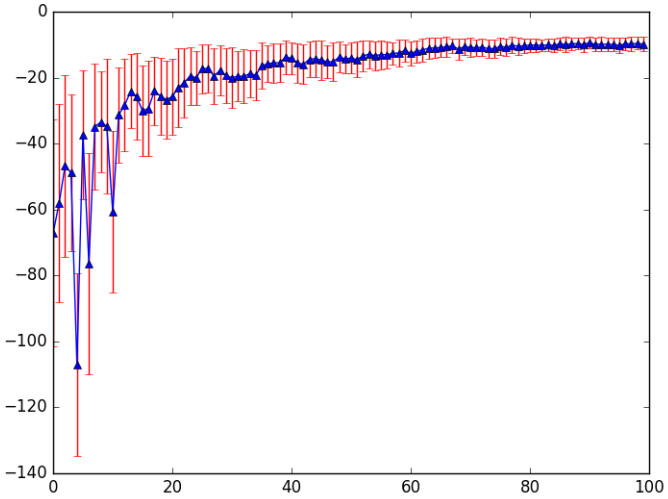
CMPSCI 687 Homework 4

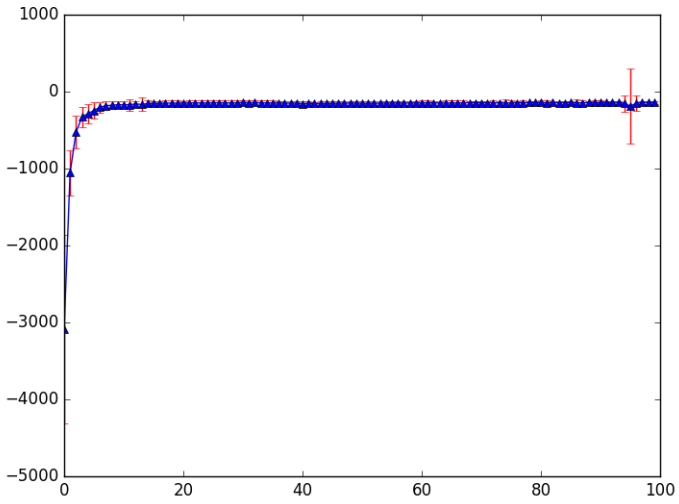
Due December 8, 2017, 11pm Eastern Time

Written Portion(75 Points Total)

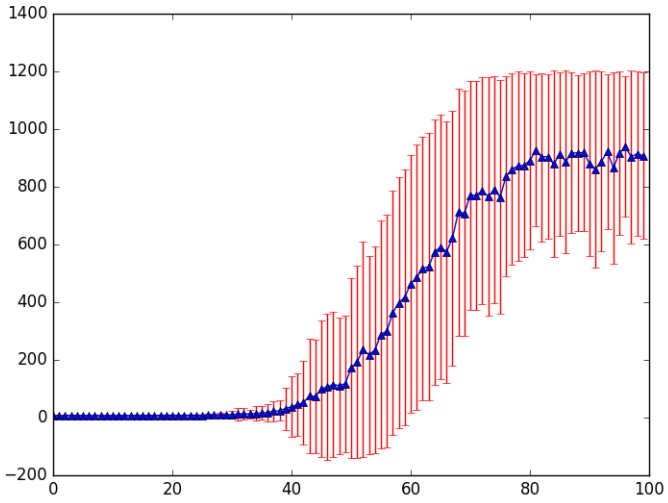
1. Ans.1 (25 Points - Q(λ))
   (A) Parameters that worked on Gridworld:
   Alpha = 0.17; Lambda = 0.3; Epsilon = 0.01; Gamma = 1



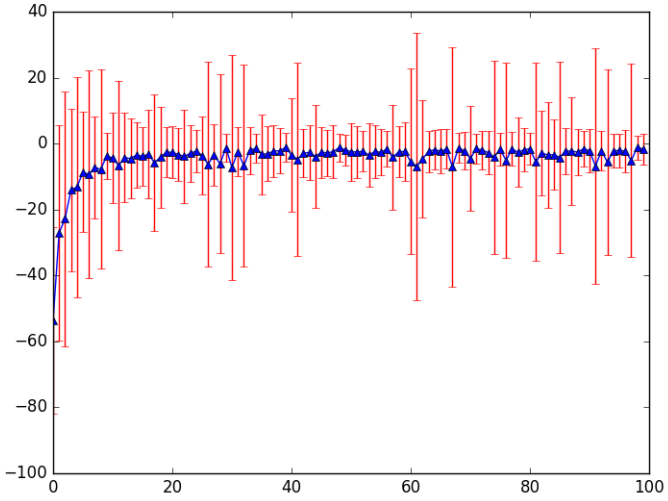
   (B) Parameters that worked on Mountain Car Domain:
   Fourier Basis = 2; Alpha = 0.01; Lambda = 0; Epsilon = 0.01; Gamma = 1



