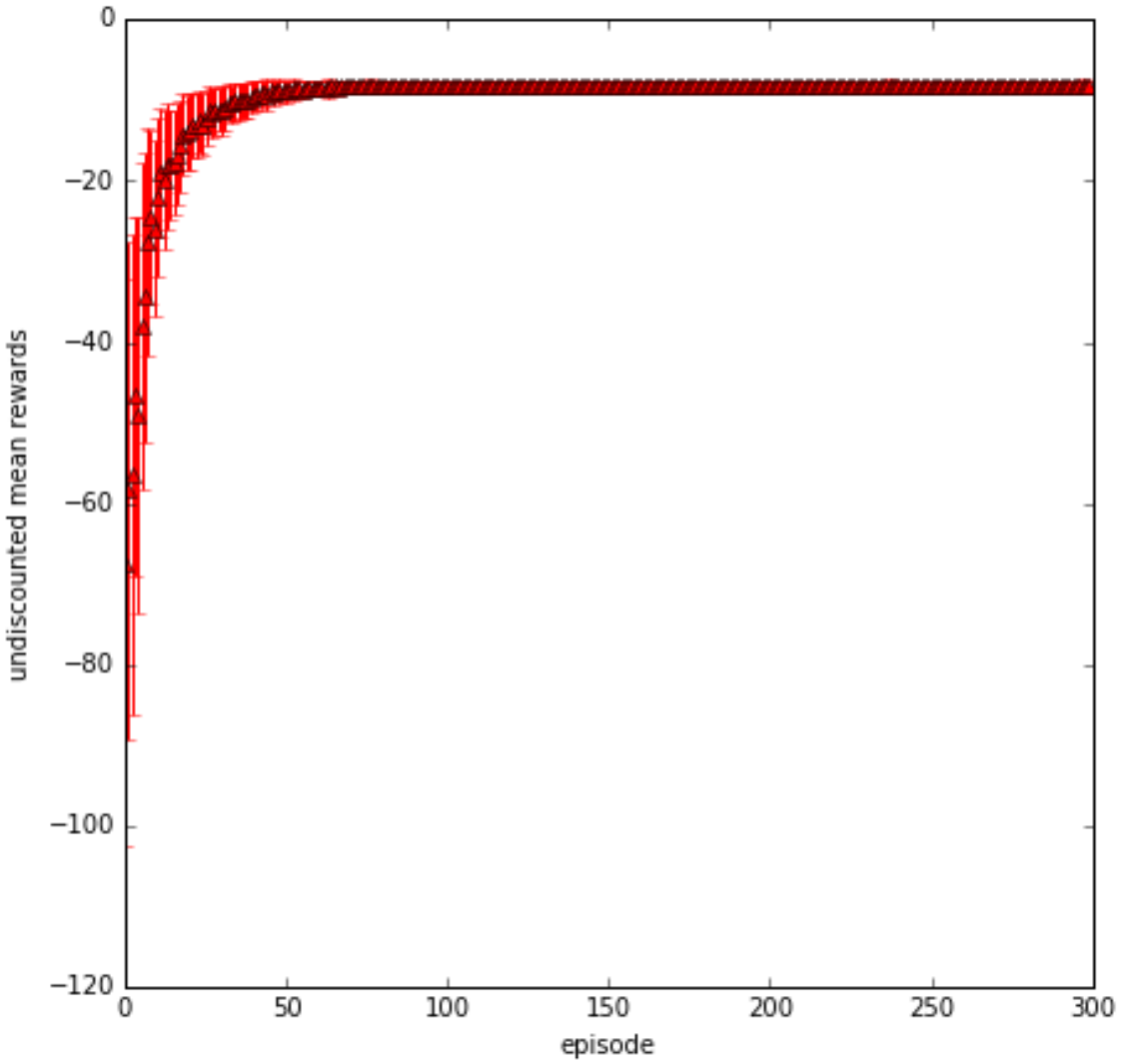
HOME WORK 4 - CS 687

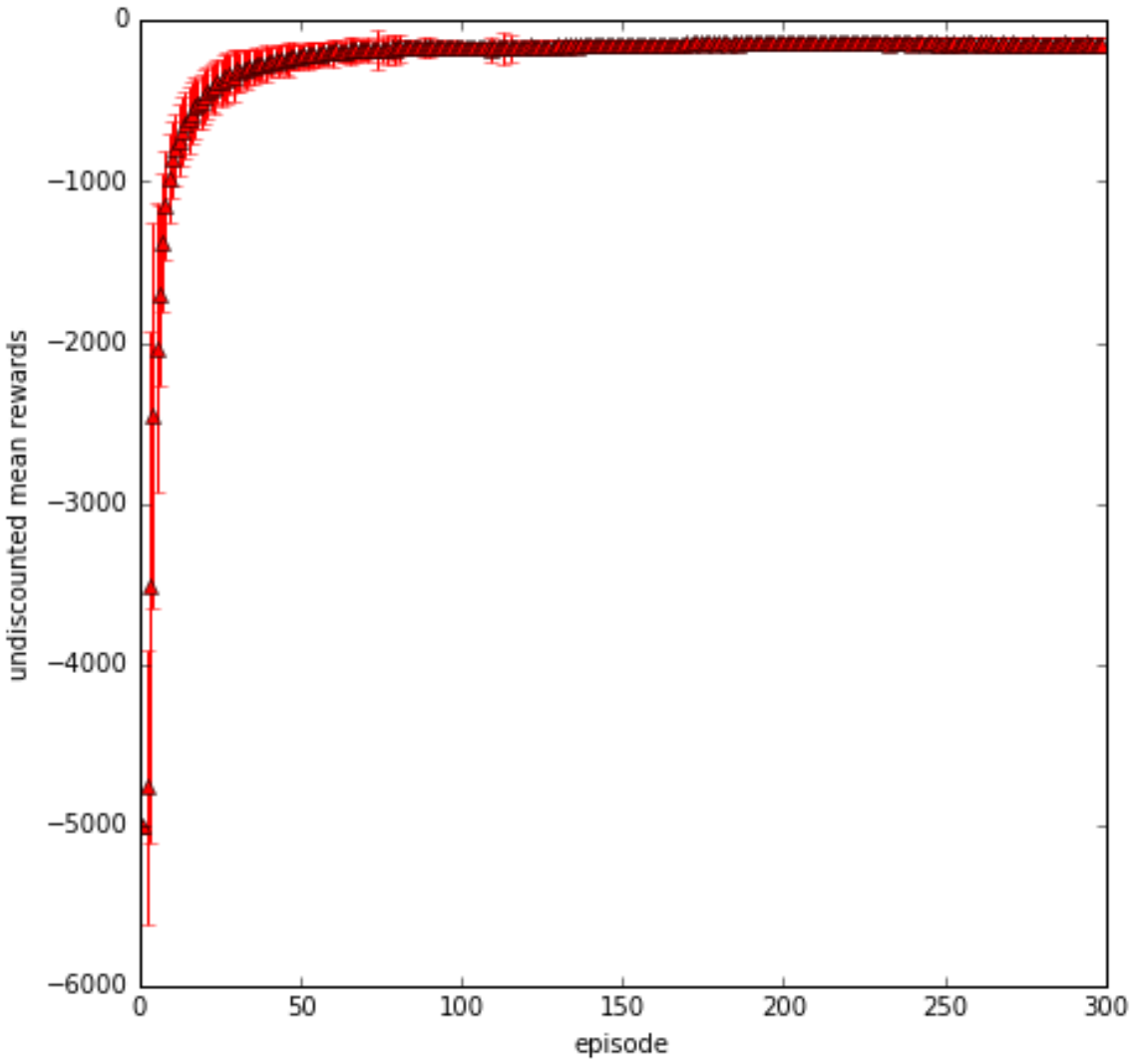
DANIEL SAM PETE THIYAGU

Qlambda Graphs. Grid World
alpha: 0.5, lambda: 0.3, epsilon: 0.01, gamma: 1
Enter number of trials to run: 200
