Lab 4

Task 2

For task 2, we discretized the angle in to 20 pieces, between –Pi, and +Pi. We then defined the reward as Pi - |angle|, this meant that a greater angle was a lower reward, and 0 angle gave the maximum reward of Pi.

For the action function, we defined 4 actions. They were: Nothing, Left rocket, Right rocket, Middle rocket.

We then implemented the update Q function based on the slides from the lectures and the Wikipedia algorithm.

This resulted in a great result, where the rocket was able to keep a steady angle near 0.

Task 2 Question 1

The value for the Q-function is based on the current value plus the future value adjusted for the learning rate. The future value is the previous reward plus the maximum value attainable from the next action, adjusted for the discount factor minus the current value. This gives a fair estimate of the future values in the table.

Task 2 Question 2

If we turn of exploration at the start, the Q table has not explored, and it does not know how to get a high reward. It simply does not work.

Task 3

For task 3, we discretized the states based on horizontal velocity, vertical velocity, and angle. We had 4 values for each velocity, and 20 values for the angle. This gave a lot of granularities to the angle, but not as much to velocity, since it is not as important.

For the reward function we created three separate rewards for the velocities and the angle. The velocities were defined as 1 / (1 + |velocity|), and the angle was defined as previously. We then added these values together.