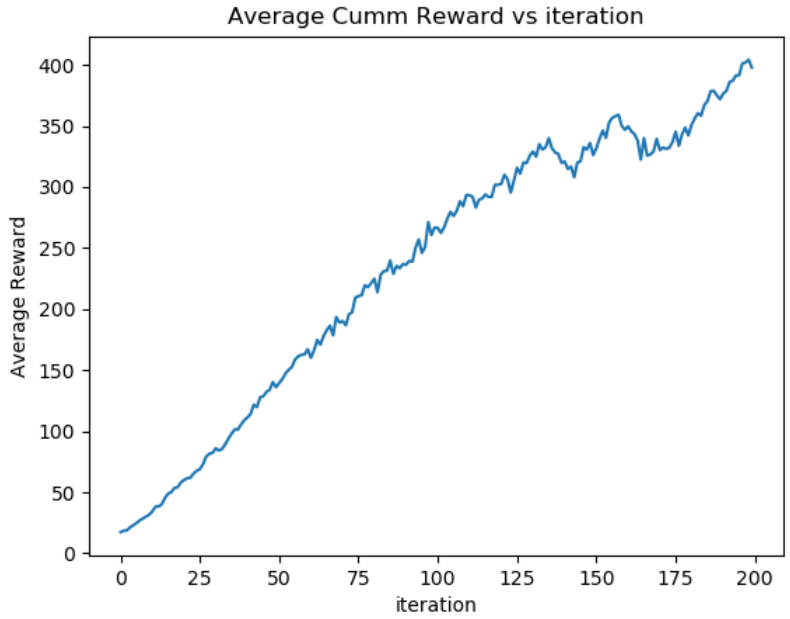
CS 593 : Robotics | Assignment 4
Sashank Modali | nmodali@purdue.edu | 0032792055

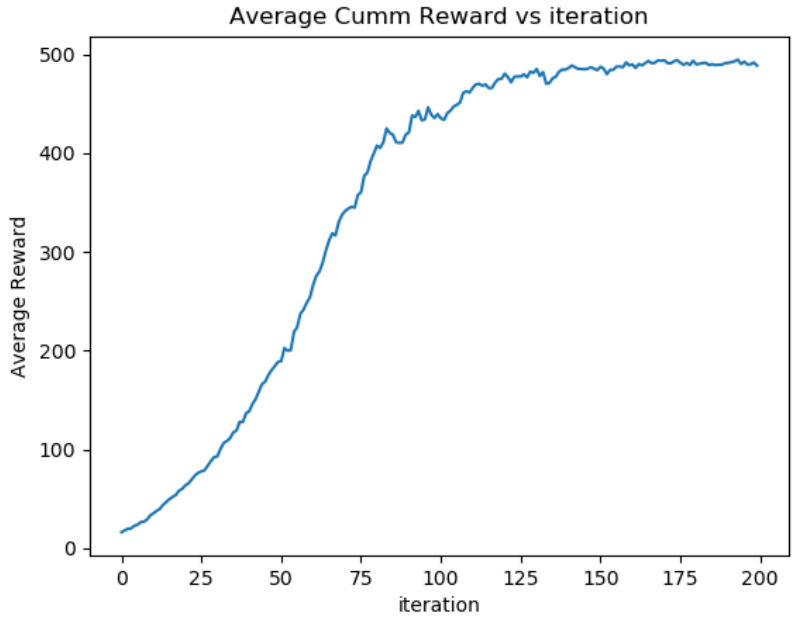
---

1. The vanilla reinforce algorithm was implemented and the results are presented below :
    a. Vanilla reinforce with full return considered in each step:

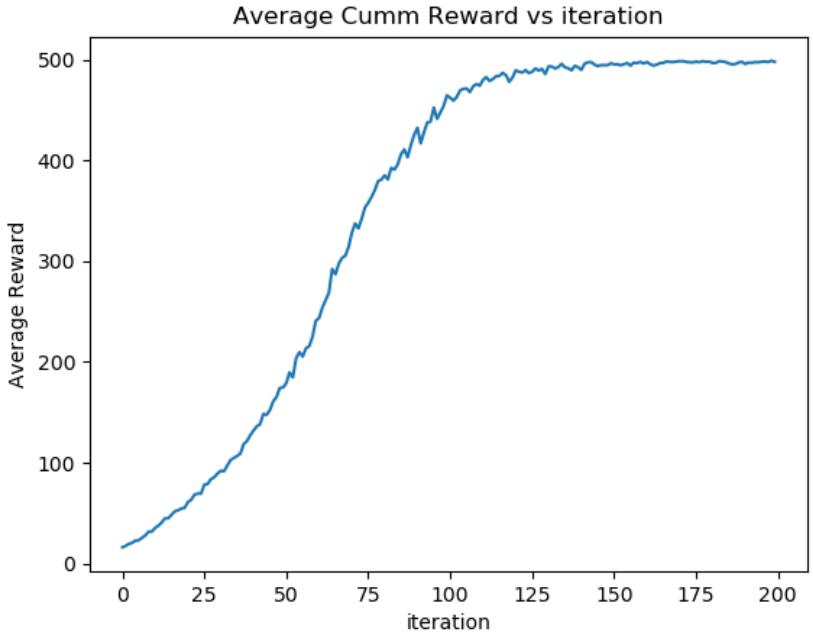
    Lr =0.001



    b. Vanilla reinforce with current return considered :