Enter the number of episodes per trial: 300



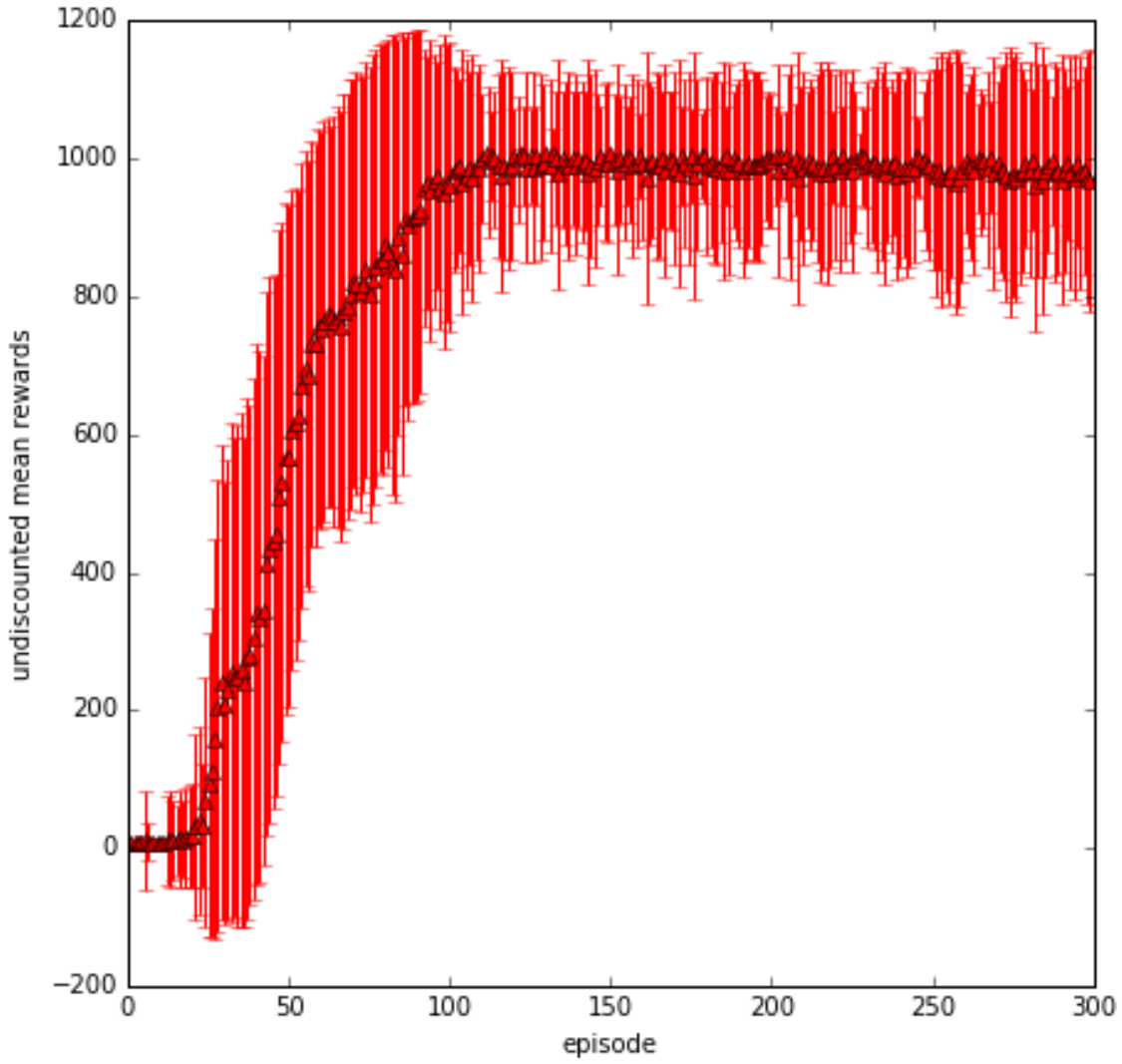
Mountain Car

Enter Fourier basis order: 2
Enter alpha: 0.001
Enter lambda: 0.1
Enter epsilon: 0.01
Enter gamma: 1
Enter number of trials to run: 200
Enter the number of episodes per trial: 300



