FDR for a1true-a2false-b1false-c1false, reward function 1 8.0 bandit-opt bandit-eps 0.6 bayes-high bayes-low FDR mlfq random 0.4 qlearning tableau ann 0.2 0.0 50 200 0 100 150 250 300 350 400

experiments run