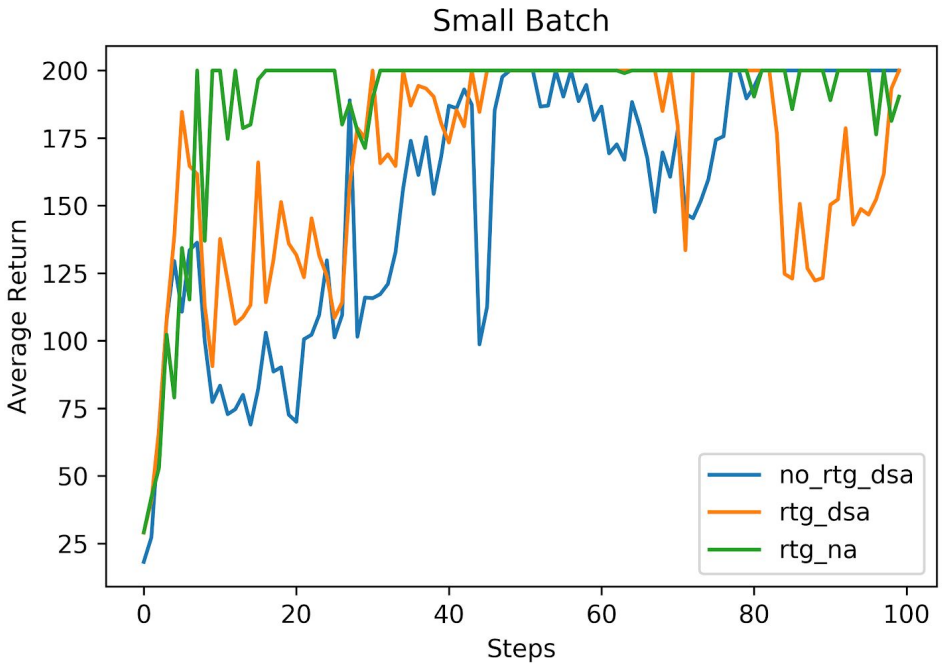
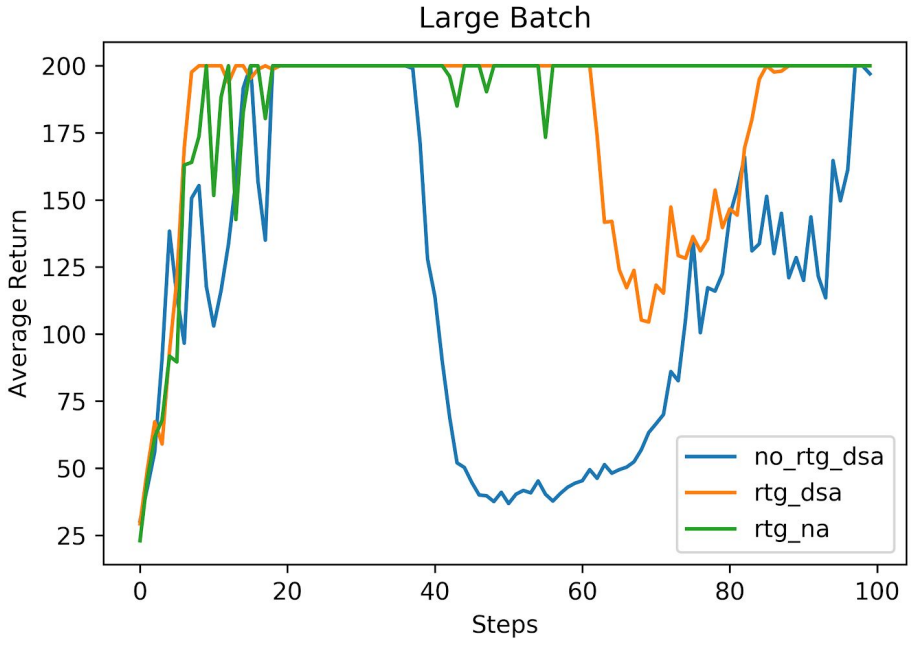
Deep RL hw2 report

(Scripts executed exactly as they appear in the problem formulations. Parameters are the same)