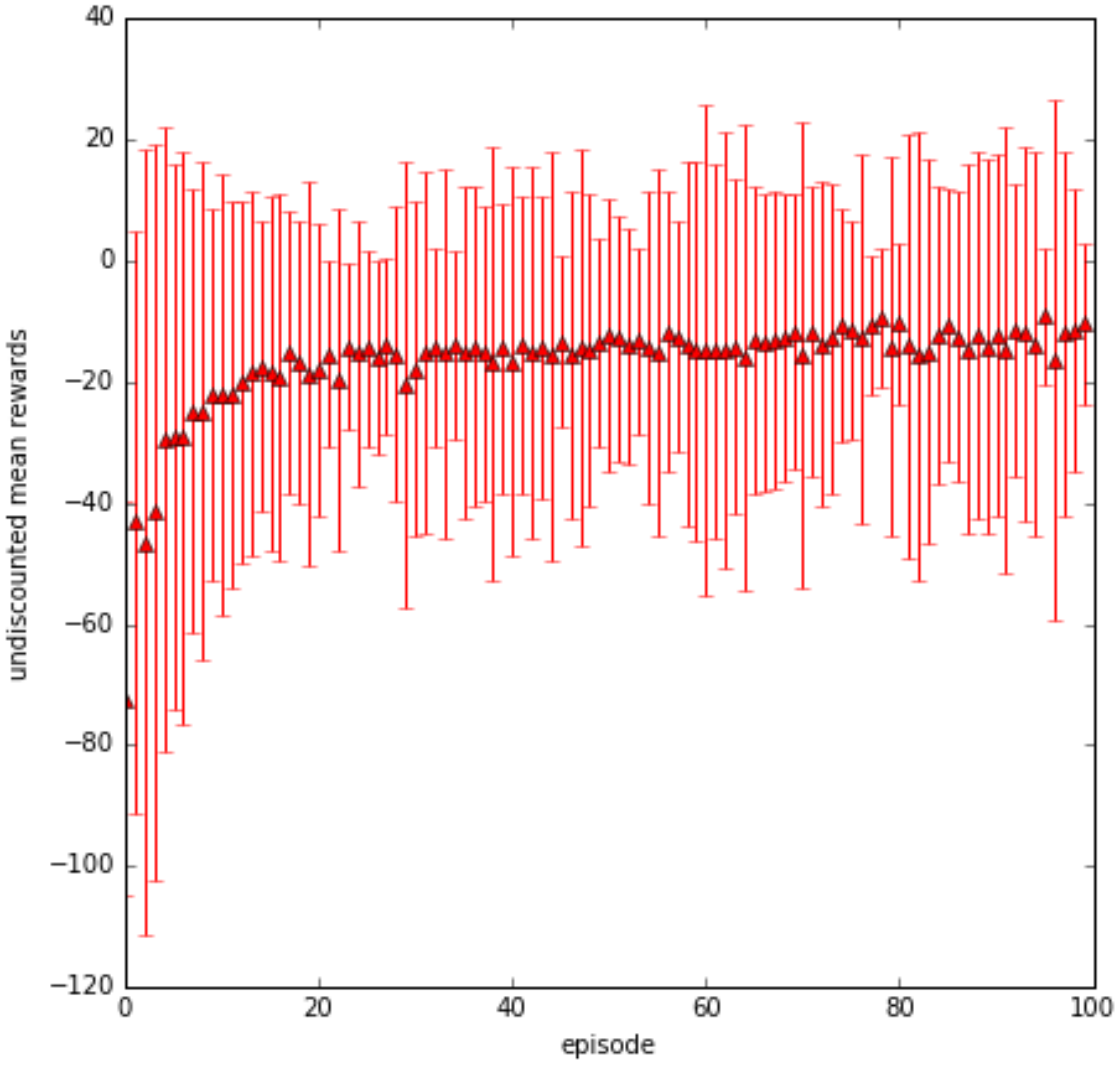
Cart Pole

Enter Fourier basis order: 3
Enter alpha: 0.001
Enter lambda: 0.07
Enter epsilon: 0.04
Enter gamma: 1
Enter number of trials to run: 200
Enter the number of episodes per trial: 300



