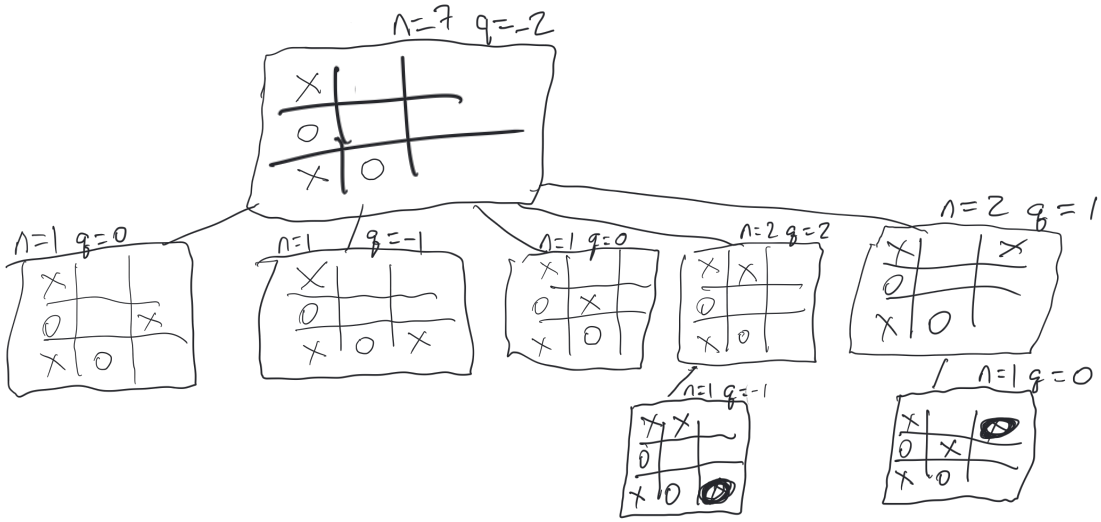
Reinforcement Learning—HW 9

April 13, 2020

Here's a partially-expanded Monte Carlo Search Tree for determining the next move in Tic-Tac-Toe:



Assume that c, the constant used for UCT (Upper Confidence Bound applied to Trees) is 0.5.

1. In this Tic-Tac-Toe game, whose turn is it (X or O)?
2. From your knowledge of Tic-Tac-Toe, what is the best move for that player?
3. For this question, you'll do one iteration of the MCST algorithm (Selection/Expansion/Simulation/Backup)
   (a) *Selection* What leaf node is the next to be selected (show your work)?
   (b) *Expansion* What child node is expanded?
   (c) *Simulation* Do a rollout (simulation). Use some form of randomization (coin flips, random number generation in Excel or on a website, coin toss, etc.) to do the simulation. What is the result? If the result is a tie, rerun the simulation until you get a decisive winner.

(d) *Backup* Backup the result through the MCST.

(e) If this were the last simulation of the MCST, what action would be chosen and why?