Reinforcement Learning
Exercise 7

Mathias Niepert, Vinh Tong

April 28, 2023

1 Linear function approximation (4P)

a) Show that tabular methods are a special case of linear function approximation. What does the feature vectors look like? (2P)

b) Give the update rule for Sarsa (λ) for: (2P)

- the tabular case
- with function approximation (general case)
- with linear function approximation

2 Mountain Car (8P)

The code template can be found on Ilias in *ex07-fa/ex07-fa.py*. It has been tested with gym version 0.18.0 (but should also be stable with version 0.18.3).

For this exercise we will use the MountainCar environment from gym: https://www.gymlibrary.ml/environments/classic_control/mountain_car/

Starting from the bottom of a valley, an underpowered car has to gain enough momentum to reach the top of a mountain. The objective is to minimize the number of time steps to reach the goal. There are three possible values of action a:

- full throttle reverse / accelerate to the left (0)
- zero throttle / don't accelerate (1)
- full throttle forward / accelerate to the right (2)

The continuous state space is defined by $x_t = (x_t, \dot{x}_t)$, where the bounded state variables $x_t \in [-1.2, 0.6]$ and $\dot{x}_t \in [-0.07, 0.07]$ are respectively the position and velocity of the car. The reward in this problem is -1 on all time steps. An episode terminates when the car moves past its goal position $x_{t+1} \geq 0.5$ at the top of the mountain, or when the episode length is greater than 200.

a) Implement Q-learning by using state-aggregation (e.g., 20 intervals for x and 20 for $\dot{x}$). Plot the value function at regular intervals (e.g., every 20 episodes). (3P)

b) Repeat the process 10 times; plot the averaged cumulative number of successes (reaching goal state), and the averaged number of steps per episode (y-axis) against number of episodes (x-axis). Analyze the learning curves. (2P)

c) Implement Sarsa with linear function approximation (e.g. tile-coding or RBFs) and compare the learning curves against the ones from b). (3P)