(C) Parameters that worked on Cart Pole Domain:
Fourier Basis = 2; Alpha = 0.002; Lambda = 0.85; Epsilon = 0.05; Gamma = 0.95



(D) Parameters that worked on Acrobat Domain:
Fourier Basis = 3; Alpha = 0.001; Lambda = 0.72; Epsilon = 0.01; Gamma = 1



(E)

- **Difficulty on finding the hyper parameters**: Since there are so many hyper parameters, finding the optimum one becomes very hard. The best strategy that worked for this case keeping gamma near to one.
- **Is it getting easier as you have more experience with RL algorithms**- For acrobat and cart pole it helped because of being related domain but not with others. So in general I am not sure which parameter can work and which cannot, but I am still trying to learn the range of values for which something worked on different domains.
-**Which parameters did the algorithm appear to be most and least sensitive to?** Alpha appears to me most sensitive and gamma appears to be least sensitive
-**Did any hyperparameter values surprise you?** Lambda =0 in the case of mountain car domain surprised me.

2. Ans.2 (25 Points - Actor-Critic)
(A) Proof:

$$\pi(a \mid s, \theta) = \frac{\exp^{\phi(s,a)^T \cdot \theta}}{\sum_{a'} \exp^{\phi(s,a')^T \cdot \theta}}$$

$$
\begin{aligned}
\frac{\partial \pi}{\partial \theta} &= \phi(s,a) \cdot \frac{\exp^{\phi(s,a)^T \cdot \theta}}{\sum_{a'} \exp^{\phi(s,a')^T \cdot \theta}} - \phi(s,a) \cdot [\ \frac{\exp^{\phi(s,a)^T \cdot \theta}}{\sum_{a'} \exp^{\phi(s,a')^T \cdot \theta}}]^2 \\
&\quad - \frac{\exp^{\phi(s,a)^T \cdot \theta}}{\sum_{a'} \exp^{\phi(s,a')^T \cdot \theta}} \cdot \sum_{a' \in A, a' \neq a} \phi(s,a') \cdot \pi(a' \mid s,\theta) \\
&= \phi(s,a) \cdot (\pi(a \mid s,\theta) - \pi(a \mid s,\theta)^2) - \pi(a \mid s,\theta) \cdot \sum_{a' \in A, a' \neq a} \phi(s,a') \cdot \pi(a' \mid s,\theta) \\
&= \phi(s,a) \cdot \pi(a \mid s,\theta) \cdot (1 - \pi(a \mid s,\theta)) - \pi(a \mid s,\theta) \cdot \sum_{a' \in A, a' \neq a} \phi(s,a') \cdot \pi(a' \mid s,\theta)
\end{aligned}
\tag{1}
$$

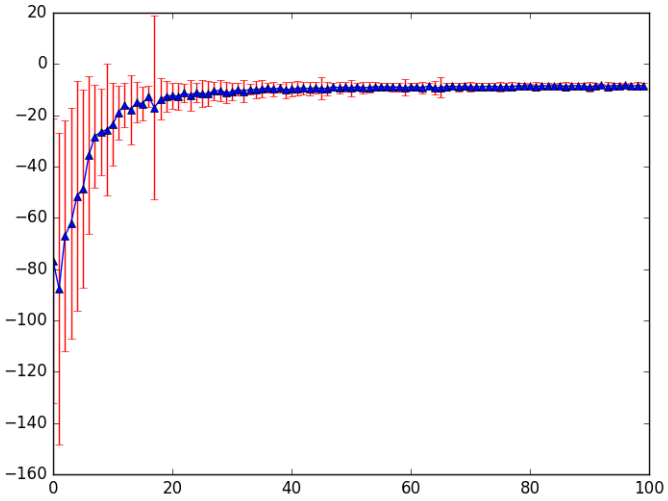
From the above result we can say that for $a' = a$

$$
\frac{\partial \ln(\pi(a \mid s,\theta))}{\partial \theta} = (1 - \pi(a \mid s,\theta)) \cdot \phi(s,a) \tag{2}
$$