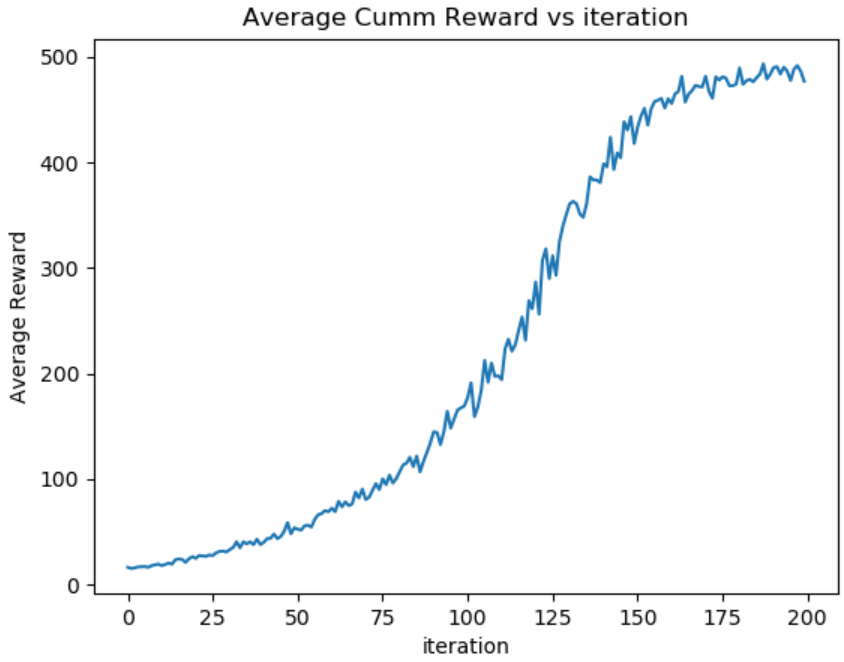


c. Vanilla reinforce with baseline considered :

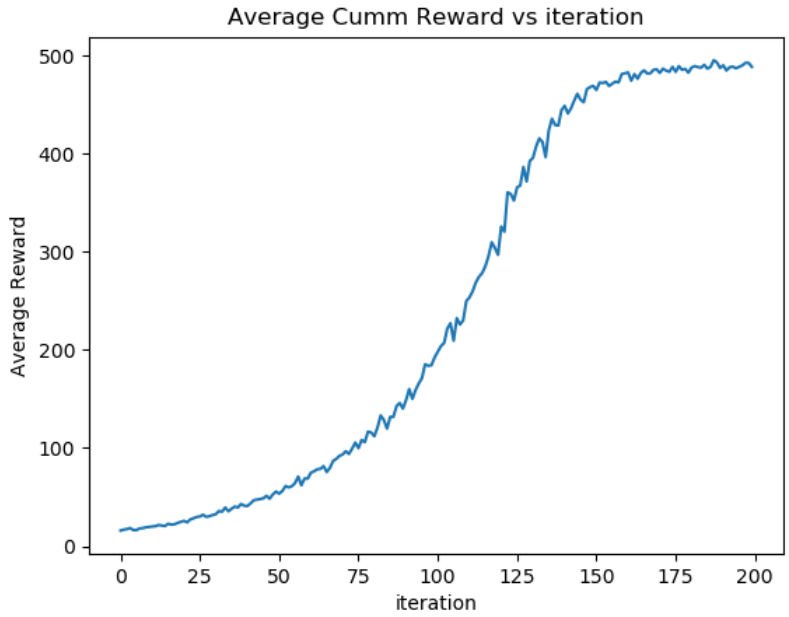


d. For a different number of episodes in each iteration, the plots are as follows:

