Home assignment 3

Task 2

The lander **was not** able to learn any useful behavior. Average number of timesteps does increase but really slowly (~ 0.03 per 200 episodes).

Reasons:

- Discretization is not effective with high dimensional space in the lunar lander problem
- Q-learning in this case does not take neighbor's state into consideration

Task 3

Yes (did not make it on time to 5000 episodes but got many episodes with positive total rewards after ~ 1200 episodes)