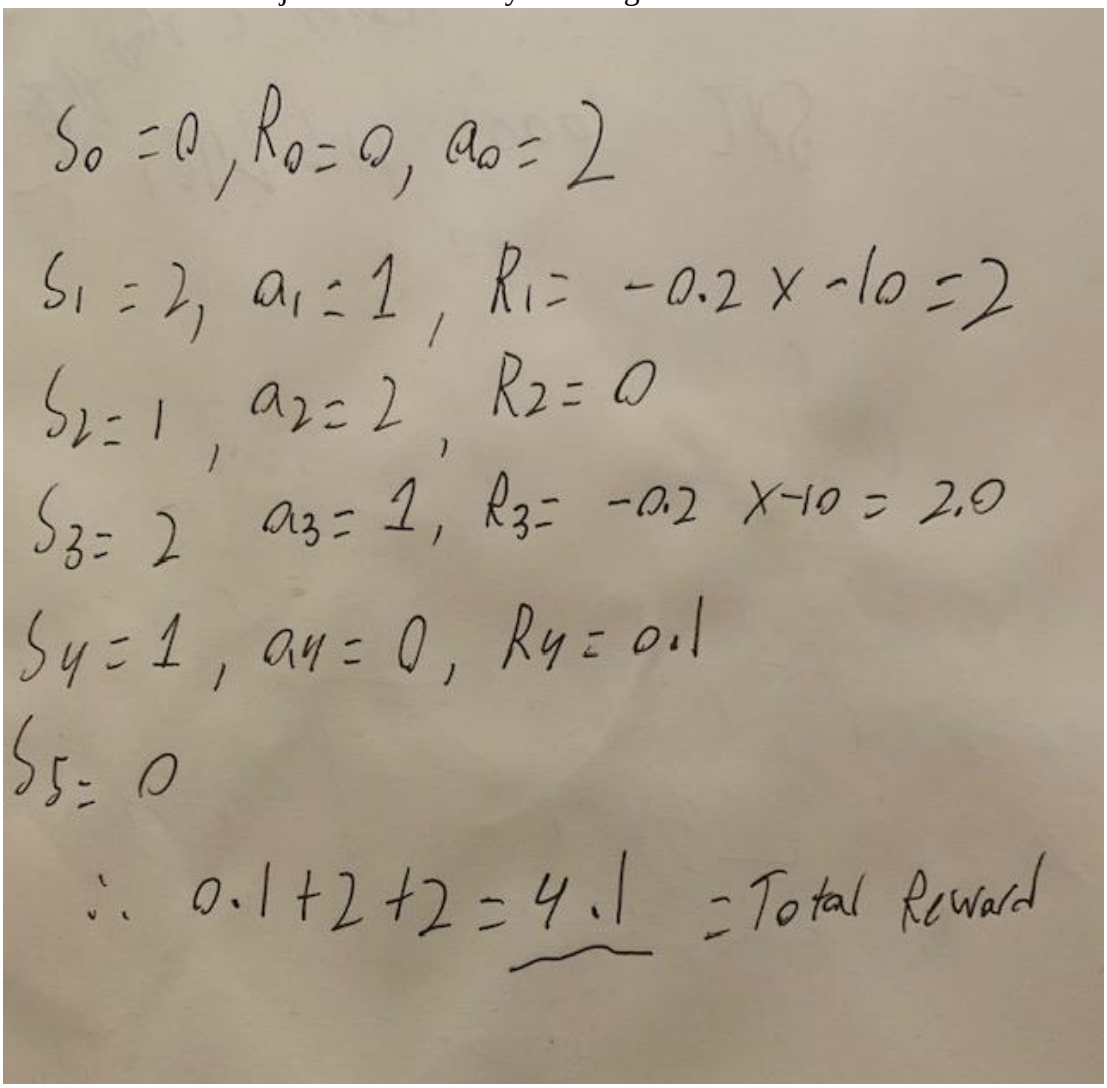
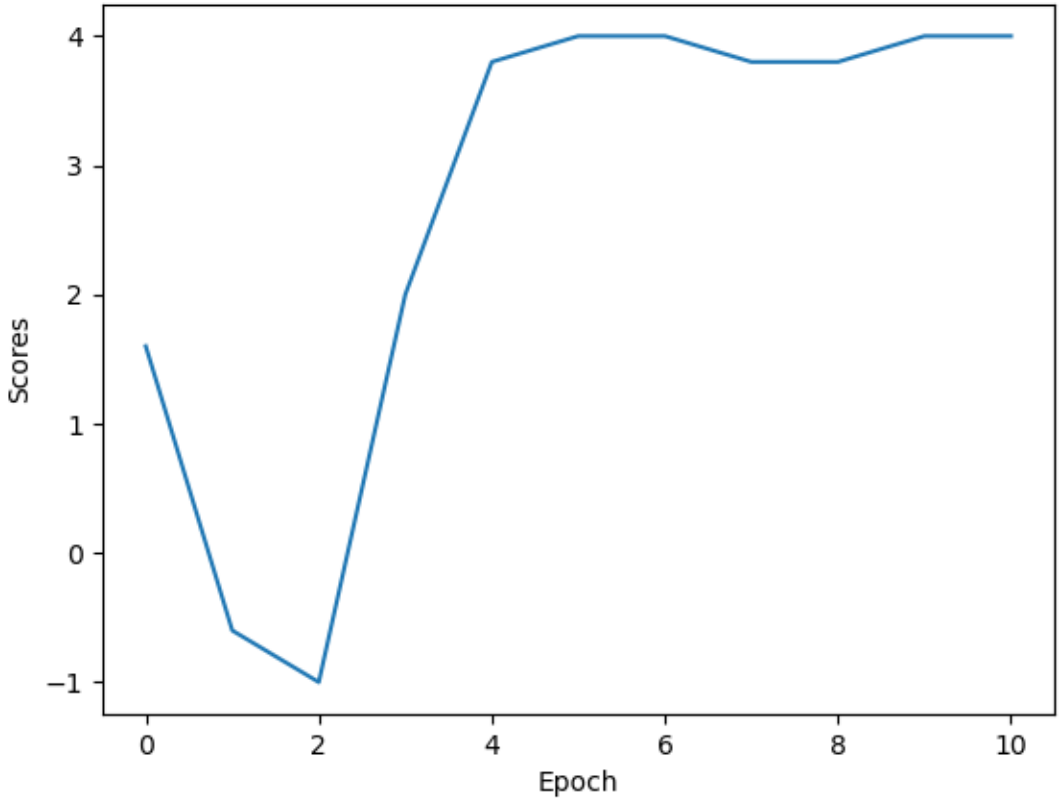
Reinforcement Learning – Assignment 3