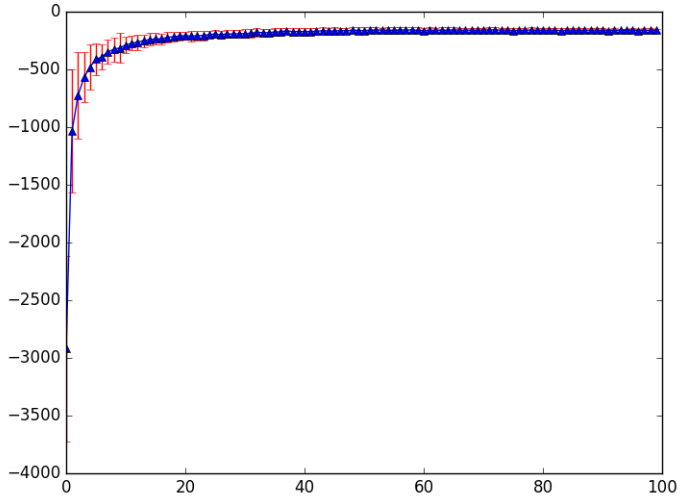
and, for all other actions, $a' \in A$, $a' \neq a$,

$$
\frac{\partial \ln(\pi(a \mid s,\theta))}{\partial \theta} = \sum_{a' \in A, a' \neq a} \phi(s,a') \cdot \pi(a' \mid s,\theta) \tag{3}
$$

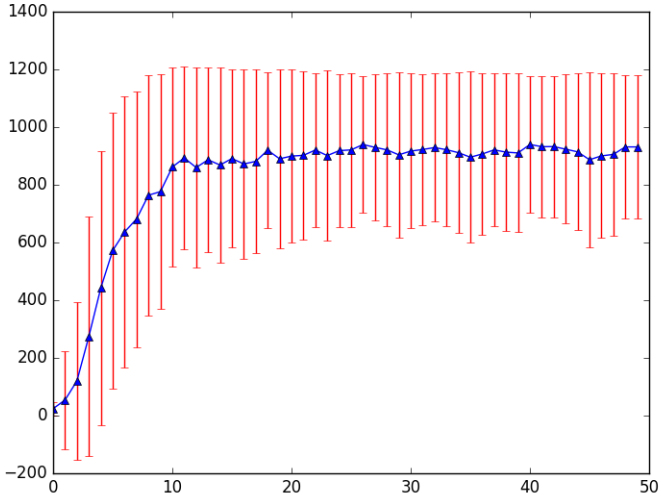
(B) Parameters that worked on Grid world:
Alpha_Actor = 0.20; Alpha_Critic = 0.10; Lambda = 0.8; Gamma = 0.95



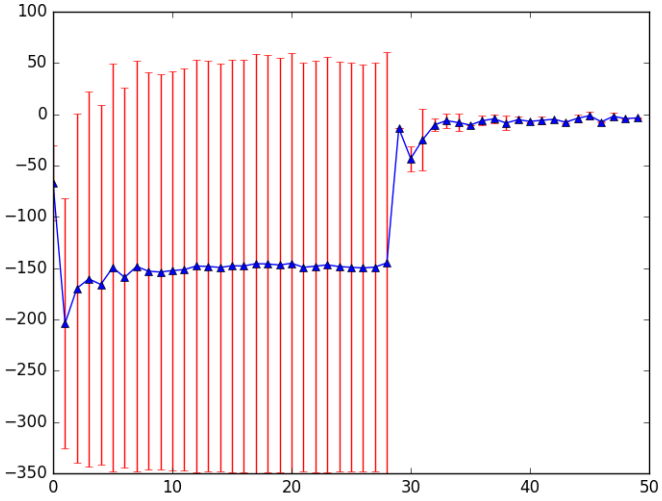
(C) Parameters that worked on Mountain Car domain:
Fourier Basis = 4; Alpha_Actor = 0.001; Alpha_Critic = 0.001; Lambda = 0.65; Gamma = 1



(D) Parameters that worked on Cart Pole domain:
Fourier Basis = 4; Alpha_Actor = 0.001; Alpha_Critic = 0.001; Lambda = 0.5; Gamma = 0.95

