Home work 1: Taxi

Author: Novokshanov Roman

Task 1: Taxi with CrossEntropy algorithm

Description

The Cross Entropy agent has been trained with different parameters set:

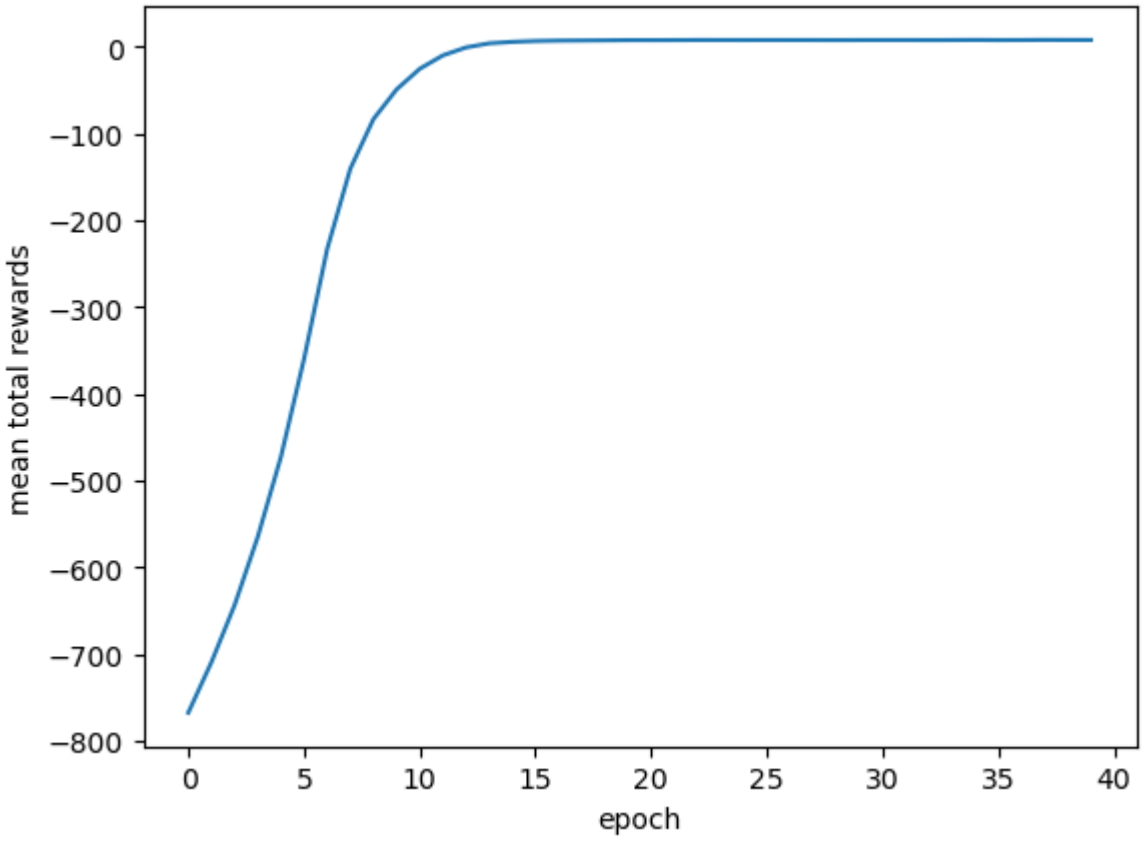
- q_param = [0.5, 0.6, 0.7, 0.8, 0.9]
- trajectory_len = [100, 200, 400, 800, 1000, 2000, 4000]
- trajectory_n = [100, 200, 400, 800, 1000, 2000, 4000]
- iteration_n = [10, 20, 30, 40]

The largest mean total rewards has been achieved at `8.01` value with the following parameters:

