TDDC17 ARTIFICIAL INTELLIGENCE

Lab 5: Reinforcement Learning

Robin Andersson (roban591)
Lawrence Thanakumar Rajappa (lawra776)

October 16, 2019

Part 2

1. In the report, a) describe your choices of state and reward functions, and b) describe in your own words the purpose of the different components in the Q-learning update that you implemented. In particular, what are the Q-values?

a) The state function is split into ten discrete pieces from angles below -2 to angles above 2. The different states the engines can be in are: none active, left active, right active, middle active and all active. The reward function gives a zero when the angle is above 2 and below -2, otherwise the following equation is used:

$$\left(1 - \frac{|\phi|}{K}\right)^2 * K \tag{1}$$

where ϕ is the current angle of the controller and K is the maximum angle that we use for the discretization. The reason we chose this formula is to make the angles around zero give a high reward and to make it just ignore angles that are nowhere near 0.

b) When we update the Qtable we use the following formula

$$Q(s,a) = Q(s,a) + \alpha(R(s) + \gamma \max_{a'} Q(s',a') - Q(s,a)),$$

where

- $Q(s,a)$ represents the Q-value of the previous state and action.
- α is calculated using the N-value of the previous state and action.
- $R(s)$ represents the reward from the previous action.
- γ is a constant.
- $Q(s',a')$ represents the Q-value of the new state and a new action where a' is chosen as the action that gives the highest Q-value.

The Q-values are used by the agent to decide which action to take next and the action that is chosen is the one with the highest Q-value. The Q-value thus tells the agent which action will yield the highest reward.

2. Try turning off exploration from the start before learning. What tends to happen? Explain why this happens in your report.

The rocket tends to spin in a circle while falling down. This is because it does not have a complete Qtable that tells it what to do to get a high reward. Therefore it tries to get a high reward which it does ocassionaly when the angle is close to zero.

Part 3

The angle controller is the same as in part 2. The vy controller is split into 8 discrete states from below -2 to above 2. The vx controller is split into 4 discrete states from below -2 to above 2. The reward function is the sum of equation (1) for all of the different controllers where ϕ now is the value of vx and vy respectively and K is the maximum value of vx and vy respectively.