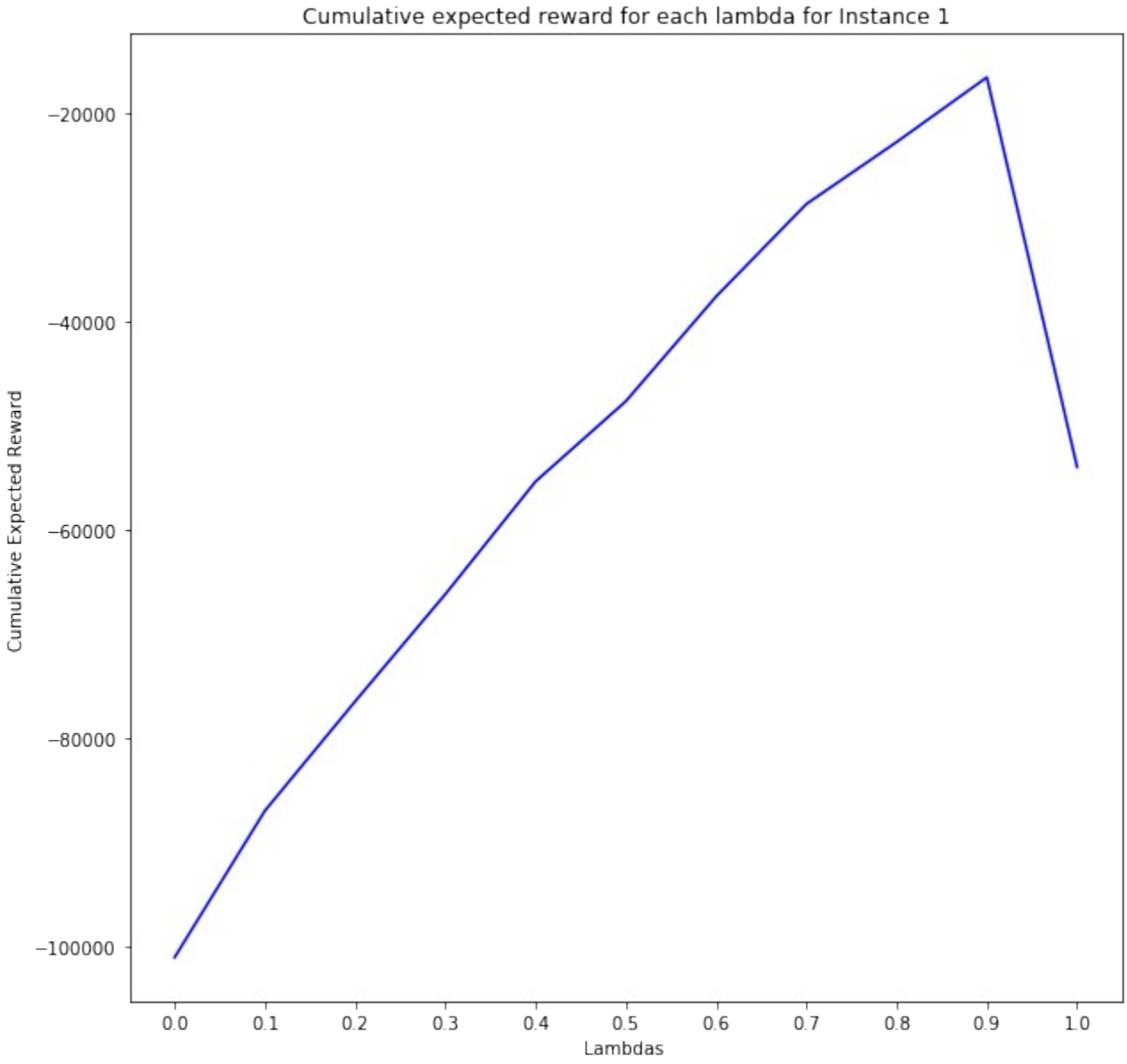
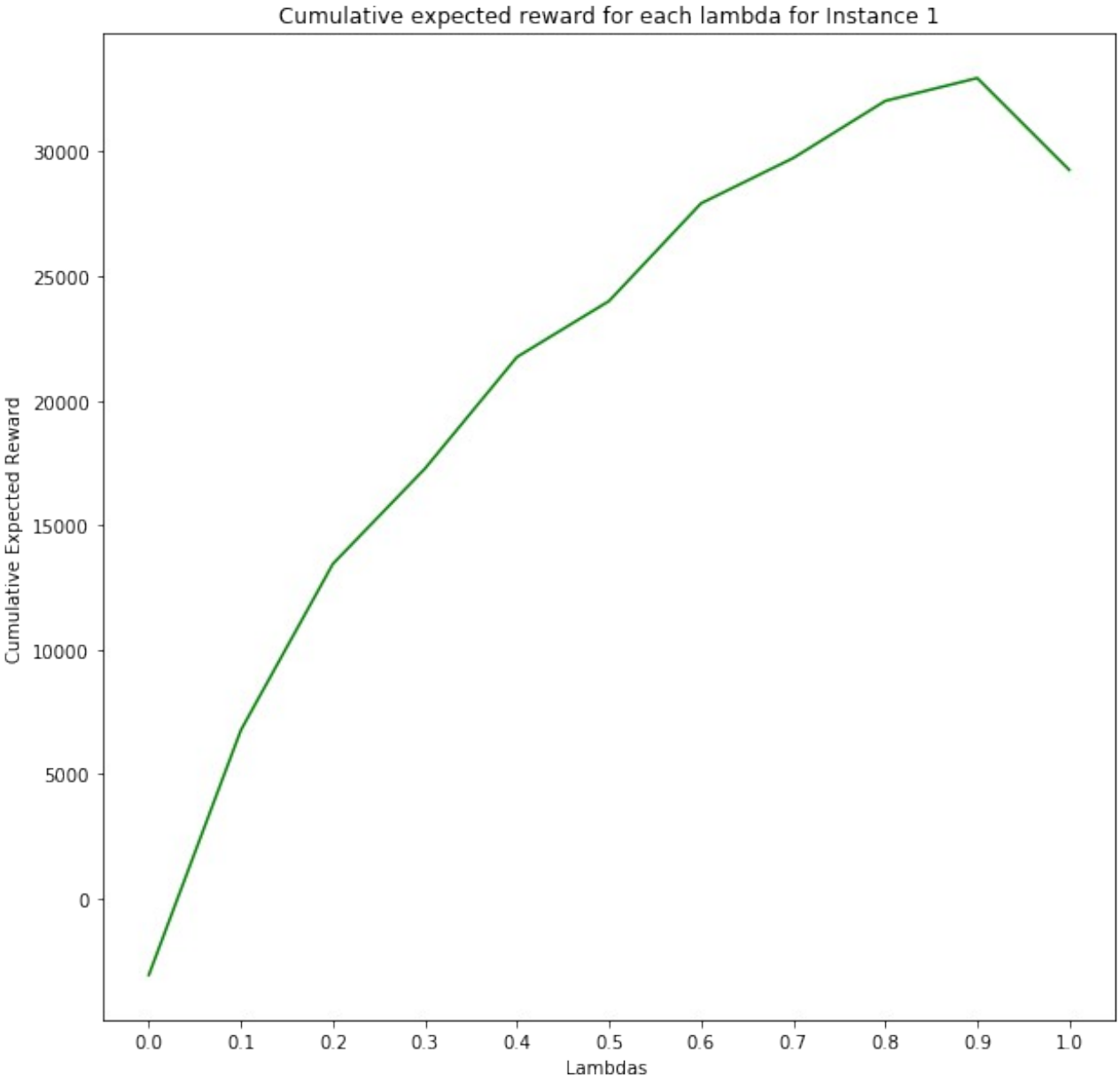
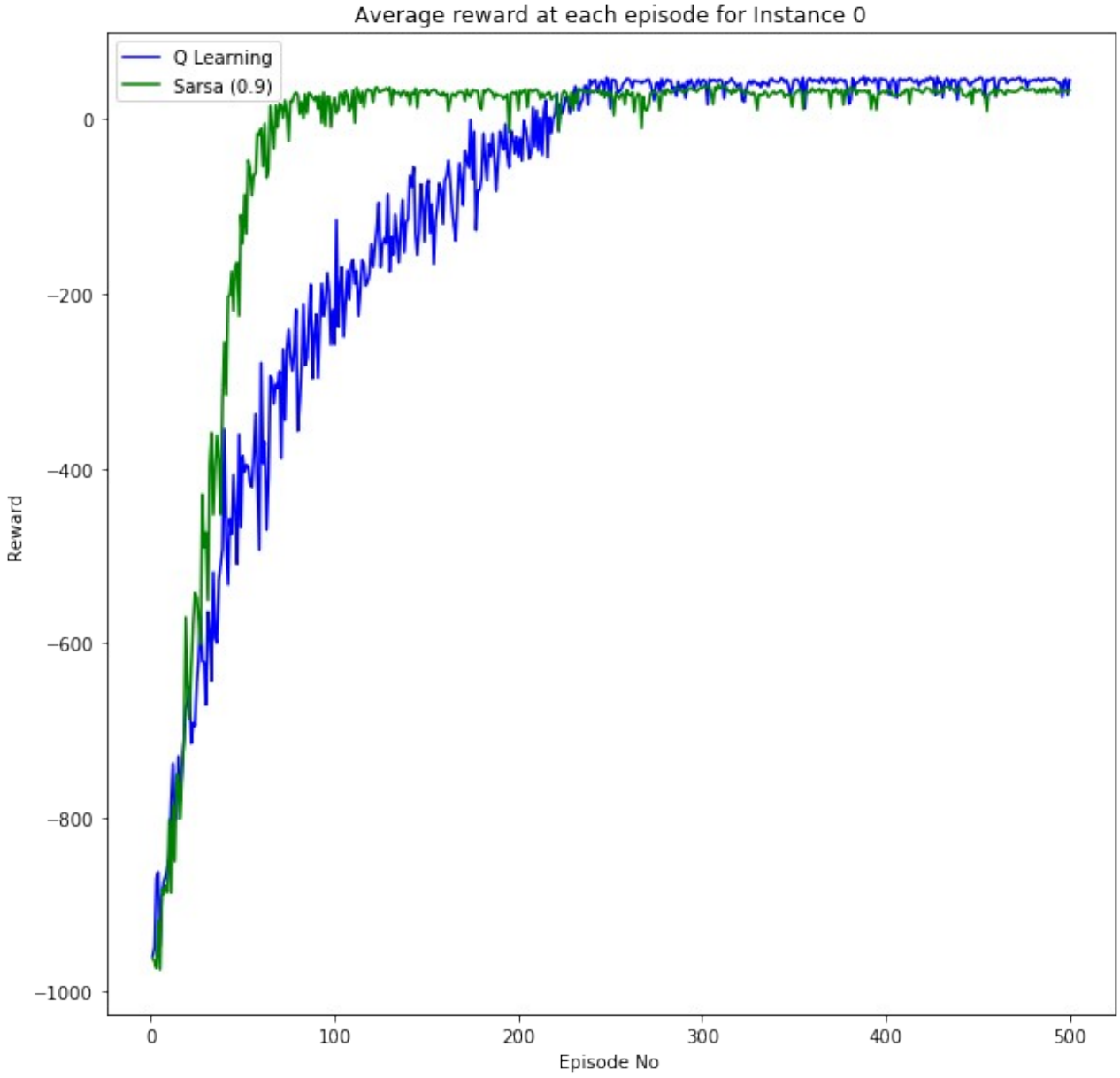
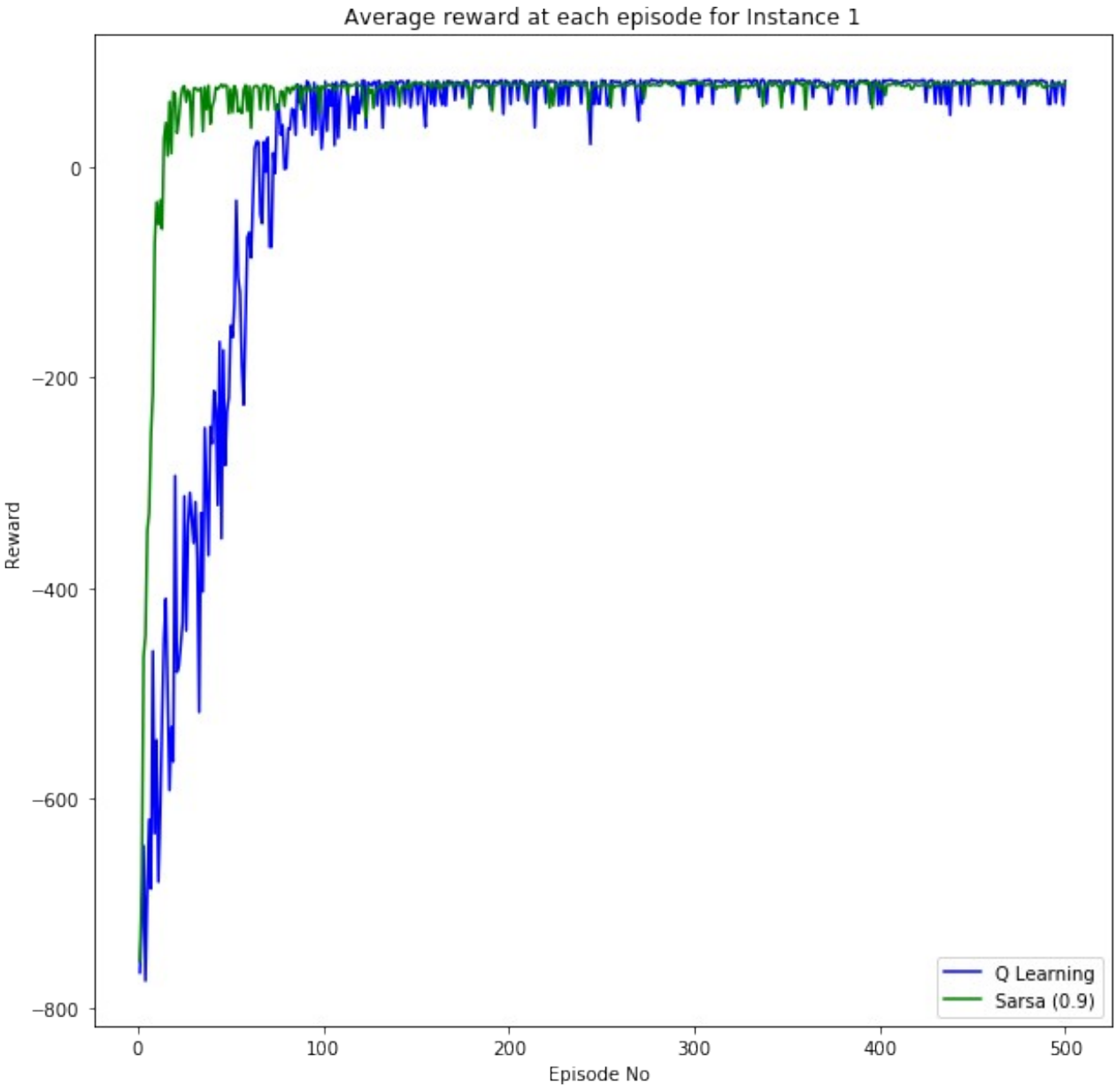
FILA Assignment 3 Report

-Utkarsh Gautam 140050009




Cumulative expected reward for each lambda for Instance 1
Cumulative Expected Reward
30000
25000
20000
15000
10000
5000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Lambdas


Average reward at each episode for Instance 0
Q Learning
Sarsa (0.9)
0
−200
−400
−600
−800
−1000
Reward
0
100
200
300
400
500
Episode No


Average reward at each episode for Instance 1
Reward
0
−200
−400
−600
−800
0
100
200
300
400
500
Episode No
Q Learning
Sarsa (0.9)

Tuned Hyperparameters for best performance*

- Sarsa
  - $\alpha = 0.5$ (No annealing.)
  - $\gamma = 0.9$
  - $\lambda = 0.9$
- Q Learning
  - $\alpha = 0.9$ (No annealing.)
  - $\gamma = 0.95$

Observations

- Increasing lambda increases the total reward until lambda becomes one when the total rewards decreases for both instances
- Sarsa(λ) learns quicker than Q Learning as evident from the average reward at ith episode but eventually learning saturates after certain number of episodes and both give similar rewards for further episodes
- Looking at total rewards for both instances we can conclude that instance zero is tougher to learn than instance one.
- The optimal value of lambda will lie between 0 and 1