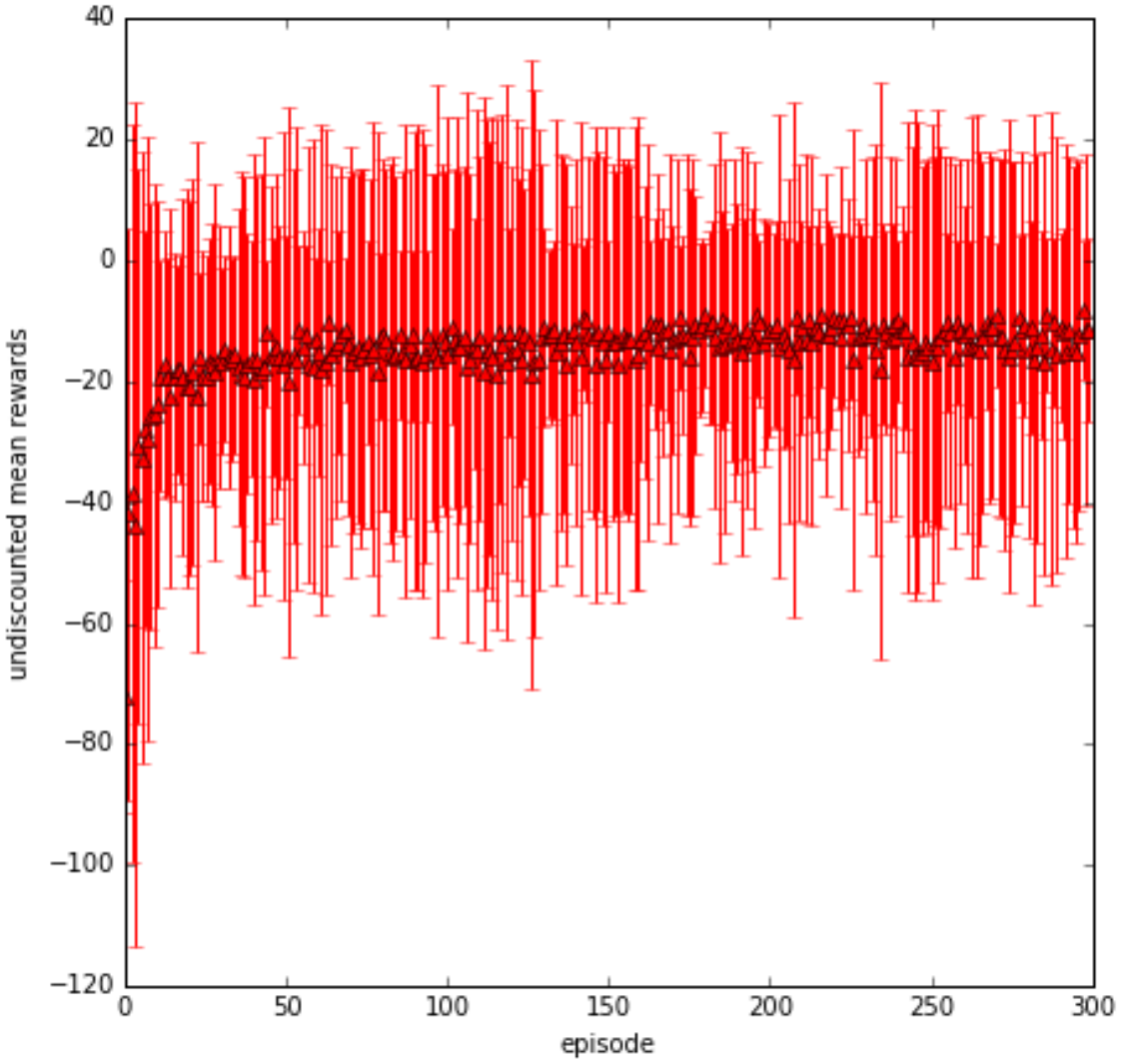
Acrobat - Graph 1

Enter Fourier basis order: 2
Enter alpha: 0.001
Enter lambda: 0.1
Enter epsilon: 0.01
Enter gamma: 1
Enter number of trials to run: 200
Enter the number of episodes per trial: 100



Acrobat - Graph 2

Enter Fourier basis order: 2
Enter alpha: 0.001
Enter lambda: 0.05
Enter epsilon: 0.01
Enter gamma: 1
Enter number of trials to run: 120
Enter the number of episodes per trial: 300



Difficulty of QLambda. Qlambda was easier to find hyper parameters. Cart Pole and Acrobat was a bit difficult to find good parameters, Decreasing lambda, decreases variance in most cases. It does get a bit easier since i have some experience. Cart Pole and Acrobat are the ones which seems to be most sensitive to. Grid World and Mountain car were easier and less sensitive. i wasn't surprised by any of the hyper parameters.

Actor Critic DLNPI Derivation.

$$lnP = \log \frac{e^{\theta_j \phi(s)}}{\sum_i e^{\theta_i \phi(s)}}$$

$$lnP = \log e^{\theta_j \phi(s)} - \log \sum_i e^{\theta_i \phi(s)}$$

$$lnP = \theta_j \phi(s) - \log \sum_i e^{\theta_i \phi(s)}$$

$$\frac{\partial lnP}{\partial \theta_j} = \phi(s) - \frac{1}{\sum_i e^{\theta_i \phi(s)}} * \phi(s)$$

If k is not equal to j ,

$$\frac{\partial lnP}{\partial \theta_k} = 0 - \frac{1}{\sum_i e^{\theta_i \phi(s)}} * \phi(s)$$

$$\frac{\partial lnP}{\partial \theta_k} = -\frac{1}{\sum_i e^{\theta_i \phi(s)}} * \phi(s)$$

$$\frac{\partial lnP}{\partial \theta} = [\frac{\partial lnP}{\partial \theta_1}, \frac{\partial lnP}{\partial \theta_2}, .., \frac{\partial lnP}{\partial \theta_{|A|}}]$$

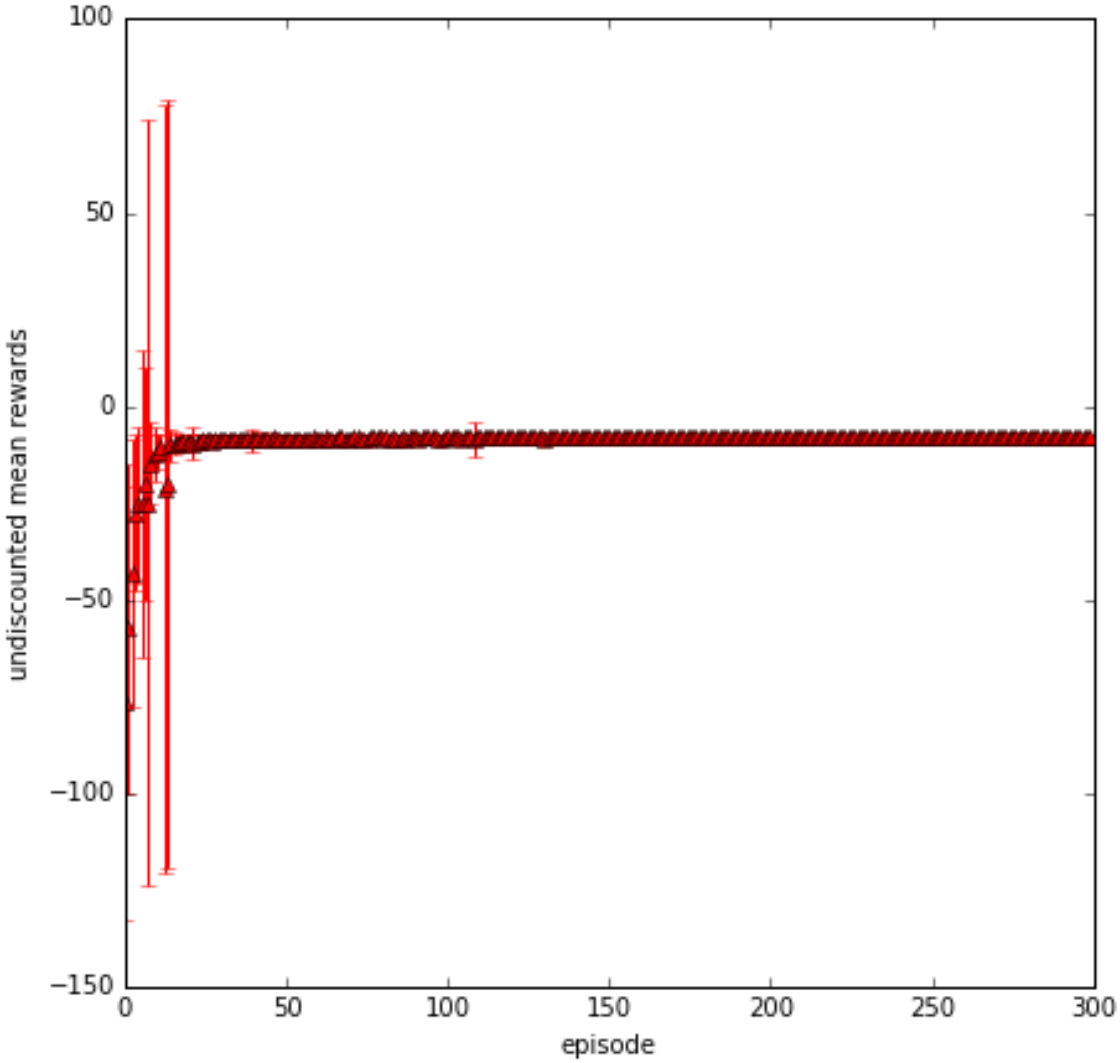
is going to be a vector for each action in the set of actions.

