Chandler Justice - A02313187

ECE-5930-004 Homework #1
Due: September 17, 2025 @ 23:59

Define the following elements of the Marcov space for this assignment:

S: This is the state of the game. Since we are playing rock paper scissors we want to represent a 3×3 array wherein each element can be occupied by either an X or O mark, or an empty string. Each player uses either X or O to mark their positions on the board, meaning the empty string represents a position where neither player has placed a mark. For simplicity of implementation, we will flatten the array to a 9 element 1D array. This will not effect the functionality of the agent or the game board.

A: Each player must mark a position on the board at every timestep, and marks can only be placed where a mark does not already exist. This means each agent has the action space of placing their respective mark on an empty position on the board.

$P(s'|s, a)$: Probability of transitioning from one game board to the next. For us, this can be thought of as a tree wherein each branch is a subsequent board state, and is associated with some probability of entering that state from the current node. Terminating nodes of the tree represent one of the players winning or a draw. While the opponent is playing randomly, our state transitions can still use this structure. The nodes where the random opponent plays will have a probability distribution characteristic of their random behavior.

$R(s, a, s')$: Rewards will be given only at the end of every game and are determined by if the agent wins, loses, or draws. This means s' should be `none` as the reward is always evaluated at the terminal state. The reward function is defined as

$$R(s, a, s') = \begin{cases} 1 & \text{winning terminal state} \\ -1 & \text{losing terminal state} \\ 0 & \text{otherwise} \end{cases}$$

Other Details: The past decisions are not discounted, meaning $\gamma = 1$. For our agent, this means past decisions are just as important as present decisions. Since we only evaluate reward at the end of the game considering the whole game as equally important seems like the right move. Additionally, a tic tac toe game will only last at most 9 moves further negating the need to discount past moves.

An episode in our environment is defined to have an initial state of an empty board, and a terminal state such that a player wins or a draw is reached. Due to the definition of the environment, the environment has the Marcov property and is the agent-environment pairing can be formulated as a Marcov decision process.