

CAP 5636: Assignment 4

Problem 1:

Implement an explicit policy for the mountain car environment without using any learning algorithm. Explain in detail your reasoning behind your policy and run several test episodes to measure its performance.

Approach:

For the approach I used, the action variable can have 3 values which can be 0, 1 or 2. They represent 'left', 'no push' and 'right' respectively. The variable observation have 2 parameters position and velocity. The explicit policy that I have used is if the velocity is negative then I return 0 where it knows that it needs to move left. Else, I return 2 where it moves right.

Problem 2:

Implement an explicit policy for the cartpole environment without using any learning algorithm. Explain in detail your reasoning behind your policy and run several test episodes to measure its performance.

Approach:

The aim of the cartpole problem is to balance the pole on the cart. The action variable may take two values 1 and 0. if you want the cart to move right the value is 1 and if you want it to move left the value is 0. The observation number will either be positive or negative. Here, I have used velocity as a parameter for the policy function. If the velocity is negative I returned 0 which indicates to go left, else I return 1 to go right.

Problem 3:

Apply the cross-entropy method to mountain car. Explain how many episodes are needed to learn a good policy. Explain which reward you use (original, modified).

Approach:

Explanation: The Cross-Entropy method is a technique frequently used for rare event simulation and optimization. The Cross Entropy method is an iterative method, which involves the following two phases:

- 1) Generation of a sample of random data according to a specified random mechanism.
- 2) Updating the parameters of the random mechanism, on the basis of the data, in order to produce a "better" sample in the next iteration.

Cross-entropy can be applied to reinforcement learning by learning (optimizing) a value function. The cross-entropy method for Mountain Car problem took 140 Episodes and 47 iterations to achieve an average score of 90.83.