Difficulty of Actor Critic. Qlambda was easier to find hyper parameters than Actor Critic Algorithm, but this seems to work really well. Cart Pole and Acrobat was a bit difficult to find good parameters. Cart Pole and Acrobat are the ones which seems to be most sensitive to different parameter values. Grid World and Mountain car were easier and less sensitive.

Difficulty of REINFORCE. It was so much more difficult to find good hyperparameters. It is very sensitive to the parameters and hence grid search and other techniques are a bit difficult. They dont seem to do extremely better than the first two algorithms

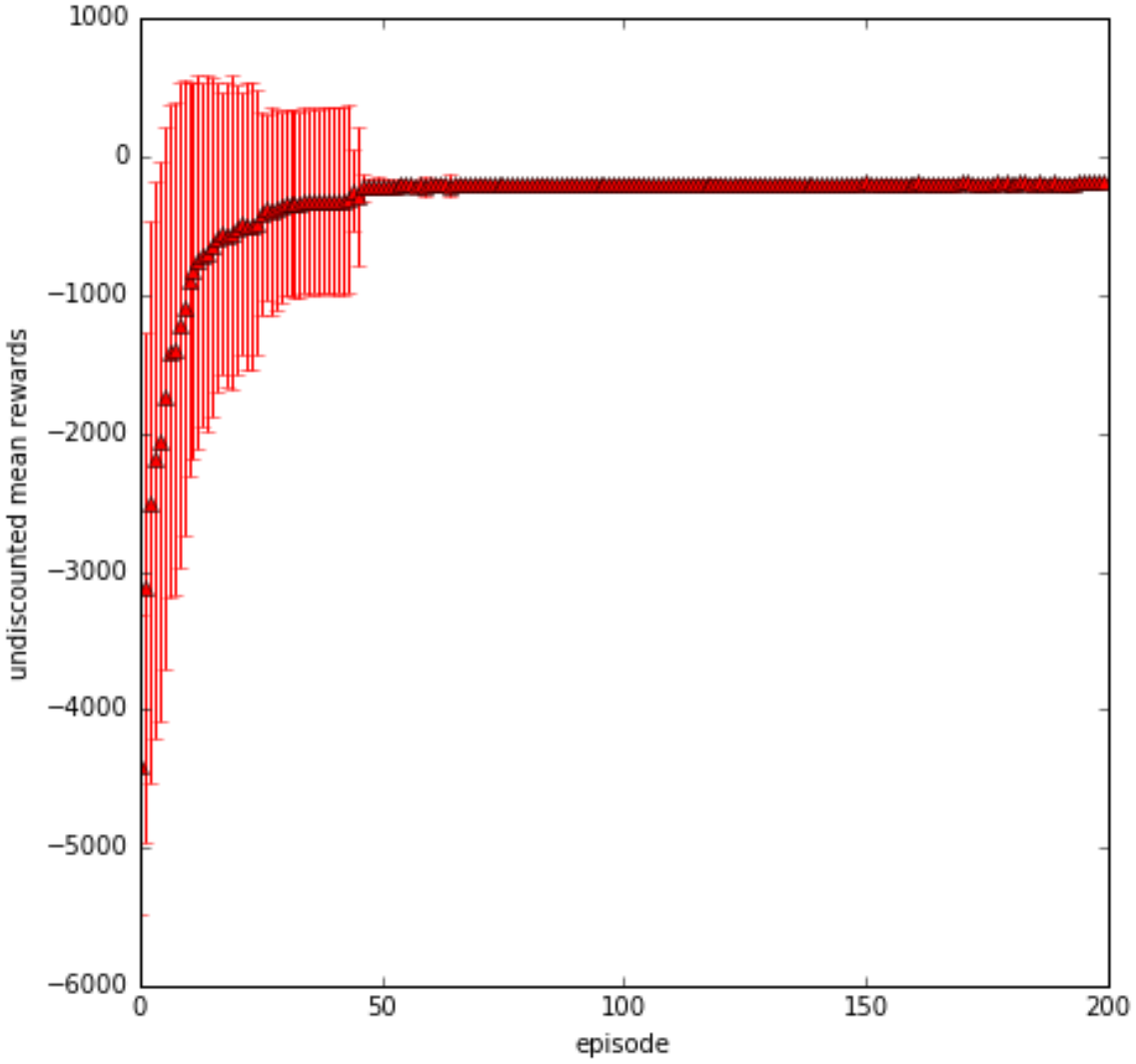
Actor Critic Graphs. Grid World

Enter alpha (actor): 0.4
Enter alpha (critic): 0.9
Enter lambda: .3
Enter gamma: 1
Enter number of trials to run: 100
Enter the number of episodes per trial: 300



