Reinforcement Learning - Series - 4

vp2504

December 7, 2024

Q-learning Algorithm

- **Initialize the episode:** Start each episode with an initial state $x_0 = [0, 0]$.
- **Policy selection:** For each step in the episode, choose an action u_n based on an ϵ-greedy policy to balance exploration and exploitation.
- **State transition:** Compute the next state x_{n+1} resulting from applying u_n.
- **Target computation:** Calculate the target:

$$y_n = g(x_n, u_n) + \alpha \min_a Q(x_{n+1}, a),$$

  where g is the immediate cost, and α is the discount factor.
- **Neural network update:** Perform a single step of stochastic gradient descent (SGD) on the neural network's parameters to minimize the loss:

$$\mathcal{L} = \big(Q(x, u) - y_n\big)^2.$$

Repeat this process for the entire episode and iterate for multiple episodes (17500) to train the Q-function effectively.

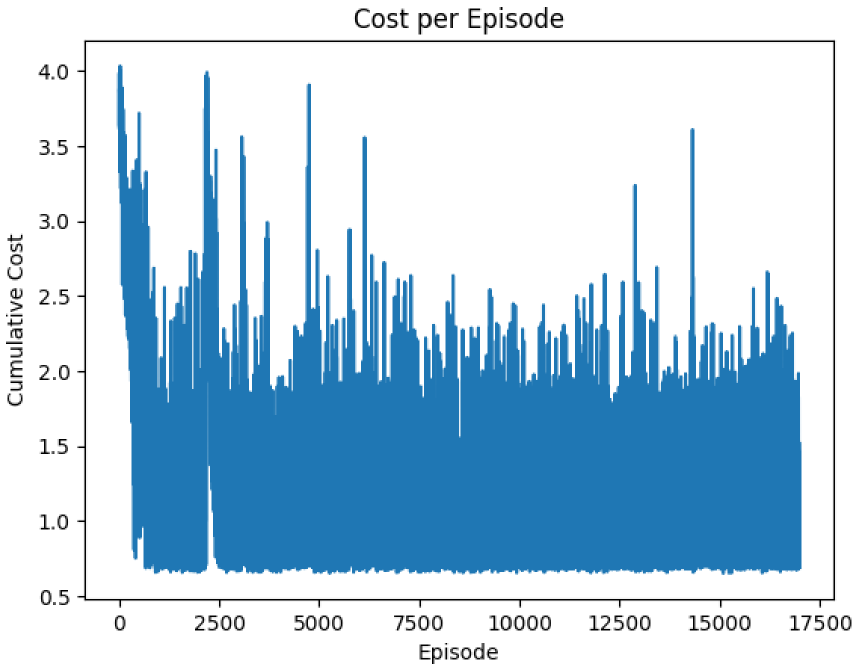

Figure 1: Cost per Episode

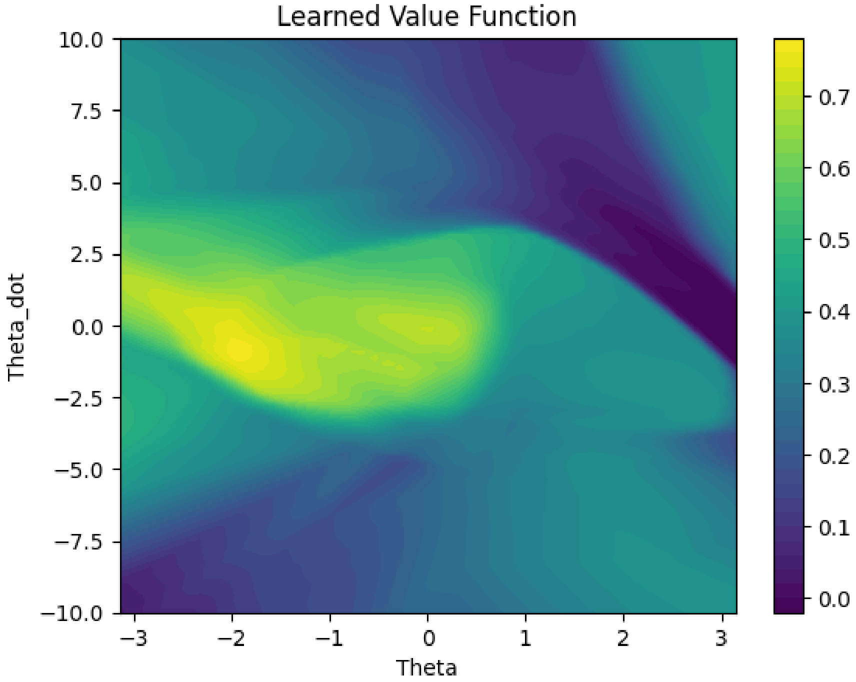

Figure 2: Learned Value Function

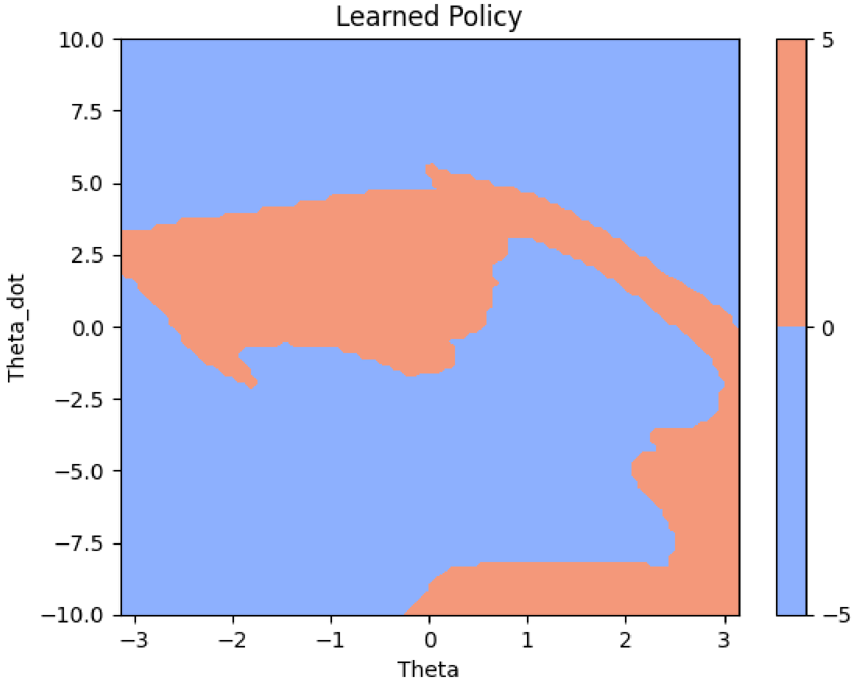

Figure 3: Learned Policy