Indian Institute of Technology Bombay
Autumn Semester 2021

Teacher: Prof. Shivaram Kalyanakrishnan

CS747: Programming Assignment 3

—

Raaghav Raaj, 180050082

Task 1

Tuning Parameters:

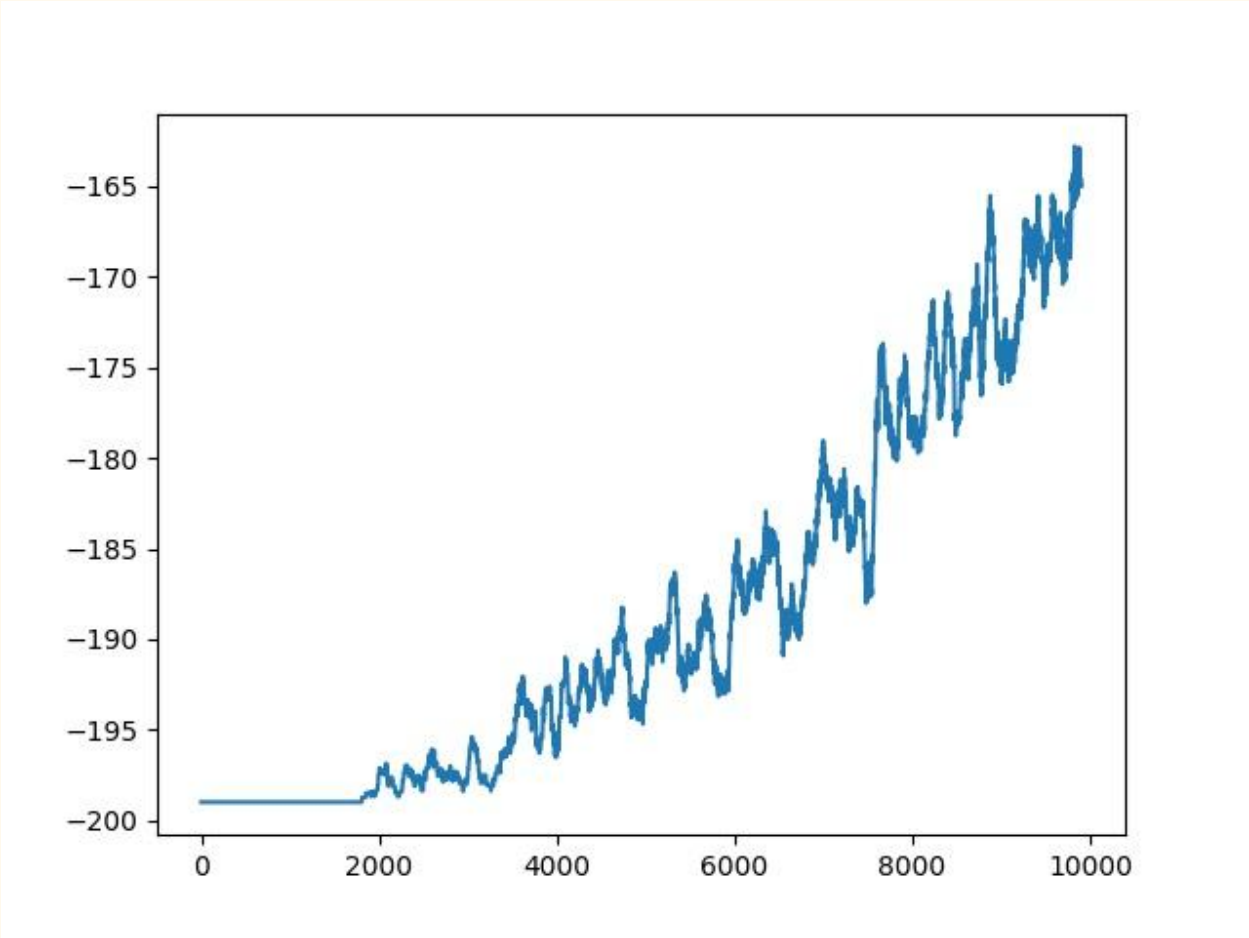
- epsilon = 0.1, learning rate = 0.3
- The range of position parameter x is divided into 51 intervals, whereas the range of velocity parameter v is divided into 101 intervals.

States and Weights initialisation:

- The weights are set as a 3-Dimensional matrix weight[i][j][a], where weight[:,:,a] denotes the weights for action a, over all the possible states.
- The state is visualised as a 2-Dimensional matrix where the cell in which the pair of position and velocity lies, is set to 1 and rest 0s.

Observations:

- The plot starts from the average reward value of -199 because the weights are initialised with all 0s that resulting in a reward of -1 for all the epochs except for one which is probably because it managed to learn a bit until the end of epochs.
- The score increases gradually and the learning rate was set by observing the plot using the range of [0.2, 0.5].
- The final reward obtained at this particular set of parameters and implementation was -153.65



Task 2 - Implemented using Tile Coding

Tuning Parameters:

- epsilon = 0.1, learning rate = 0.5
- The range of position parameter x is divided into 31 intervals, whereas the range of velocity parameter v is divided into 41 intervals.

States and Weights initialisation:

- The weights are set as a 4-Dimensional matrix weight[t][i][j][a], where weight[:,:,:,a] denotes the weights for action a, over all the possible states, over all the tiles. The first axis is the tile axis.
- The state is represented in a similar manner as the earlier one only this time using tile coding. Thus, the corresponding (i, j) cell is set 1 for all the tiles.

Observations:

- The plot starts from the average reward value of -199 but starts increasing much before that in the previous task.
- The score continues to increase on an average but also drops at some points. The learning rate was set by observing the plot using the range of [0.4, 0.6].
- The final reward obtained at this particular set of parameters and implementation was -119.22