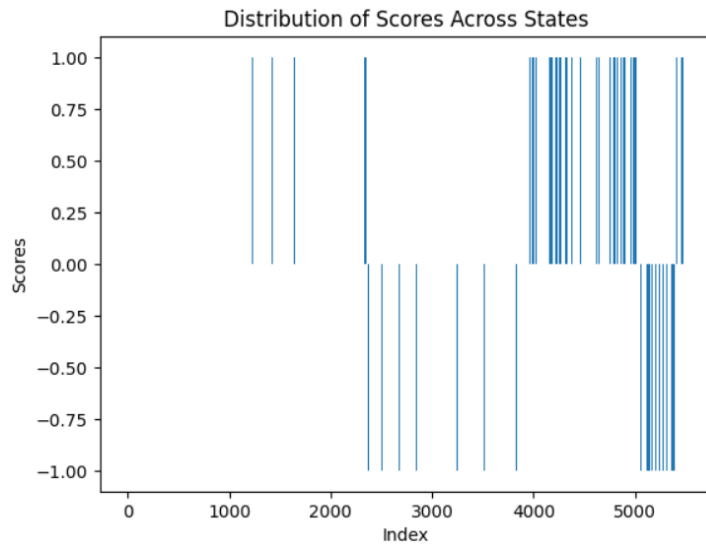# REPORT

**Purpose:** Purpose of this report is to show the analysis of the change of score values in Reinforcement learning in tic tac toe game and to compare the efficacy of AI game play against a human player using Mini-Max and Reinforcement Learning.

## Analysis for reinforcement algorithm:



Distribution of Scores Across States

Initially the values for every state were zero for the draw terminals and other states which were not terminals.The winning states had value 1 and losing states had value -1. Therefore there were only three values either 0,1 or -1.



Distribution of Scores Across States
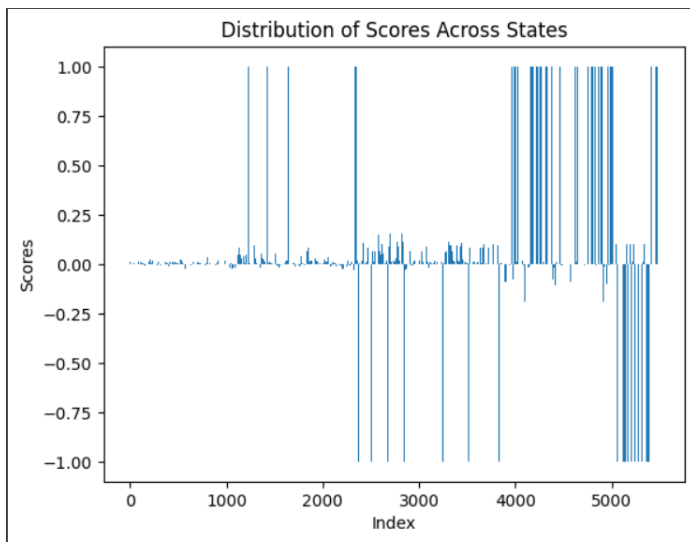
Value of alpha = 0.10;
No of iterations = 1
It can be observed that values has changed but not as much as initial values; they are still almost the same as the initial values.

Distribution of Scores Across States
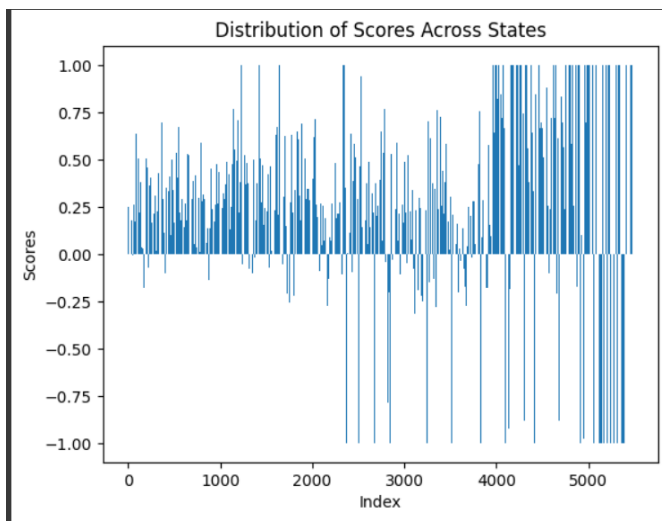
Value of alpha = 0.8
No of iterations = 1
On taking the value of alpha quite high I observed that in only one iteration there was drastic change in the values of the states.



Distribution of Scores Across States
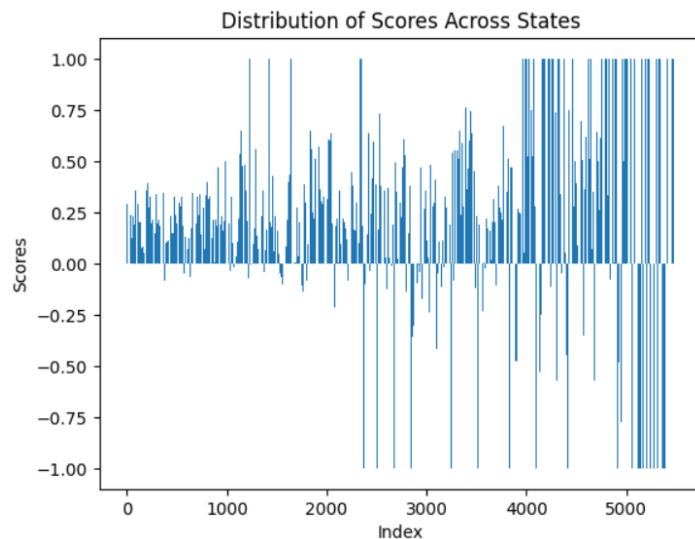
Value of alpha = 0.10
No of iterations = 100
It can be seen the values are being distributed in the range(-1,1).I stopped my training after these many iterations.
Reason for stopping the training:

1. The graph was not changing much.
2. All the values were distributed in the graph and were taking multiple values and not the same As they were there when I took the less number of repeated iterations.
3. After 100 more iterations on changing the value of alpha there was no change in the graph/ values .

We made tic tac toe game by two ways:

1. Using Minimax Algorithm
2. Using Reinforcement Algorithm

Observations of gameplay when game was played by minimax

1. I was never able to win from the computer, each match has either draw, or win from the computer.
2. The computer was giving the best moves and was tackling my tricks and stopping me every time from winning the game.

Observations of gameplay when game was player using reinforcement learning

1. In most of the cases I was winning the game and in few cases the computer was winning the game.This was different from minmax where I was not winning even once.
2. On changing the value of alpha and on changing the values of the states the experience was changing. Sometimes the game was becoming easy to play, sometimes it was hard to play.

| Aspect | Minmax | Reinforcement |
|---|---|---|
| Winning rate | Greater than Reinforcement | < 30%(in my case) |
| Losing rate | 0 | > 60% |
| Adaptability | Limited adaptability since it relies on a fixed set of rules and predefined strategies | Highly adaptable, capable of learning and evolving strategies over time through trial and error. |
| Training cost | Low training cost as there is no learning phase; computation is mainly used during gameplay. | High training cost in terms of computational resources and time, especially for deep reinforcement learning in complex environments. |
| Human-like behavior | Tic tac using minwax wasn't human-like gameplay. | Gameplay was human-like behavior. |
| Exploration vs exploitation | Does Not uses exploration | Balanced exploration and exploitation,allowing to make it unpredictable and behaves like human |