CS747

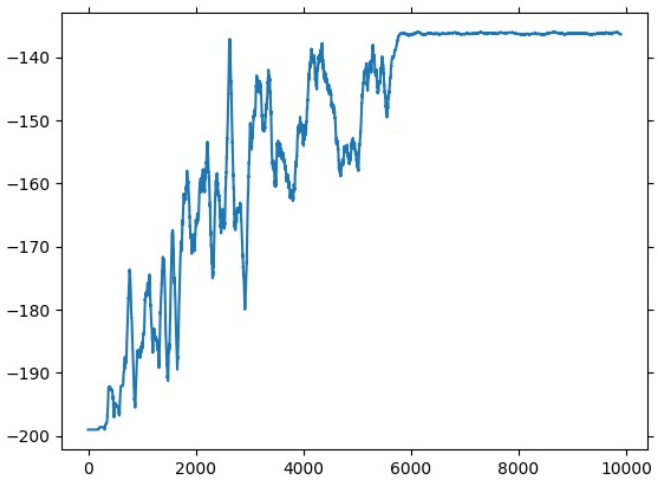
Assignment #3
Report

Sibasis Nayak (190050115)

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

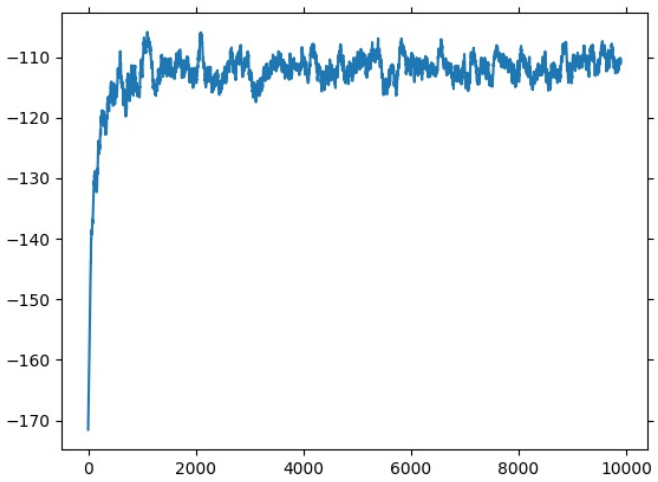
2021-2022

Task 1 Observations



- Implemented 19 way discretization for x, i.e bins of 0.1 size and 15 way discretization for v, i.e bins of 0.01 size. Therefore the size of my state vector is $15 \times 19 = 285$.
- ϵ has been set to e^{-5} and learning rate to 0.1. This achieves a reward of -136.24 on the default seed.
- We can observe from the graph that, the algorithm first tries to explore, hence there are a lot of ups and downs,and then it converges to the optimal reward value.
- I have kept ϵ to a number such that is sufficient for the algorithm to explore, but once it learns it should use the best action that is learnt from previous episodes.
- The learning rate is kept sufficiently large so that we move forward, and is not kept too large which might result in not converging to the optimum value.
- We can observe from the graph that the reward has converged after about 6000 episodes.

Task 2 Observations



- To obtain the feature vector I have used 2D tile encoding, the tiles were as per suggestions of course textbook, i.e 8 tiles of 9*9 size were used.
- The diagonal offset while moving from one tile to the next is set to be 1/8th of the size of the squares in each tile in both directions.
- ϵ has been set to e^{-5} and learning rate to 0.1. This achieves a reward of -98.79 on the default seed.
- The intuition behind choice of ϵ and learning rate is same as the first task.
- We can observe a faster training and convergence than the previous case which can be explained to the sophistication of the state vector with tile encoding which involves generalisation and parameter sharing.
- Convergence is observed even before 2000 episodes which is a lot of improvement from the previous task.