1. A) The following trajectory has been calculated below; the maximum sum would be 4.1. No other can get a higher reward is since the max reward we can get is if in state 2, you take action 1 which yields a reward of 2.0. You can do this two times since there are only 5 actions per episode. The remaining actions in the last episode can yield a reward of 0.1. There are no other trajectories that can yield a higher reward.

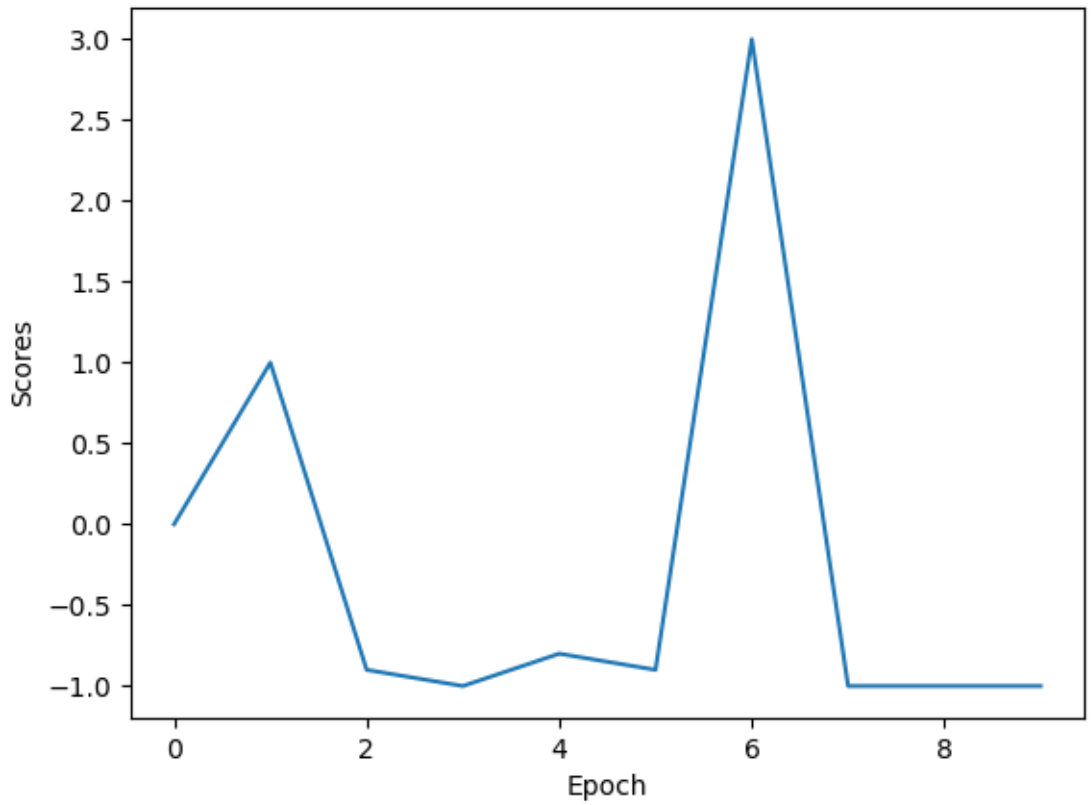


2. A) Representing the Q function as such makes it easy to link our state with the available actions. It is a faster way to get all possible rewards for all actions of one specific state instead of returning all rewards for each state manually every time.
   B) All 3 steps pass every time, no extra work was needed.

3. A) I reach right under the optimal reward at around 4.0, sadly I could not get it right up to 4.1 but was very close, not sure as to why. I have provided the generated graph below:



B) The deep-Q network does not run as well as the linear approximation network. This could be do to randomness or lack of training. I noticed the deep-Q network takes a long time on my machine to run as well. I have attached the result graph below:


3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
Scores
0
2
4
6
8
Epoch