- q_param = 0.5
- trajectory_len = 200
- trajectory_n = 4000
- iteration_n = 40

Experiments

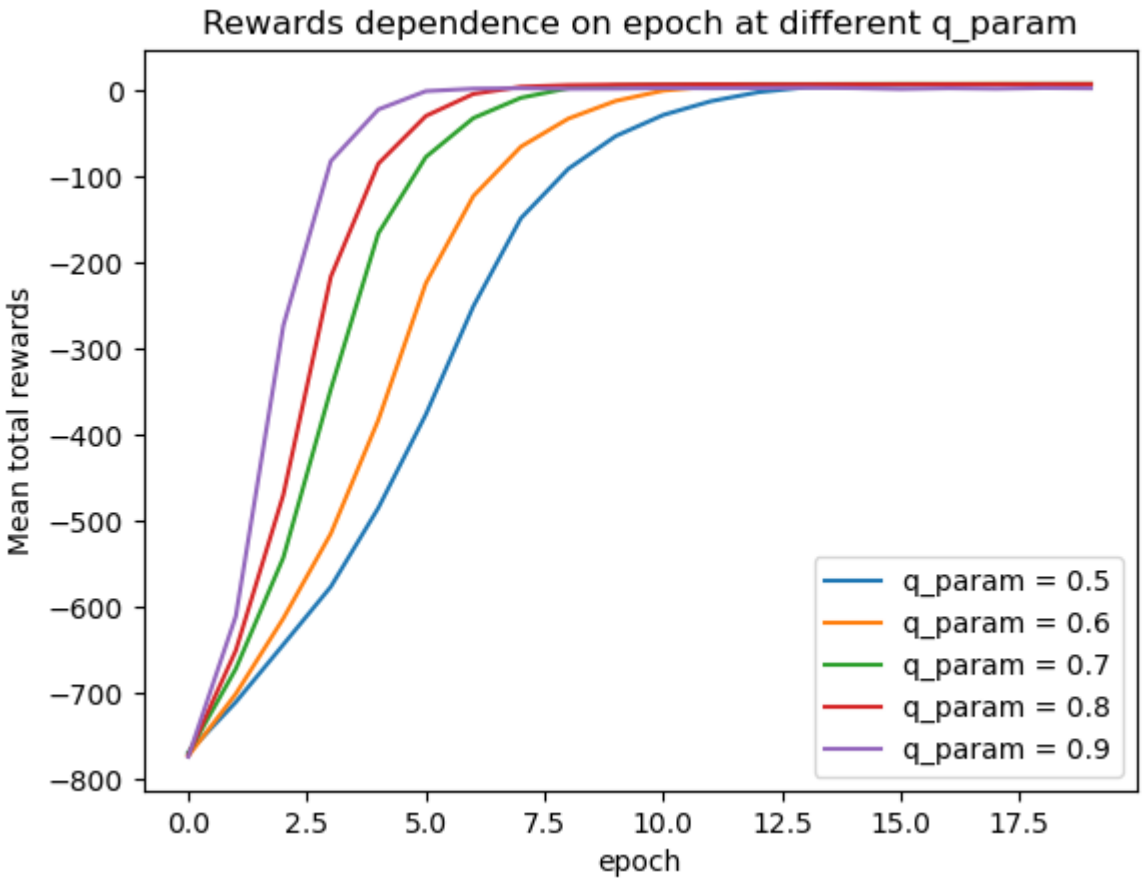
Learning curve (max mean total rewards at 8.01 achieved at 37 epoch):



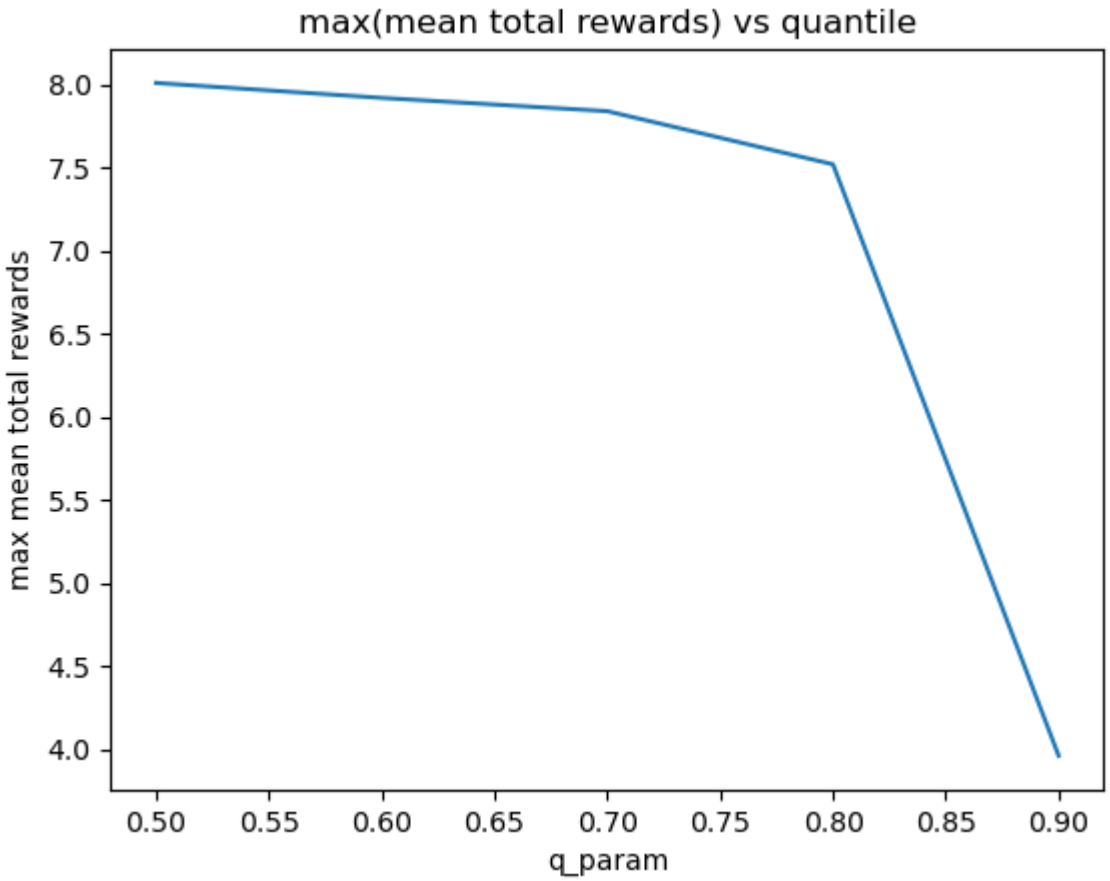
Mean total rewards dependency on epochs at different quantiles:

Other parameters are fixed at:

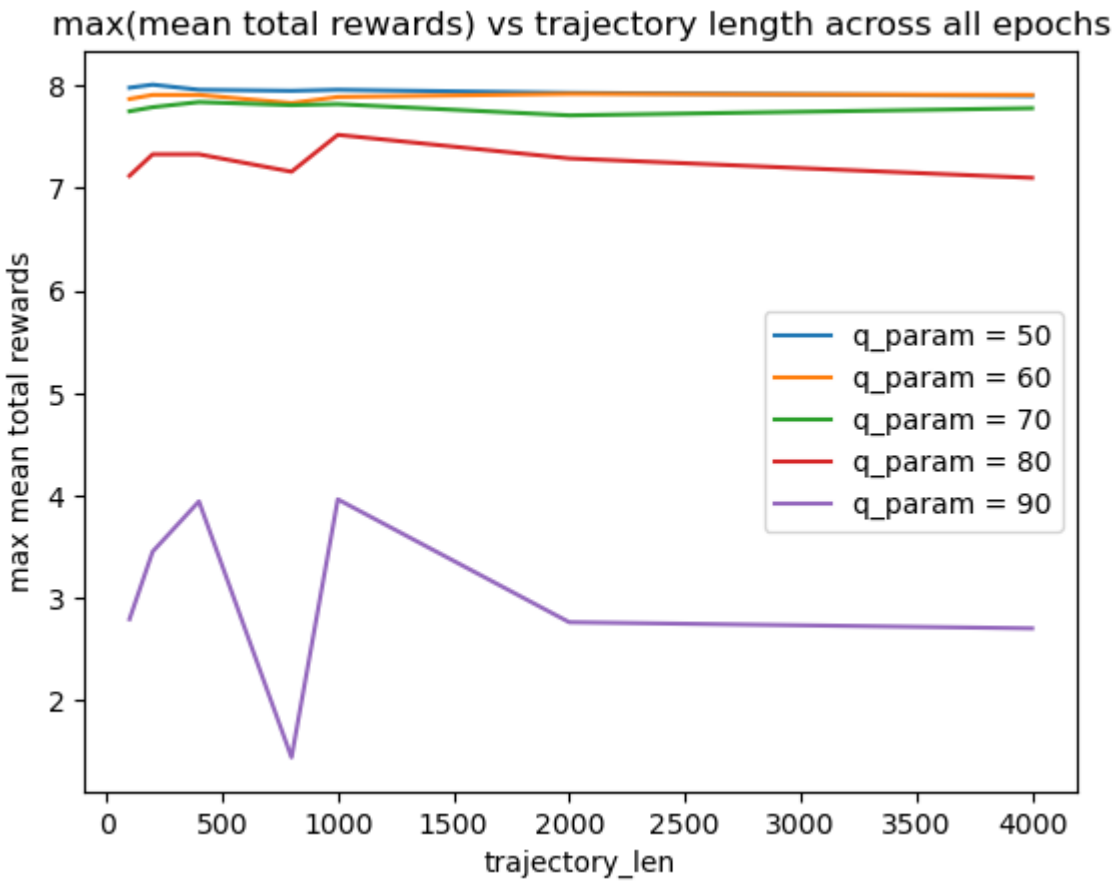
- iteration_n = 20
- trajectory_len == 4000
- trajectory_n == 4000