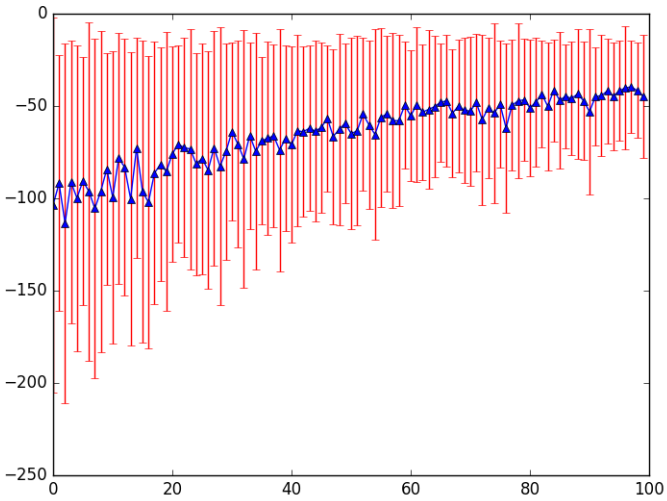


(E) Parameters that worked on Acrobat domain:
Fourier Basis = 4; Alpha_Actor = 0.001; Alpha_Critic = 0.001; Lambda = 0.15; Gamma = 1



(F) Hyper parameter: Except the Grid world domain, alpha for other domain need not be changed much. The most sensitive hyper parameter in this domain is lambda.
Comparison with Qlambda: It is very slow in training.

3. Ans.3 (25 Points - Q(λ))
(A) Parameters that worked on Gridworld:
Alpha_Actor = 0.004; Alpha_Baseline = 0.004; Gamma = 0.9

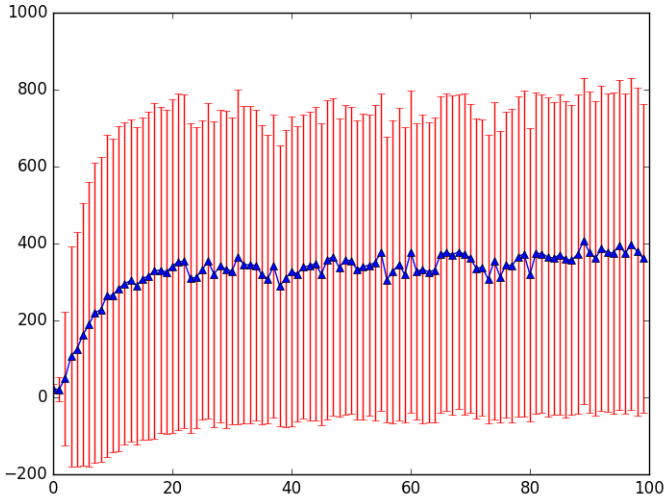


(B) Best paramaters that worked on Mountain car domain:

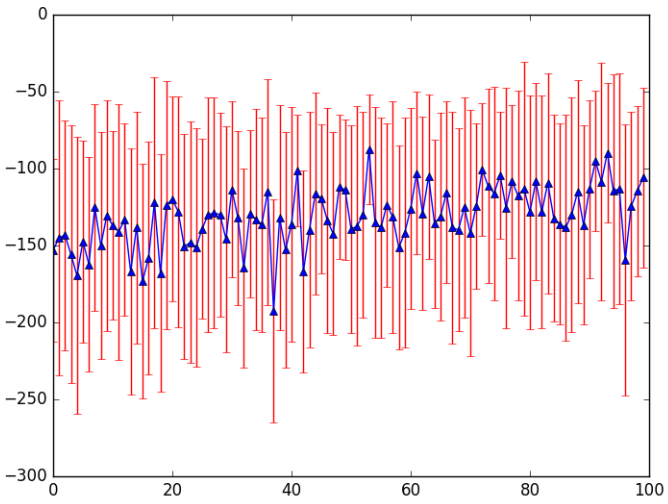
Fourier Basis = 2; Alpha_Actor = 0.0001; Alpha_Baseline = 0.0001; Gamma = 0.91

Reason for this not working well : Since Reinforce is similar to Monte-Carlo in terms of high variance, in order to get good parameters we have to run very high number of episodes per trial. Due to time constraint we are not able to run such high number of experiments.

(C)Parameters that worked on Cart Pole domain:
Fourier Basis = 2; Alpha_Actor = 0.01; Alpha_Baseline = 0.06; Gamma = 0.91



(D)Parameters that worked on Acrobat domain:
Fourier Basis = 2; Alpha_Actor = 0.00001; Alpha_Baseline = 0.00001; Gamma = 0.95



(E)
-**Difficulty on finding the hyper parameters**: It has been the most difficult experience in finding optimum parameters for this algorithm in the whole exercise.
-**Comparison with previous two**: From what I saw with experiments this is the highest unstable methodology and is very sensitive to hyper parameters.