Reinforcement Learning: Chapter 1 Exercises

Name: Eli Andrew

(a) **Exercise 1.1: Self-Play** Suppose, instead of playing against a random opponent, the reinforcement learning algorithm described above played against itself, with both sides learning. What do you think would happen in this case? Would it learn a different policy for selecting moves?

- The random opponent described in the example is assummed to play with a constant policy. So, the first change in this new scenario is that the algorithm is playing against an opponent that changes its policy over time. Now that the opponent will learn from its experience and update its policy, both players will be simulataneously adjusting their policy based on what the other is doing. This will lead to both players learning the optimal policy for an optimal opponent which in tic-tac-toe means a draw. However, since in this scenario the opponent is not just another algorithm but the *same* algorithm that is trying to win for one side, we will probably see the algorithm playing strong moves for its side and weak moves for its opponent in order to make it more likely that it will win.

(b) **Exercise 1.2: Symmetries** Many tic-tac-toe position appear different but are really the same because of symmetries. How might we amend the learning process described above to take advantage of this? In what ways would this change improve the learning process? Now think again. Suppose the opponent did not take advantage of symmetries. In that case, should we? Is it true, then, that symmetrically equivalent positions should necessarily have the same value?

- To amend the learning process to take advantage of symmetries we can modify our state representations to view symmetric states as the same state. This would allow us to have a smaller search space and to therefore learn the optimal policy faster. In tic-tac-toe this would result in our state space being reduce from 9 states to 6 states. If we do not take advantage of symmetries then we are allowing our learning process to pick up on alternate outcomes of symmetric states. For example, it could be the case that our opponent plays a different way in symmetrically equivalent states and if we had a reduced state space then we wouldn't pick up on those differences. So, it likely does not make sense to assume that symmetrically equivalent positions should necessarily have the same value, but should be considered on a case by case basis.

(c) **Exercise 1.3: Greedy Play** Suppose the reinforcement learning player was *greedy*, that is, it always played the move that brought it to the position that it rated the best. Might it learn to play better, or worse, than a nongreedy player? What problems might occur?

- A greedy player would essentially never explore a different policy and would always take what it currently views as the best possible action. Originally, this means that from any given state the player would always go to the same state.