Name: Shaun Mathew

Assignment #1

**Question 3:**

*Exploratory vs Greedy:*

During training, we trained the exploratory and greedy TicTacToe AIs with an alpha function that produced a constant value of 0.1. Furthermore, both AIs are trained till convergence is achieved (i.e. the delta between state values at the start and beginning of an episode is <= 0.﻿00001) or the number of iterations taken exceeds 20,000. The epsilon value for the exploratory AI was set at 0.2. Also, during test time, exploratory moves are not performed, and only optimal action as determined by the AI is chosen, since it does not make sense for an AI to perform exploratory moves during a competition.

The results are as follows:

Greedy

﻿Training Row AI

Converged in 20000 iterations

Training Col AI

Converged in 92 iterations

Training Diagonal AI

Converged in 262 iterations

Number of row opponent victories with exploration 0.0: 500

Number of col opponent victories with exploration 0.0: 500

Number of diag opponent victories with exploration 0.0: 500

Exploratory:

﻿Training Row AI

Converged in 12676 iterations

Training Col AI

Converged in 20000 iterations

Training Diagonal AI

Converged in 2984 iterations

Number of row opponent victories with exploration 0.2: 500

Number of col opponent victories with exploration 0.2: 500

Number of diag opponent victories with exploration 0.2: 500

As we can see from the results, exploratory moves significantly increase the number of iterations it takes for the AI to converge. This is expected since more values of the state value table are perturbed and hence it will take more iterations for the values to reach a steady value. We also expect the state value table to be a better representation of the win probabilities since more states are likely to be visited. With regard to the number of victories, the number of wins accrued by the exploratory and greedy AIs are the same.

*Comparing Convergence of Different Sequences:*

We tried three different sequences of a\_n. a\_n\_1 is a sequence of solely 0.1 for every time step. a\_n\_2 is the sequence mentioned in the book i.e. 1/t where t is the episode number. a\_n\_2 is constructed the following way ﻿(1/((t+1)/50))\*0.1 if (t+1) >= 50 else 0.1. So the sequence starts at 0.1 and decreases to half that previous value every 50 steps. Furthermore, had an exploration probability of 0.1 during training.

The results are as follows:

**a\_n\_1 (i.e. [0.1, 0.1, 0.1 …. ]**

Training Row AI

Converged in 11640 iterations

Training Col AI

Converged in 20000 iterations

.

Training Diagonal AI

Converged in 8344 iterations

**a\_n\_2 (i.e. [1.0, 0.5, 0.333, 0.25, …. ]**

Training Row AI

Converged in 1868 iterations

Training Col AI

Converged in 3269 iterations

Training Diagonal AI

Converged in 5517 iterations

**a\_n\_3 (i.e. [0.1,…0.05,…0.025… ]**

Training Row AI

Converged in 3460 iterations

Training Col AI

Converged in 6376 iterations

Training Diagonal AI

Converged in 10241 iterations

Number of row opponent victories: 100

Number of col opponent victories: 100

Number of diag opponent victories: 100

Number of row opponent victories: 100

Number of col opponent victories: 100

Number of diag opponent victories: 100

Number of row opponent victories: 100

Number of col opponent victories: 100

Number of diag opponent victories: 100

From the results we see that for an exploratory AI with probability 0.1, sequence a\_n\_1, the sequence with a constant step size of 0.1, converges the slowest; whereas, sequence a\_n\_2, the sequence whose alpha value was computed by 1/t, converges the fastest.