Reinforcement Learning
4x4x4 Tic-Tac-Toe

The game is first setup with a 4x4x4 array of empty values. From there, each player takes turns making a move.
During each move the player first checks if the game is about to be over. If so, it makes the move to win the game, or makes the move to stop the opposing player from winning.
If the game is not about to be over, the player moves based on exploring or exploiting. After each move, the utility values are updated.

Explore

The explore function is called when the game is in early stages and curiosity and learning is more rewarding than going to the square with the highest utility. The explore function finds the square that has been visited the least number of times and makes that move.

Exploit

The exploit function happens in later stages of the game when winning outweighs curiosity or learning for the future. The exploit function finds the maximum utility value, from the current utility grid for the player at that state, and places the move at that given spot as long as it is empty.

Updating Utility

After each move the utility is updated for each block of the board for each of the states. The utility's for each player are stored in 'x_score' and 'o_score.' The utility values range from -1 to 1. The game is initialized with each block having 0 utility. States that lead to a loss for the player get negative values and ones that lead to a win for the player get a positive values. When the player is exploiting, the utility values are used to make the move for the given board state.

Contribution

Ahmet Hatip- game setup/playing game, updating utility values, explore vs. exploit
Dominick Giordano- exploring, checking if game is about to be over
Nick Egnatz- exploiting, shell conversion, write-up
All- brainstorming, creating helper functions