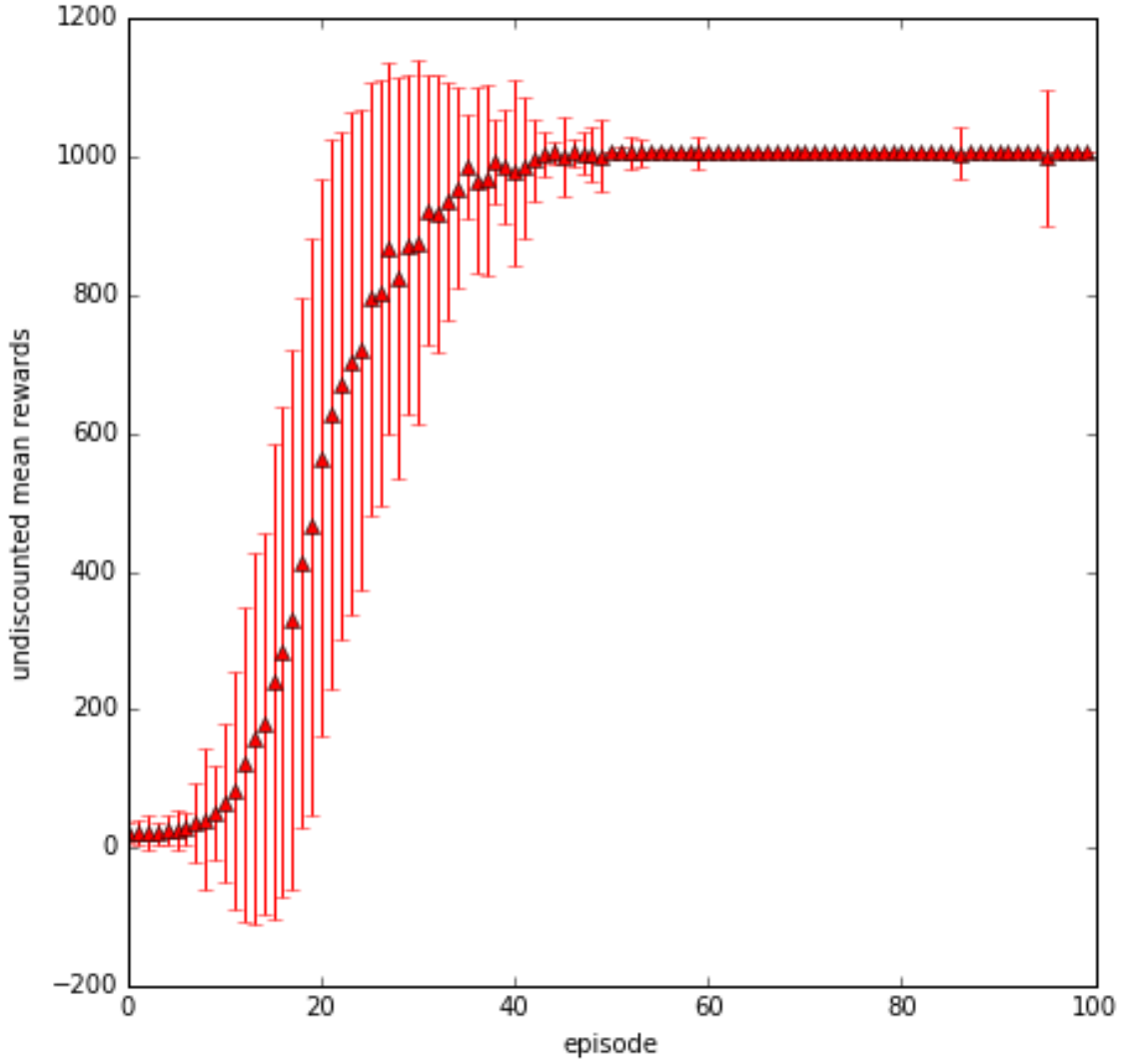
Mountain car

Enter Fourier basis order: 1
Enter alpha (actor): 0.1
Enter alpha (critic): 0.8
Enter lambda: 0.01
Enter gamma: 0.8
Enter number of trials to run: 100
Enter the number of episodes per trial: 200



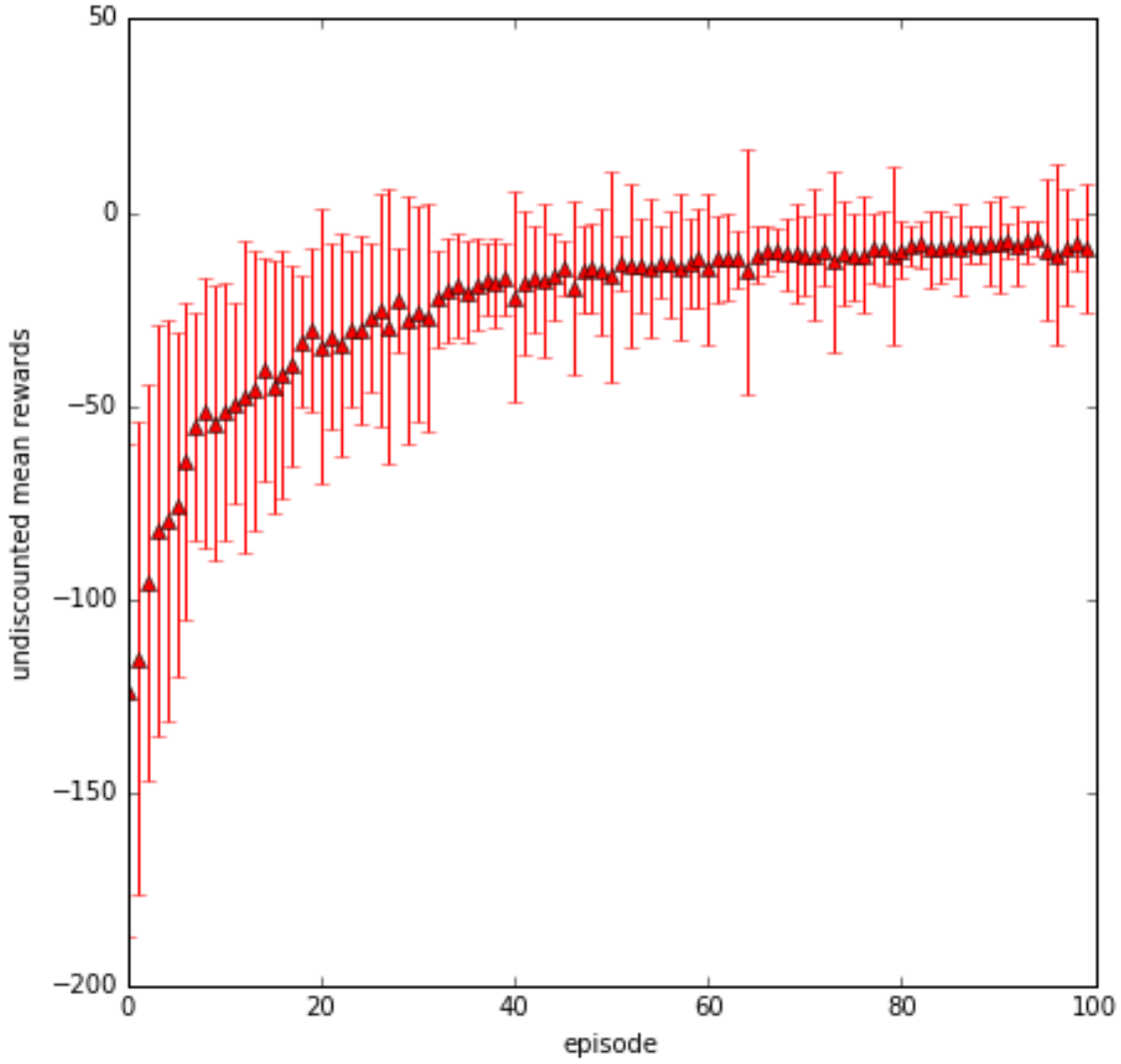
Cart Pole

Enter Fourier basis order: 3
Enter alpha (actor): 0.001
Enter alpha (critic): 0.001
Enter lambda: 0.1
Enter gamma: 0.9
Enter number of trials to run: 100
Enter the number of episodes per trial: 100



