page2image568

Reinforcement Learning  
 Lesson-End Project Problem Statement



**Solve the CartPole Environment Problem Using Q-Learning**

**Problem Statement**: Solve the control problem in the CartPole environment using the Q-learning method. While solving the CartPole environment, run it for 500 episodes where the epsilon is a fraction of 1/sqrt(n+1) and the discount factor is 0.99. Also, plot all the rewards and their running averages for all the episodes.

**Hyperparameters:**

* gamma = 0.99
* epsilon = 1/sqrt(n+1)  
    
  n= It is a random value in the list of total rewards, i.e., N.

**Prerequisites:**

* Python
* OpenAI Gym

**Steps to perform:**

1. Preprocessing steps:
   1. Minimize a regularized empirical loss
   2. Preprocess
   3. Optimize the environment using sklearn
2. Also, use SGDRegressor to transform features of the environments and design a model (initialize the environment, predict the states, update the states, actions and rewards, and define a sample action with the epsilon)
3. Find the total rewards:
   1. Initialize the number of steps and total rewards
   2. Limit the number of steps (e.g. fixing it to <2000)
   3. Assign the reward value of -200 when the episode is done
   4. Update the model based on rewards while considering the gamma
4. Write function to plot the rewards and run averages of the rewards
5. Run the environment for 500 episodes where epsilon is a fraction of 1/sqrt(n+1) and discount factor is 0.99 and then plot the rewards and their running averages