# Programming Assignment 2: CS 747

Md Kamran

Due: 11 October 2021

## Contents

# Task 1

Epsilon = 0.01
Learning rate = 0.15

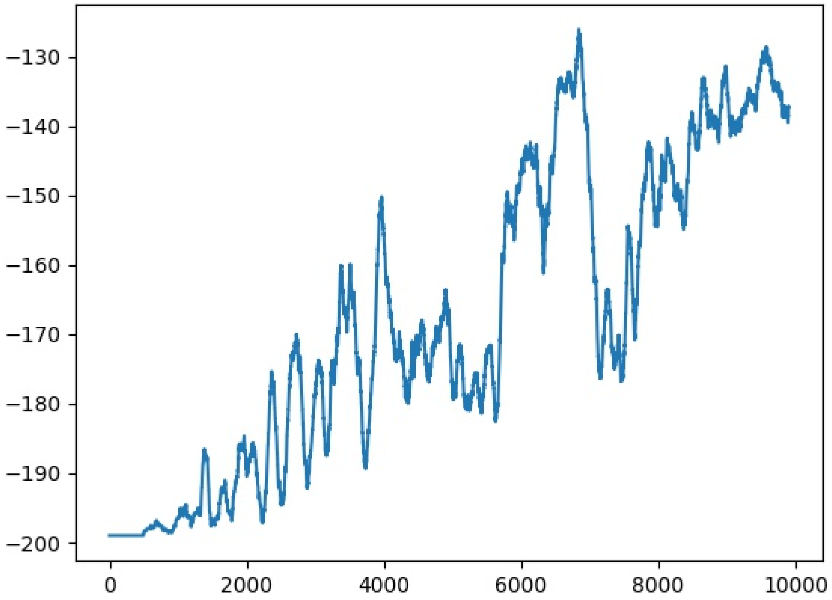

Figure 1: Task-1

Reward obtained (averaged over 100 episodes) = -132.77
Best reward obtained for a single episode = -112.0

In Task-1, the reward values are quite unstable and fluctuating a lot. Although as we proceed with the number of iterations, the overall reward value on the graph seems to increase in the long run. Despite the instability in the graph, the value of reward at the end of 10000 iterations seems to be well above -160. Tabular sarsa seems to be a fair enough approximation for the infinite state space of position and velocity. Since we see the reward to be increasing this ensures that the decrease in the number of timestamps taken by the car to escape the valley.

# Task 2

Epsilon = 0.01
Learning rate = 0.01
Number of Tiles = 10
Tile width for parameter x = 0.1
Tile width for parameter v = 0.01

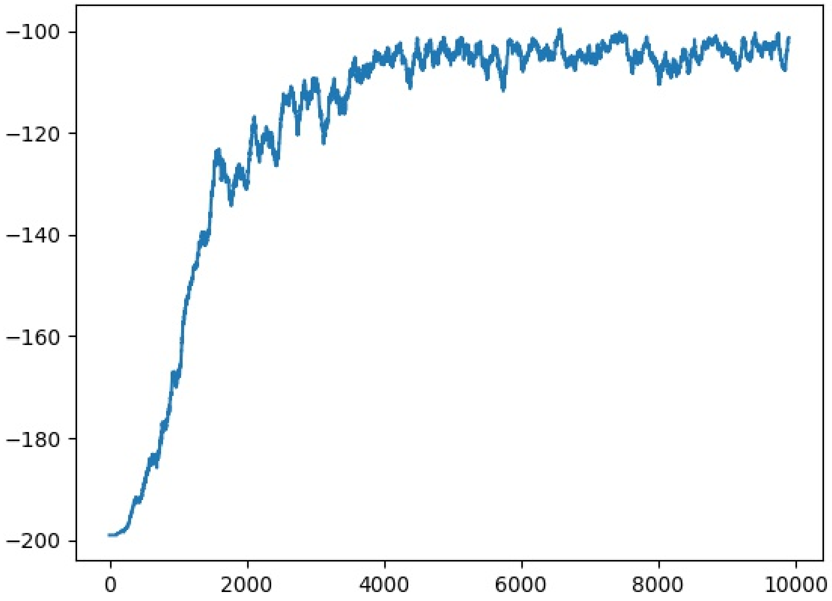

Figure 2: Task-2

Reward obtained (averaged over 100 episodes) = -98.61
Best reward obtained for a single episode = -83.0

In Task-2, the fluctuations observed are much smaller than that compared to Task-1. Also here, the graph is seen to climb very quickly in first 2000 iterations and after that, the reward is observed to stay above -120 and is also stable. Tile coding is observed to perform much better than tabular method because it stores weights corresponding to every tile and weights for all the tiles are updated for different episodes and weights for all the tiles are used for choosing the action. The value of reward at the end of 10000 iterations seems to be well above -130. Since we see the reward to be increasing this ensures that the decrease in the number of timestamps taken by the car to escape the valley.