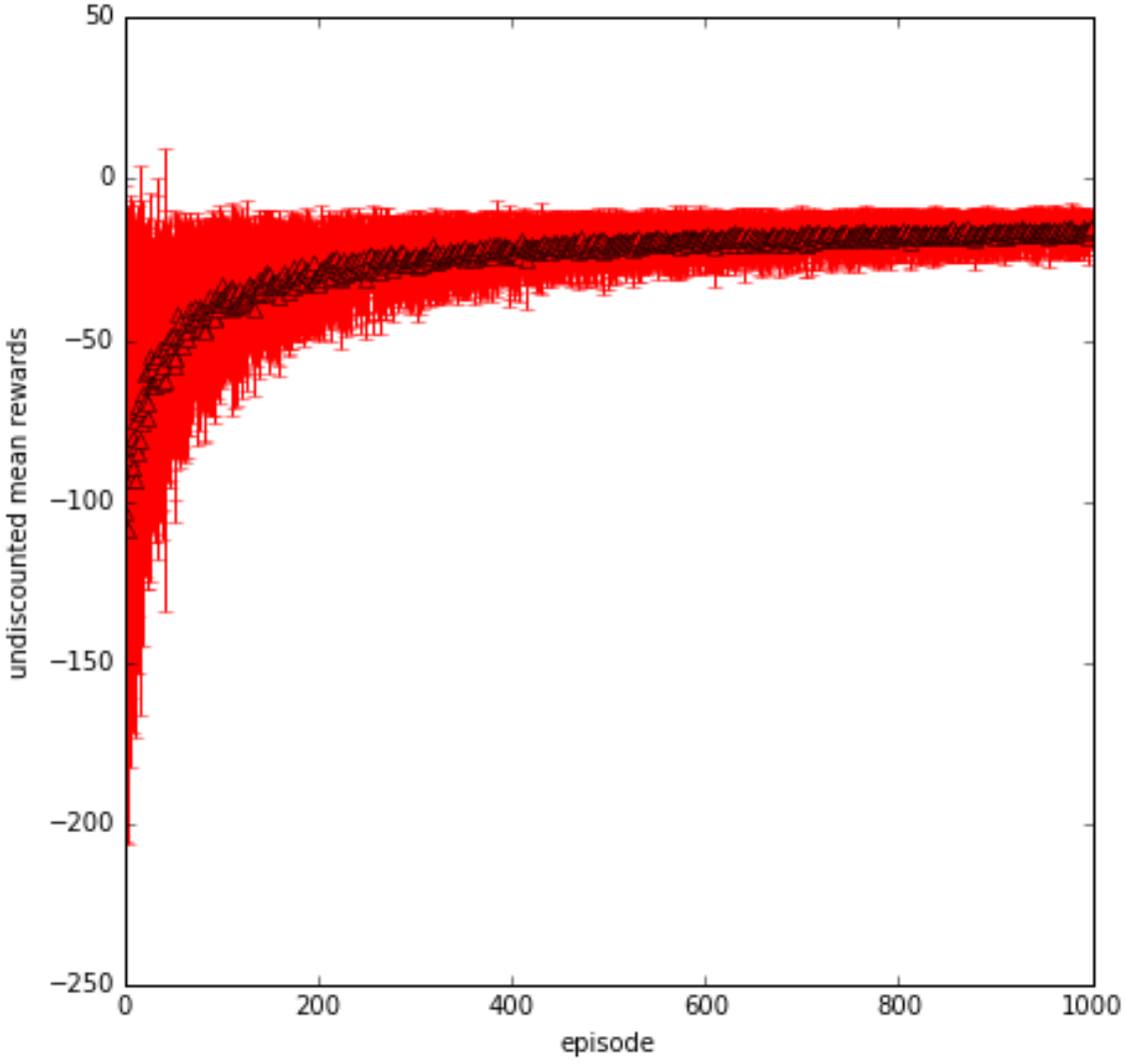
Acrobat

Enter Fourier basis order: 3
Enter alpha (actor): 0.0001
Enter alpha (critic): 0.01
Enter lambda: 0.3
Enter gamma: 1
Enter number of trials to run: 100
Enter the number of episodes per trial: 100



REINFORCE Graphs. Grid World

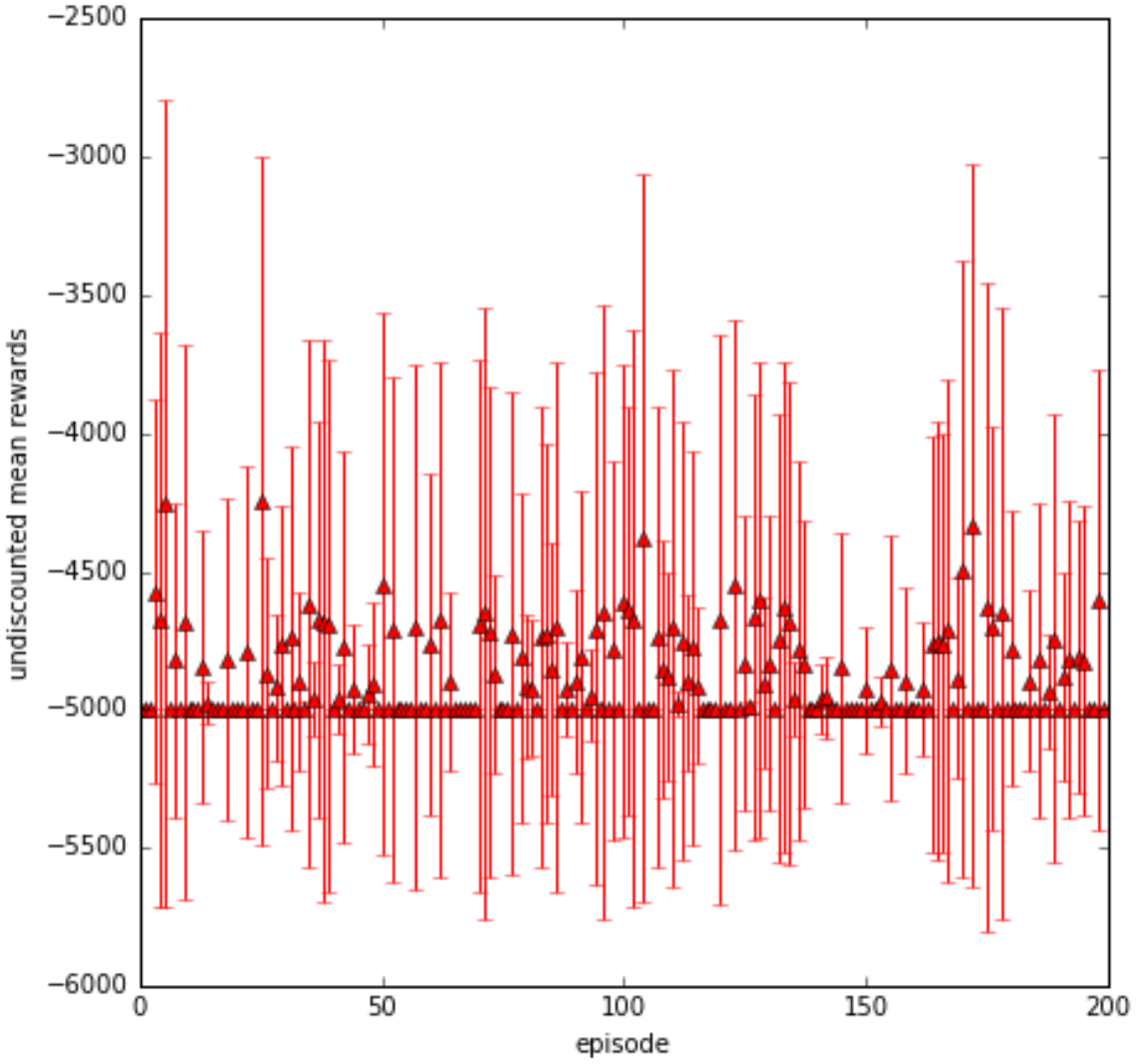
Enter alpha (actor): 0.001
Enter alpha (baseline): 0.01
Enter gamma: 1
Enter number of trials to run: 100
Enter the number of episodes per trial: 1000



