

Homework 1

Cora Zhang, beileiz@cs.cmu.edu

1 Reinforcement Learning via Finite Difference

1.1.2 Todo: Run CartPole with Linear Policy

- $[1, 0, 0, 0]$ with all weight on location parameter, the pendulum fell very soon, the sum of reward is 9.0
- $[0, 1, 0, 0]$ with all weight on linear velocity, the pendulum fell very soon, the sum of reward is 9.0
- $[0, 0, 1, 0]$ with all weight on the angle parameter, the pendulum swung left and right for a while then fell. the sum of reward is 47.0
- $[0, 0, 0, 1]$ with all weight on the angular velocity, the pendulum maintained balance in a small angle until the cart moved out of the frame. the sum of reward is 253.0
- $[0, 0, 1, 1]$ with half weight on the angular velocity, half on the angle, the sum of reward is 500.0

1.1.3 Todo: Sample random policies and observe behavior

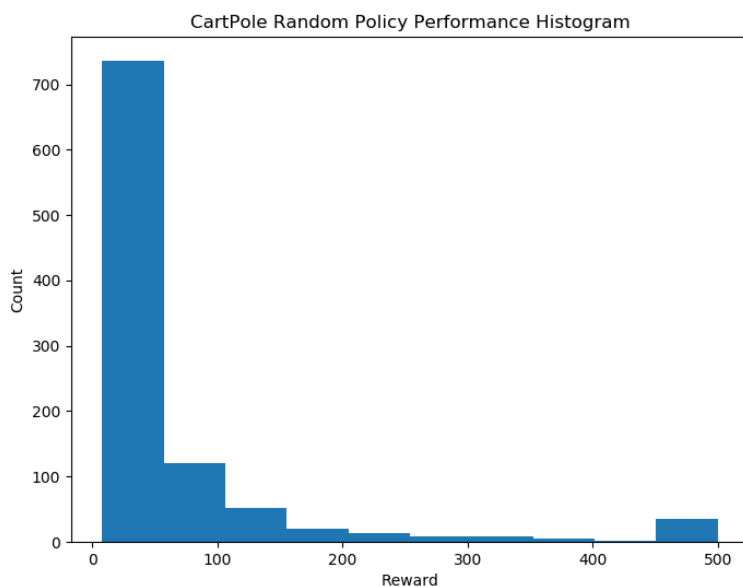
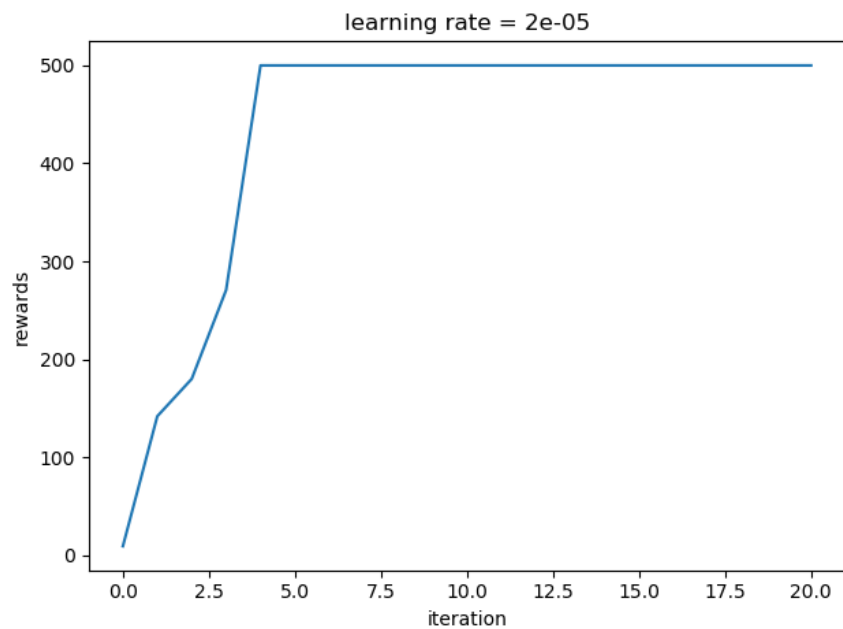


Figure 1: Percentange of policies that achieved full reward: 0.033

1.2.4 Todo: Tune Learning Rate and Plot Learning Curve



- Learning rate: 2e-05
- Final params: [-0.01165 0.0887 0.11715 0.156]

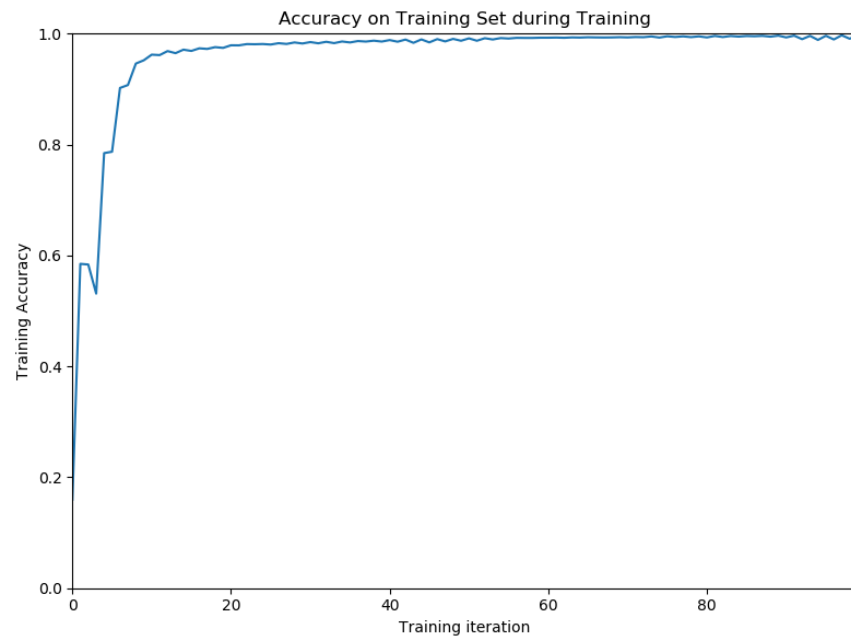
2 Supervised Learning: Classification via Logistic Regression

2.1.1 Todo: Derive the Logistic Regression Gradient Update Formula

$$\frac{\partial l(\tilde{w})}{\partial w_j} = \sum_{i=1}^N [y_i - \sigma(\tilde{w}^T \tilde{x}^i)] x_j^i \quad (1)$$

where j denote the jth feature in ith sample (x), also the jth weight in (w)

2.3.2 Todo: Run Training Script



- Achieved final training accuracy: 0.997
- Achieved test accuracy: 0.980