Problem 3

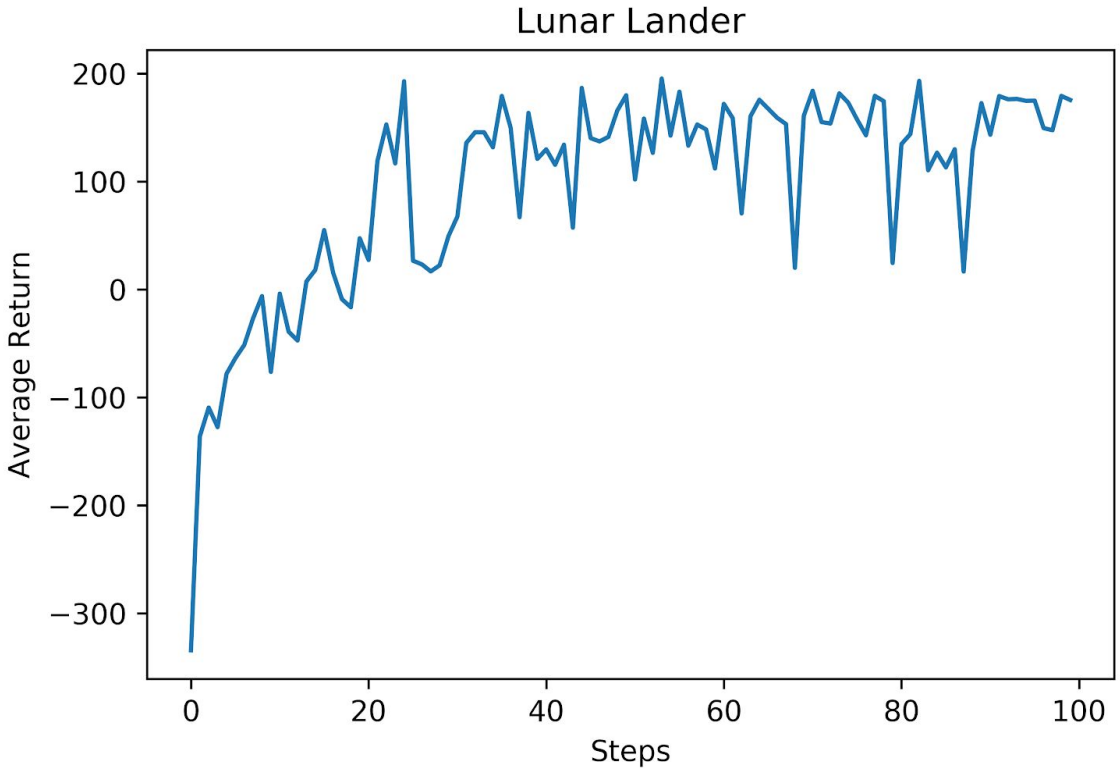
- Using reward-to-go improves convergence and stability without advantage-standardization.
- Advantage-standardization stabilizes the performance
- Batch size does not seem to have any impact in the presence of both advantage-standardization and reward-to-go





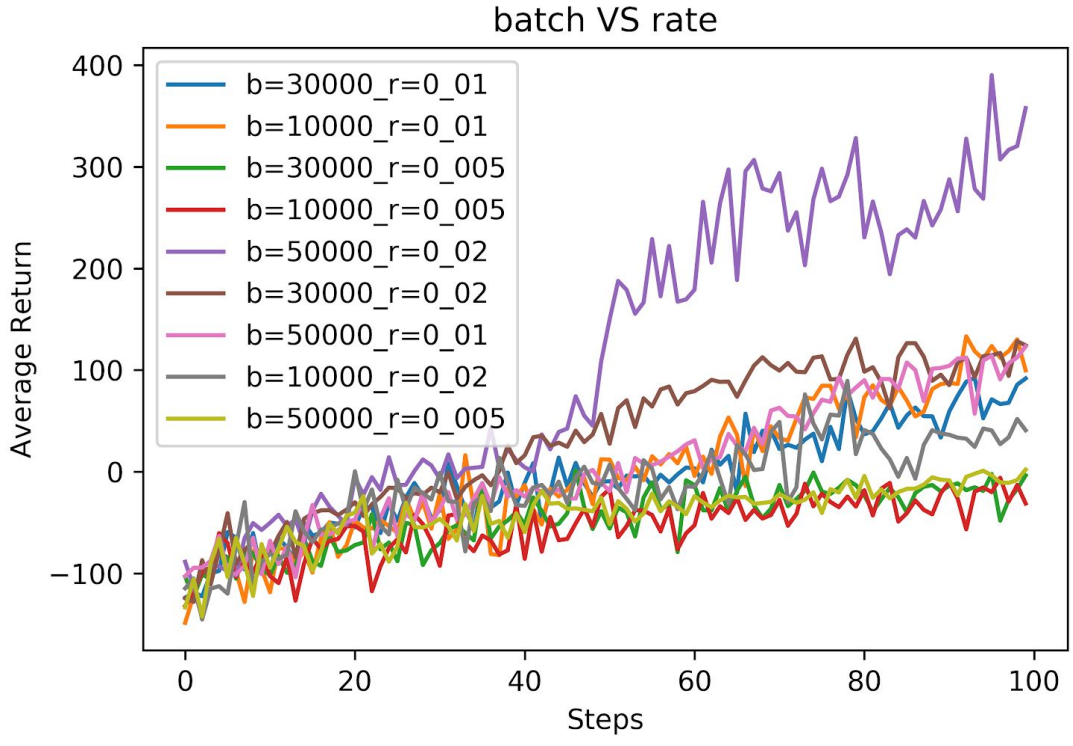
Problem 4 <b*> = 400; <r*> = 0.07

Problem 6



Problem 7.a

In this parameters set, the performance increases with both the batch size and the learning rate



Problem 7.a

<b*> = 500000

<r*> = 0.02