Mountain car

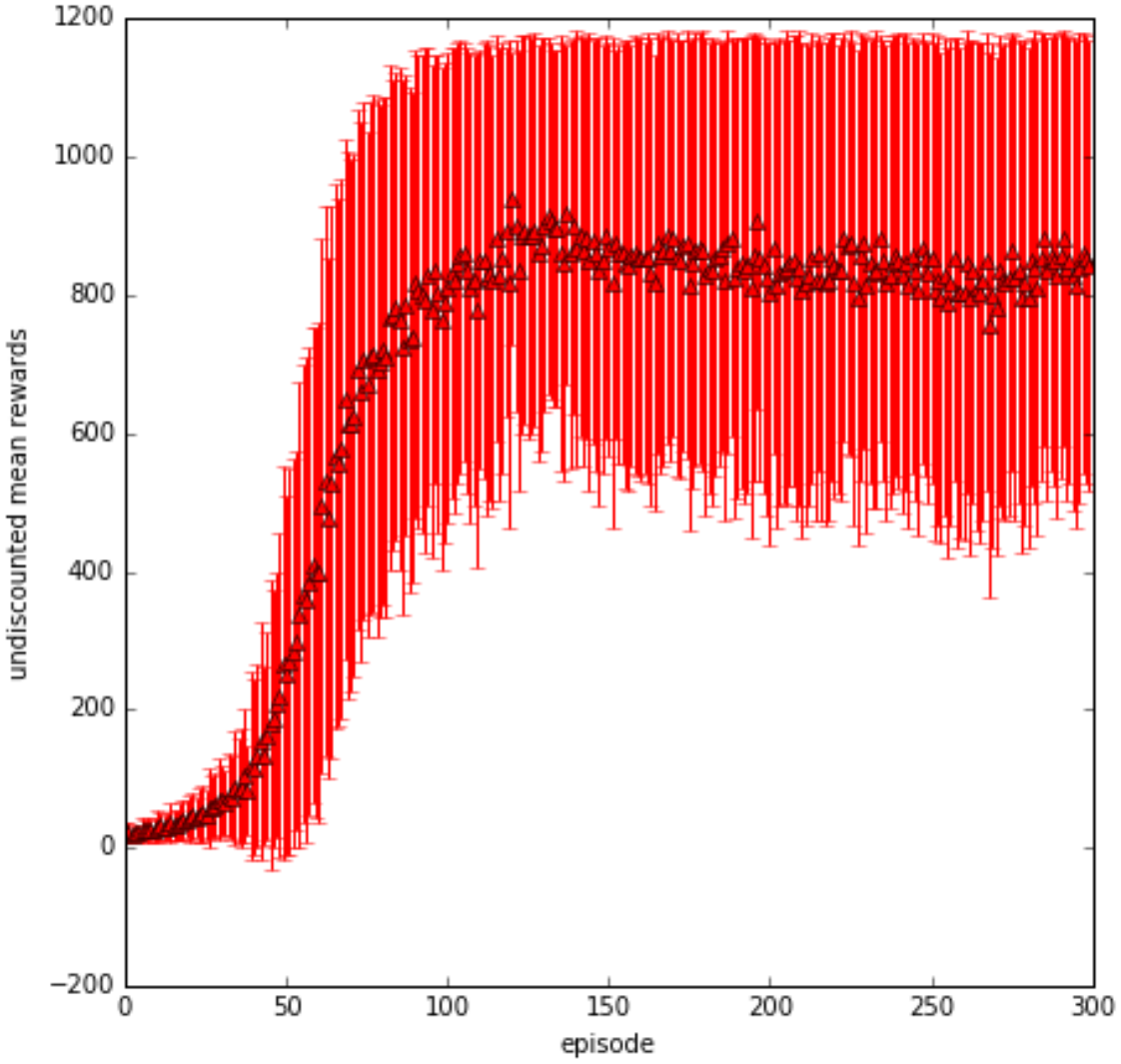
I could not find good hyper parameters for this environment.

Enter Fourier basis order: 4
Enter alpha (actor): 0.9
Enter alpha (baseline): 0.001
Enter gamma: 1
Enter number of trials to run: 10
Enter the number of episodes per trial: 200



Cart Pole

Enter Fourier basis order: 4
Enter alpha (actor): 0.00008
Enter alpha (baseline): 0.2
Enter gamma: 0.9
Enter number of trials to run: 100
Enter the number of episodes per trial: 300



Acrobat

Enter Fourier basis order: 3
Enter alpha (actor): 0.0000009
Enter alpha (baseline): 0.4
Enter gamma: 1
Enter number of trials to run: 120
Enter the number of episodes per trial: 300