REINFORCEMENT LEARNING

ASSIGNMENT 2

Policy Gradient - Reinforce

SAYAN SINHA (16CS10048)

This project has been taken up as a part of Assignment 2 of Reinforcement Learning course (CS60077).

1. Learning algorithm

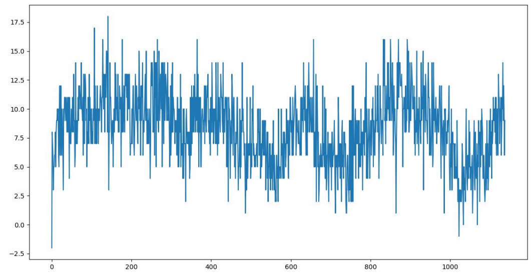
Monte Carlo method works on the basis of sampling of episodes. REINFORCE is a Monte Carlo method which works in a similar fashion. The policy is supposed to learn a function to maximise the cumulative reward. The reward might be normalised in order to reduce variance. REINFORCE is an on-policy algorithm, that is, it samples from the policy that the algorithm updates. It works using gradient ascent which computes the required derivative by partial differentiation of the objective wrt the parameter of the policy. The function itself, however, is parameterised by a neural network. The discounted rewards (or optionally up to a horizon) is calculated and the parameters are updated.

2. Network architecture

The neural network features 37-dimensional input and a 4-dimensional output. There are 2 hidden layers, with 32 and 16 neurons in each. The activation function at the hidden layers is ReLU while that in the output layer in Softmax.

3. Results

The average value obtained was 8.33. The inference curve looks like:



During training, stability was

reached after around 4000 episodes.

4. Future work

It is apparent that an on-policy solution does on suit the environment. We need to start from scratch using an off-policy method. Usage of Q learning could be attempted.