CS440 MP4 Report

Andong Jing ajing2 - 3 credit In charge of

Siping Meng smeng10 - 3 credit In charge of 1.1/1.2

Siyu Tao siyutao2 - 3 credit In charge of

# 1. Q-Learning

## 1.1 Single-Player Pong

1. Report and justify your choices for α, γ, exploration function, and any subordinate parameters. How many games does your agent need to simulate before it learns a good policy?

2. Use α, γ, and exploration parameters that you believe to be the best. After training has converged, run your algorithm on 200 test games and report the average number of times per game that the ball bounces off your paddle before the ball escapes past the paddle.

3. Include “Mean Episode Rewards vs. Episodes” plot for both Q-Learning and SARSA agents and compare these two agents.

## 1.2 Environment Changed

1. Describe the changes you made to your MDP (state space, actions, and reward model), if any, and include any negative side-effects you encountered after doing this.

2. Describe your method of training agent A and tell us why it works.

3. Include two “Mean Episode Rewards vs. Episodes” plots and compare these two agents.

## 1.3 Extra Credit

# 2. Deep Learning (Pong)

## 2.1 Cloning the Behavior of an Expert Player

1. Answer the following question in the report: What is the benefit of using a deep network policy instead of a Q-table (from part 1)? (Hint: think about memory usage and/or what happens when your agent sees a new state that the agent has never seen before).

2. Implement the forward and backwards functions of a neural network and give a brief explanation of implementation and architecture (number of layers and number of units per layer).

3. Train your neural network using minibatch gradient descent. Report the confusion matrix and misclassification error. You should be able to get an accuracy of at least 85% and probably 95% if you train long enough. Report you network settings including the number of layers, number of units per layer, and learning rate.

4. Plot loss and accuracy as a function of the number of training epochs. You do not need to do a train-validation split on the data, but in practice this would be helpful.

5. Report the number of bounces your agent gets. It should be around 8 bounces.

## 2.2 Deep Q-Learning

## 2.3 Extra Credit