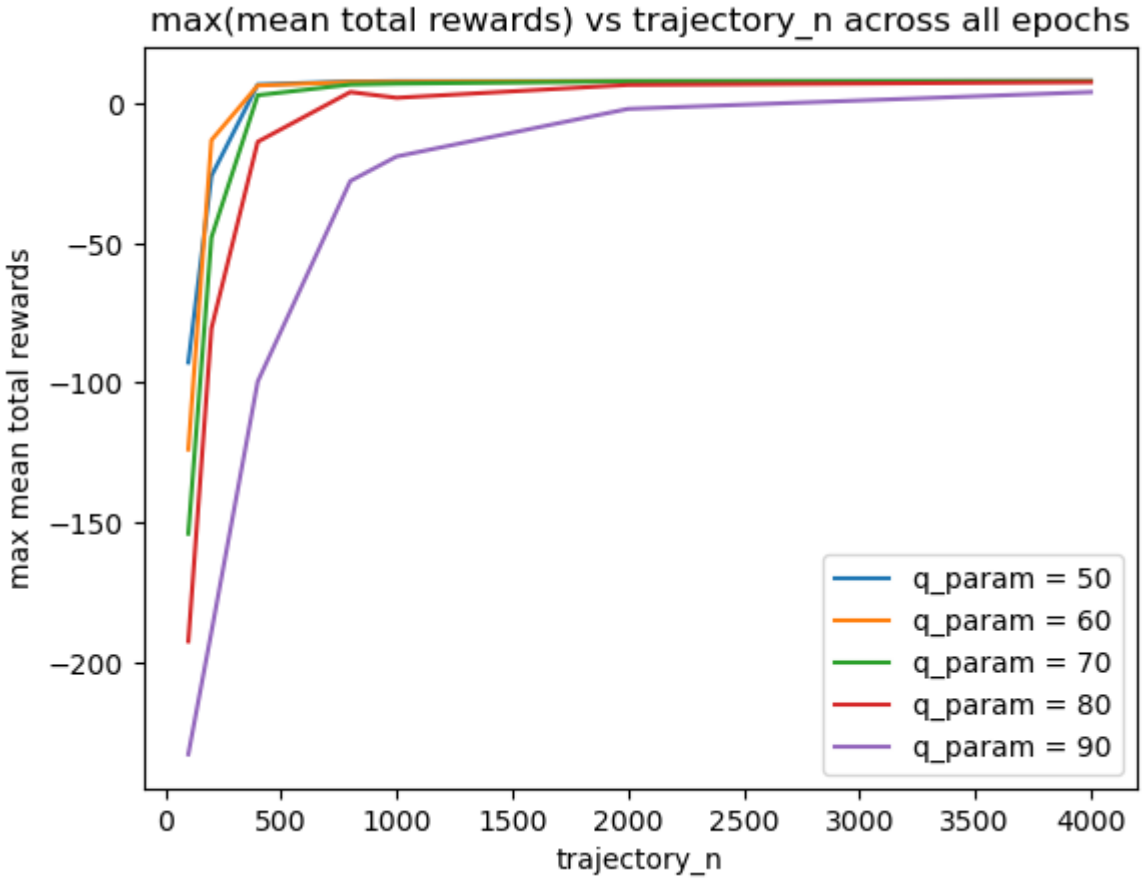
Max mean total rewards dependency on quantile across all epochs:



Max mean total rewards dependency on trajectories length (trajectory_len) across all epochs:



Max mean total rewards dependency on trajectories quantity (trajectory_n) across all epochs:



Conclusion

1. The Cross Entropy agent learns relatively fast. It starts getting mean total rewards above zero after 30-40 steps.
2. Higher quantile requires larger number of epochs.
3. Long trajectories could accumulate larger negative total rewards before it receives positive reward hence are likely to be excluded from pool of elite trajectories at high quantiles.
4. Lower quantiles produce higher mean total rewards. It can be explained by the fact that low quantiles would have larger pools of elite trajectories and include long trajectories with large negative rewards.
5. Trajectories length might be 2-3 times the number of states. After that length of trajectories does not affect much the mean total rewards.