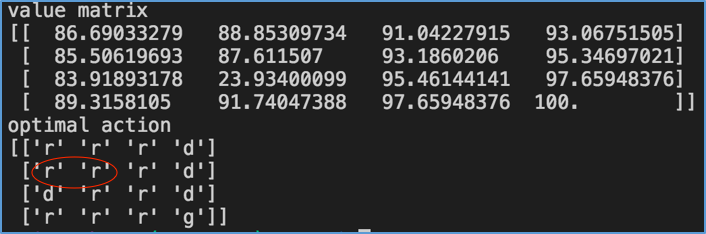
Ruixue Zhang

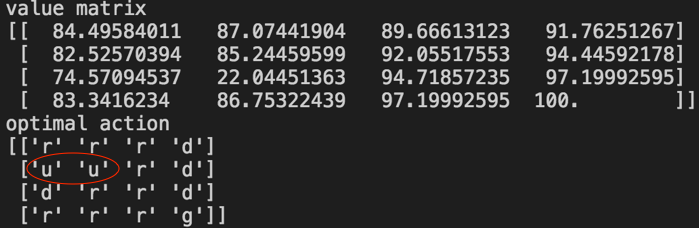
20619404

Problem 1

a = 0.9, b = 0.05



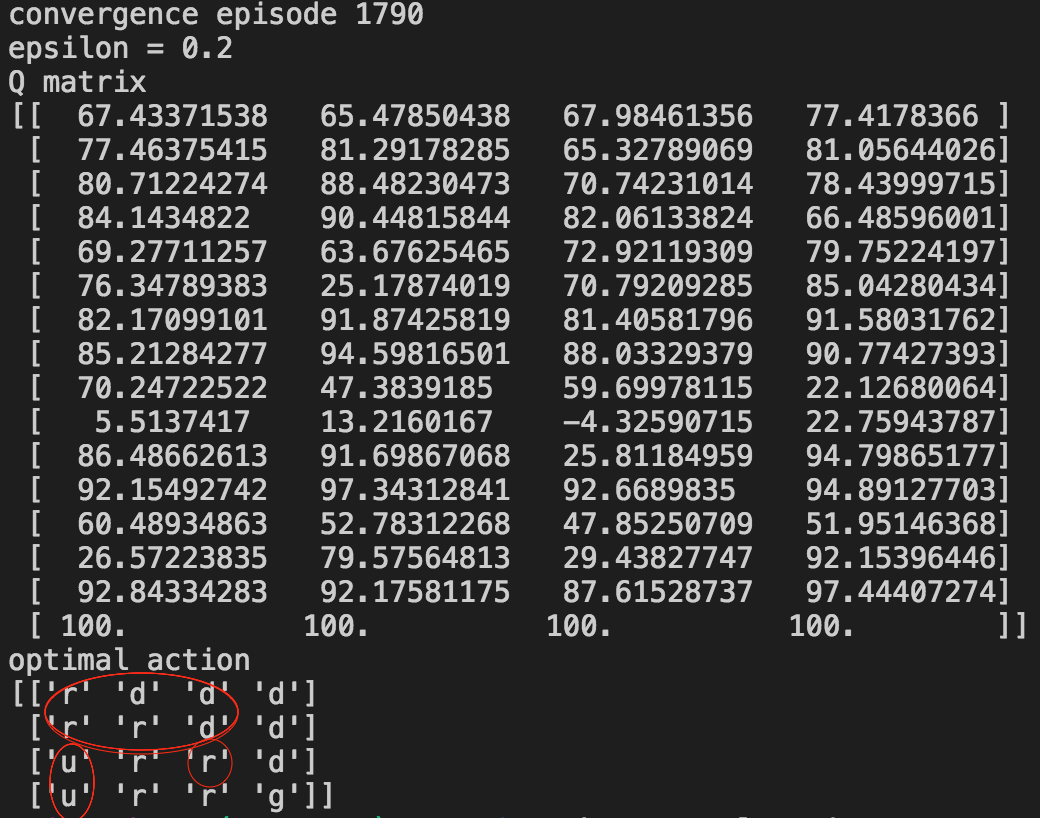
a = 0.8, b = 0.1

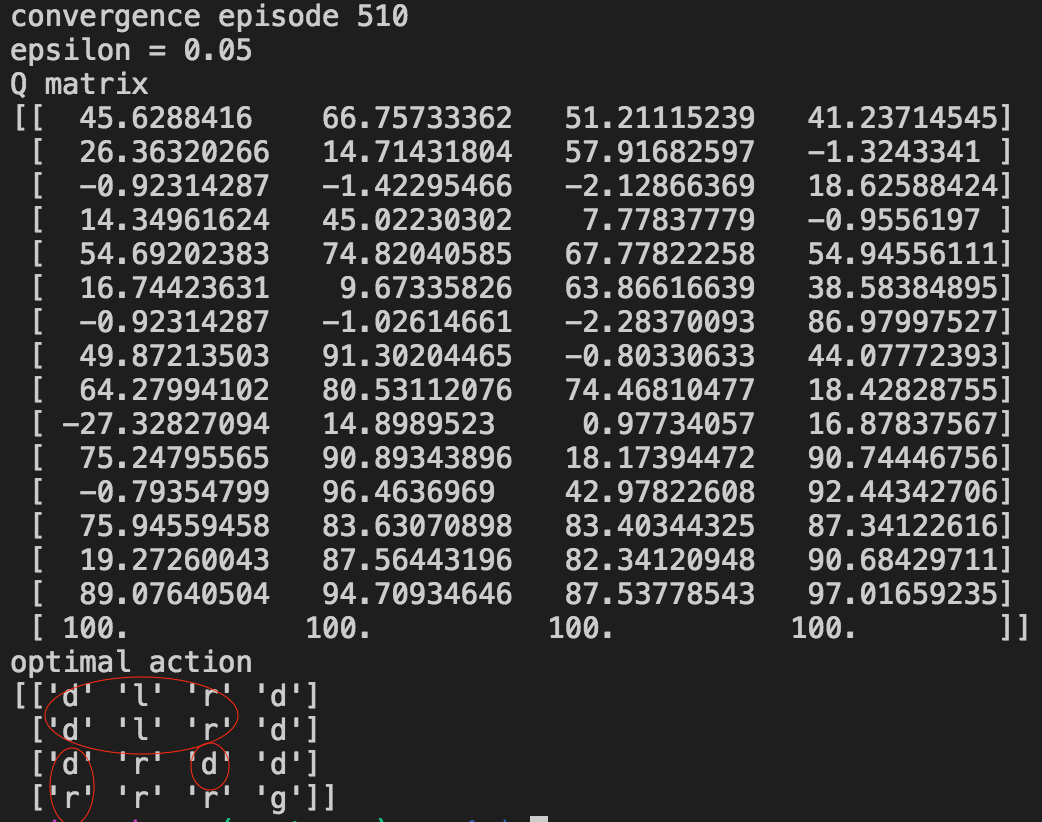


Small a value means a low probability to execute an expected action. Difference of optimal action is marked in the two figures. When we approach the bad state, we need to get rid of the bad state in case that we drop into it and get the worst reward.

In the whole, the value will be small when a is small. Since when it is definitely move to intended location, it will get the location with highest value. When a decreases, the probability for intended move gets lower and the total value decreases. Also, it may choose other orientation with lower value.

Problem 2





Episodes needed to convergence is shown in the two figures. Large epsilon value indicates large probability to randomly choose action. This helps to do exploration but will consume more time to get convergence.

The difference of primal action is marked in red circles. With small epsilon to choose random action, the primal action is hard to find out and some actions to go back may be chosen. The training with larger epsilon will tend to face to the goal state to avoid negative reward for too many steps. Also primal action also tends to get away from bad state when epsilon is small because next step may be randomly chosen and it is dangerous to keep close to bad state.

Problem 3

Experience replay and the target network are of almost equal importance to the deep Q network. It seems only one of them won’t be much useful to train the network with help of another. The first figure is the reward per episode respect to the episode. The second one is another figure under a modified reward (time steps per episode with respect to the episode for 500 episode). The cart pole can be quite balanced under the changed reward.