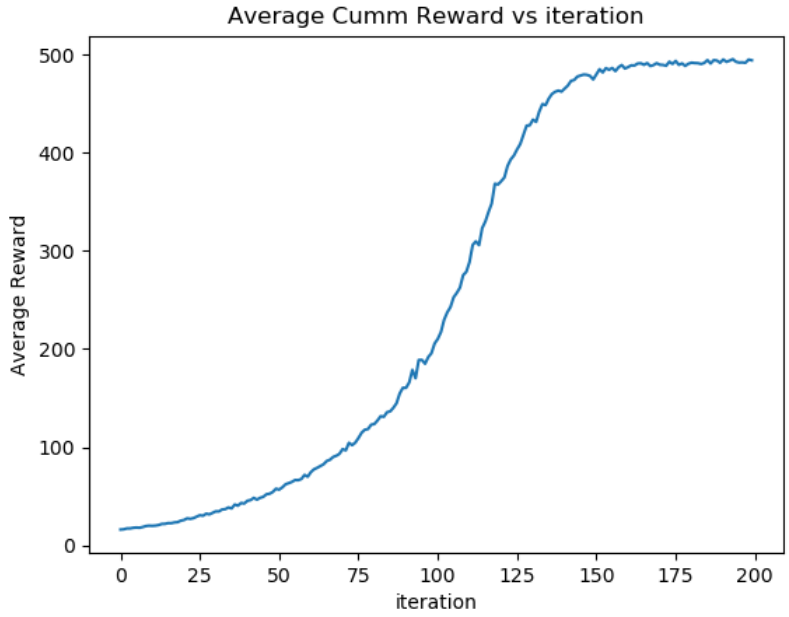
100 episodes:



300 episodes:



1000 episodes:

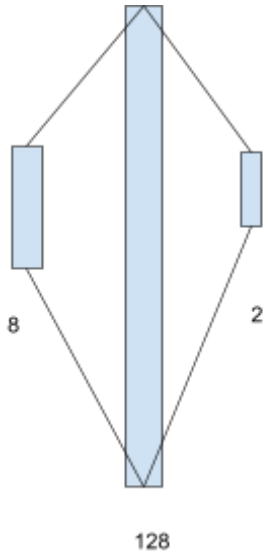


It has been observed that increasing episode count decreases the variance of the update. This leads to better training. We see that this helps in a slightly faster convergence of the algorithm.

2. For the continuous problem, the results are as follows:

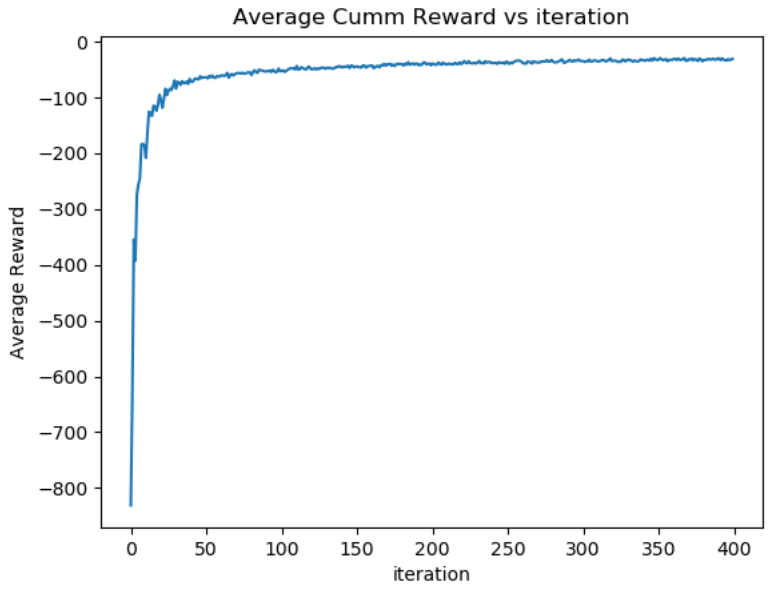
   The average cumulative reward didn't go above -22. It hovered around -22, for a large number of iterations and I was unable to get it to go above that. I tried changing the neural network, from 2 layers of 128 hidden layer size to 3 layers of 32 length hidden layer features. Both didn't allow the neural network to go above this cumulative reward. Therefore, I have attached the results of my plots for both networks below:

   a. For the same neural network structure fro both mean and sigma as follows:

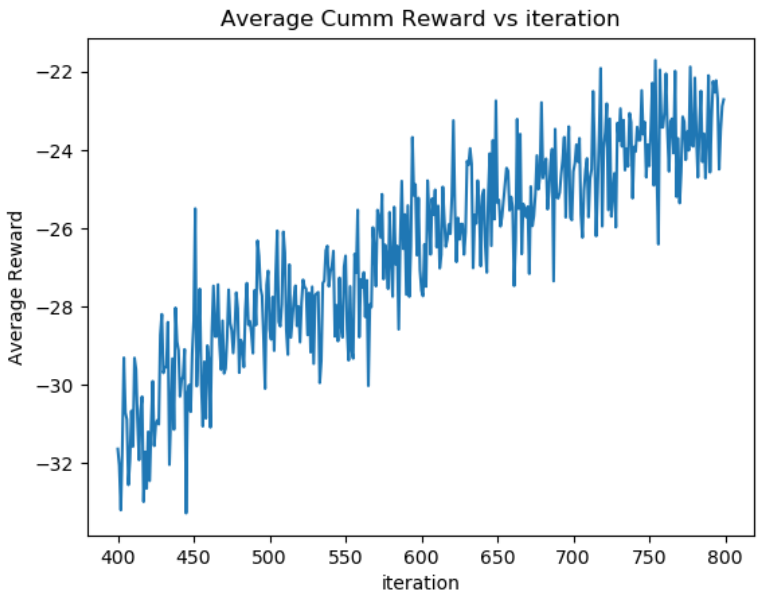


   Episodes: 100 iterations : 1600, type : current_reward formulation

   From 0 to 400 iterations :



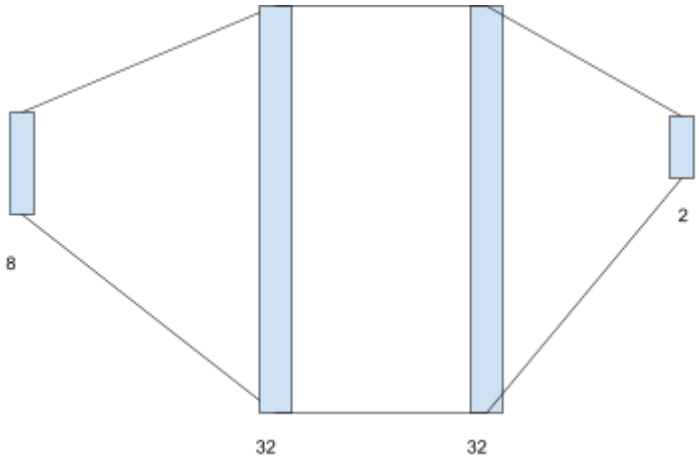
From 400 to 800 iterations:



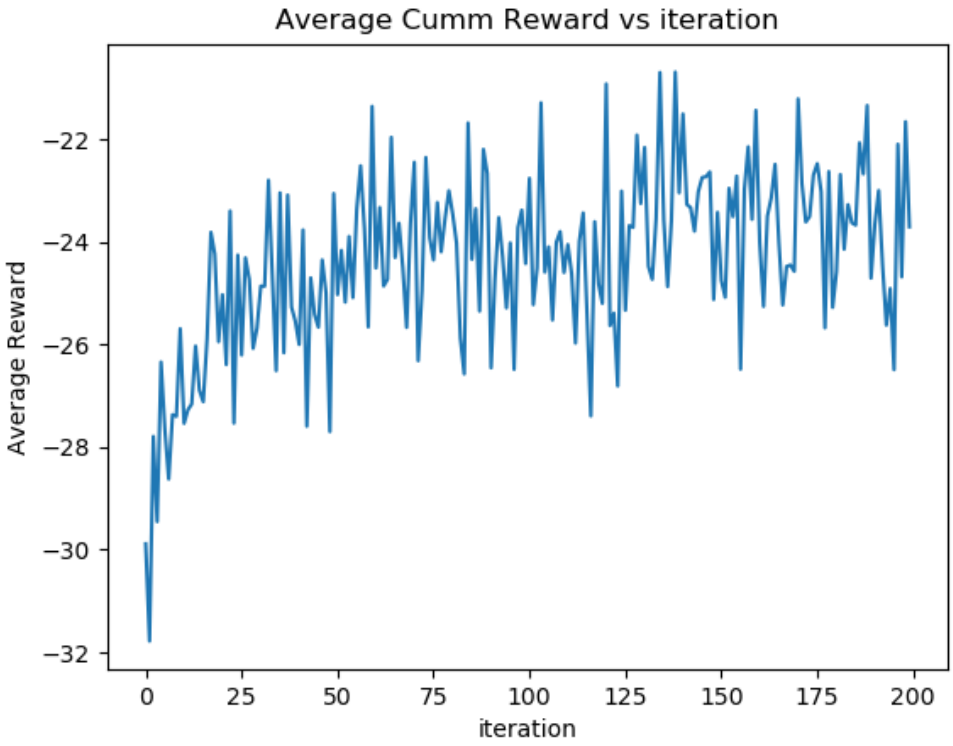
This was continued to 1600 iterations but the average cumulative reward stayed near -22.

Thus it wasn't possible to improve after this, from my code.

b. For the following neural network structure of the mean , and constant variance =0.1 for both actions:



Episodes : 50, iterations : 200, advantage formulation, lr =0.001, the rewards obtained were as follows:



Thus. I was unable to make the neural net converge. The resultant gif has been uploaded along with the code.