL lower bounds hold upon convergence of the CQL update rule, or also during each Q iteration?
- How is the policy learned? Is there another way it could have been computed?
- What directions for future work would you consider in order to improve, extend, or make use of this method or the lower bounds that it guarantees?

Submit a file called `a2_firstname_lastname_studentid.pdf`, containing your summary, on Quercus.

[1] `https://github.com/aviralkumar2907/CQL/blob/master/d4rl/rlkit/torch/sac/cql.py`